



HAL
open science

Structure and dynamics of the penguin synnomes : understanding seabird life history and response to climate change through population genomics

Robin Cristofari

► To cite this version:

Robin Cristofari. Structure and dynamics of the penguin synnomes : understanding seabird life history and response to climate change through population genomics. Populations and Evolution [q-bio.PE]. Université de Strasbourg, 2016. English. NNT : 2016STRAJ005 . tel-01423915

HAL Id: tel-01423915

<https://theses.hal.science/tel-01423915>

Submitted on 1 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École Doctorale Vie et Santé

CNRS-IPHC-DEPE

THÈSE

présentée par :

Robin CRISTOFARI

soutenue le : 23 février 2016

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline / Spécialité : Biologie / Écologie, Génétique des populations

Structure and dynamics of the penguin synnomes

Understanding seabird life history and response to climate change
through population genomics

THÈSE dirigée par :

Dr LE MAHO Yvon
Dr ANCEL André

Professeur émérite, CNRS-Université de Strasbourg
Directeur de recherches, CNRS-Université de Strasbourg

Co-encadrée par :

Dr LE BOHEC Céline
Dr TRUCCHI Emiliano

Chargée de recherches, CNRS-Université de Strasbourg
Associate researcher, Universität Wien

RAPPORTEURS :

Dr BEAUGRAND Grégory
Dr DAVEY John

Directeur de recherches, CNRS-Université de Lille
Principal investigator, University of Cambridge

PRESIDENT DU JURY :

Dr POTIER Serge

Directeur de recherches, Université de Strasbourg

1 *Acknowledgements*

2 If ever a PhD was the fruit of an enthusiastic and supportive team, it is the work I present here. It
3 was only made possible thanks to the unfailing dedication and patience of Dr Céline Le Bohec
4 and Dr Emiliano Trucchi, who assumed the doubly essential roles of supervisors and friends. I
5 sincerely thank you both, with the joyful certitude that this work is just the beginning of a long
6 series..!

7 I warmly thank Pr Yvon Le Maho for giving me the chance to enter the insular world of polar bi-
8 ology, and for his continued guidance throughout this endeavour. His support made it possible
9 not only to design this work, but also to realise it with the help of the Ecophy team: André An-
10 cel, who co-directed this work, Julien Courtecuisse and Nicolas Chatelain, who invented every
11 possible penguin-tracking contraption used here and there in this work, Aymeric Houstin,
12 Matthieu Boureau, Gildas Lemonnier and Anne-Catherine Klein on the field. And, of course, the
13 whole DEPE team in Strasbourg for their support and welcome every time I visited.

14 I spent the best days of this work on the research stations of Dumont D'Urville, in Adélie Land,
15 and Alfred Faure, in Crozet Archipelago: I sincerely thank the French Polar Institute for this op-
16 portunity, as well as for their essential technical support that made this work possible, and Marie
17 Pellé, Quentin Schull, Gaël Cardinal, Caroline Bost and Paco Decina for making these days the
18 happiest ones.

19 I whole-heartedly thank Pr Nils C. Stenseth and the CEES team, who welcomed me at the Uni-
20 versity of Oslo, where almost all this work was done. Nanna Winger Steen and Emelita Rivera
21 Nerli had the patience and kindness to help through the long and messy lab-work. The Norwe-

1 gian Sequencing Centre team's creativity and skill made this work possible: I am extremely grate-
2 ful to Morten Skage's ideas and hard work, Marianne Selander Hansen's skill and optimism, and
3 Ave Tooming Klunderud's support, as well as to Lex Nederbragt's experience and help.

4 This work was finally achieved thanks to the support of Pr Denis Allemand and the Centre Sci-
5 entifique de Monaco, who welcomed me for the final months of this work, and provided contin-
6 ued collaboration and support for its whole duration. I thank the CSM team for its warm wel-
7 come - and particularly Leila Ezzat for proofreading this thesis.

8 I happily thank Pr Guillermo Luna Jorquera and the merry team of the EDAM, at the Universi-
9 dad Catolica del Norte, in Coquimbo - Claudia Fernandez Zamorra, Diego Miranda Urbina,
10 and Nicole Licuime Castillo for their invaluable help on the field, Nicolas Gouin and Rasme
11 Hereme from the CEAZA in the lab, and Paula Plaza and Maritza Cortes Labra for their ideas
12 and experience.

13 I am sincerely grateful to Dr John Davey and Dr Grégory Beaugrand who took the time to re-
14 view in great depth this hefty manuscript, to Dr Serge Potier who also presided the jury during
15 its defense, and to Dr Giorgio Bertorelle, who, after contributing largely to this work, travelled a
16 long way to discuss it at length. Defending this thesis in front of such a jury has been an honour
17 - but also a great and unexpected pleasure.

18 Finally, although a PhD thesis is a rather dreary thing to dedicate, all my thanks go to my family
19 - and in particular to my parents, to Cécile, who gave me the empirical demonstration of how to
20 defend a PhD, and to Hélène, who is nearly there too - and to the crew of the schooner La
21 Sonate - Norith Eckbo, Jean-Christophe Gairard, Florian Maury and Marc Danger - for opening
22 the windows of the world for a breath of fresh air after all this exhausting work.

1 *Table of contents*

2 Acknowledgements3
3 Table of contents.....5
4 Articles and communications.13
5 Foreword.....15
6
7 **Chapter 1. Introduction21**
8 Populations, philopatry and dispersal.....21
9 §-1. *Species: type and repetition*.....21
10 §-2. *Population, deme, colony*.....25
11 §-3. *What is a Synnome?*.....28
12 §-4. *Dispersal and migration*.30
13 §-5. *Philopatry and site fidelity*.....32
14 §-6. *Philopatry and dispersal in oceanic systems*33
15 **Antarctica and the Southern Ocean.....35**
16 §-7. *Oceanography and geography*.....35
17 §-8. *Climate variability in the Southern Ocean*.....40
18 §-9. *Climatic history of the Southern Ocean*.....42
19 §-10. *Biogeography of the Southern Ocean*.48
20 **Species' responses to anthropogenic climate change in the Southern Ocean.50**
21 §-11. *Climate is changing*.....50
22 §-12. *Particularities of climate change in the polar regions*.52
23 §-13. *Species' responses to climate change*.....55
24 §-14. *Micro-evolution and phenotypic plasticity*.55
25 §-15. *Habitat displacement and range shift*.....57
26 **Aims of this work.....60**
27 §-16. *The scales of climate change*.....60
28 §-17. *Understanding the scales of organisation in seabirds*.....62
29
30 **Chapter 2. Concepts and methods65**
31 **A brief introduction to Seabirds.65**
32 §-18. *Seabirds, and Penguins in particular, as sentinels for polar ecosystems*.....65
33 §-19. *Philopatry and dispersal amongst seabirds*.68
34 §-20. *Taxonomy of Penguins, and their relationship to other seabirds*.71
35 §-21. *Origin, evolution and radiation of sphenisciforms*.....73
36 §-22. *Aptenodytes penguins in the historical period*.....75
37 **Study sites and species.....77**
38 §-23. *The Emperor penguin*.78
39 §-24. *The Pointe Géologie colony and its regional and continental ties*.79
40 §-25. *The King penguin*.....80
41 §-26. *The Baie du Marin colony and long-term monitoring design*.81
42 §-27. *The Peruvian diving-petrel*.....83

1	Laboratory methods	84
2	§-28. <i>Utility of classical markers</i>	84
3	§-29. <i>The RADseq approach</i>	85
4	§-30. <i>Whole-genome resequencing</i>	86
5	From DNA libraries to polymorphism data	87
6	§-31. <i>Particularity of RAD data for analysis</i>	87
7	§-32. <i>The Stacks pipeline</i>	89
8	§-33. <i>SNP-calling approach</i>	90
9	§-34. <i>Genotype-free approach</i>	92
10	§-35. <i>Exploring the duplicate bias</i>	93
11	§-36. <i>The duplicate bias in our data</i>	96
12	§-37. <i>Phasing of the WGS data</i>	97
13	§-38. <i>A point on data management</i>	98
14	Genetic analysis framework	99
15	§-39. <i>Populations, random mating and neutral drift</i>	99
16	§-40. <i>The Coalescent theory</i>	101
17	§-41. <i>Coalescence, population size and migration</i>	103
18	§-42. <i>The Ancestral Recombination Graph</i>	107
19	§-43. <i>What is a summary statistic?</i>	110
20	§-44. <i>The Extended Bayesian Skyline Plot approach</i>	113
21	§-45. <i>The Allele Frequency Spectrum approach</i>	117
22	§-46. <i>The Stairway plot approach</i>	121
23	§-47. <i>Sequentially Markovian Coalescent approaches</i>	122
24	§-48. <i>Rate of evolution</i>	124
25	Habitat and climate analysis	127
26	§-49. <i>Habitat and niche</i>	127
27	§-50. <i>Climate projections in the Coupled Model Intercomparison Project</i>	130
28	§-51. <i>The Representative Concentration Pathways</i>	132
29	§-52. <i>Palaeoclimate experiments</i>	133
30		
31	Chapter 3. Fine-scale colony structure	135
32	Context	135
33	§-53. <i>Local structure and random sampling in population genetics</i>	135
34	§-54. <i>Impact of local habitat on populations</i>	136
35	§-55. <i>Aims of this chapter</i>	137
36	Abstract	137
37	Introduction	138
38	§-56. <i>Coloniality and population structure</i>	138
39	§-57. <i>The King penguin as a model</i>	139
40	§-58. <i>Expectations and hypotheses</i>	140
41	Materials and Methods	141
42	§-59. <i>Permits and ethics statement</i>	141
43	§-60. <i>Study site and sampling</i>	142
44	§-61. <i>DNA extraction, PCR and genotyping</i>	143
45	§-62. <i>Analysis of microsatellite data</i>	145
46	§-63. <i>Spatial autocorrelation analyses</i>	145
47	§-64. <i>Inbreeding and relatedness</i>	146

1	§-65. Pairwise cluster analysis.	147
2	§-66. Site-quality descriptors analysis.	147
3	Results	149
4	§-67. Microsatellite data.	149
5	§-68. Spatial analysis.	151
6	Discussion	156
7	§-69. A near-panmictic population.	156
8	§-70. Heterogeneity in inbreeding probability.	157
9	§-71. Pairwise cluster differentiation.	157
10	§-72. Heterogeneity in breeding site quality.	158
11	§-73. Effects of demography and life-history.	159
12	§-74. Heterogeneity as a mixing mechanism.	161
13	Acknowledgements	162
14		
15		
16	Chapter 4. The King synnome	163
17	Context	163
18	§-76. From local to global gene flow.....	163
19	§-77. Understanding the dynamics of contemporary structures.....	164
20	§-78. Integrating genetics with climate models.....	165
21	Abstract	166
22	Results	166
23	§-79. Expected responses to climate change.	166
24	§-80. Habitat constraints for the King penguin.	167
25	§-81. A Synnome species.	168
26	§-82. Palaeodemography of the Aptenodytes penguins.	169
27	§-83. 21st century range shifts.	173
28	§-84. Uncertainties and perspectives.	174
29	Supporting information	177
30	S-0: Supplementary methods: from sample collection to SNP typing.	
31	§-85. Sample collection and DNA extraction.	177
32	§-86. Genome-wide Single Nucleotide Polymorphism (SNP) typing.....	178
33	§-87. Sequence alignment and genotyping.....	179
34	S-1: Analysis of genetic data.	
35	§-88. Ancestral state reconstruction.	180
36	§-89. Summary statistics.....	181
37	§-90. Clustering analysis.....	181
38	§-91. Principal component analysis.	182
39	§-92. Analysis of molecular variance.	183
40	§-93. Pairwise Hamming distance network.....	184
41	§-94. HVR comparison with Macquarie.....	184
42	§-95. Demographic reconstructions: the Stairway plot method.	186
43	§-96. Demographic reconstructions: the Extended Bayesian Skyline Plot (EBSP) method....	187
44	§-97. Demographic reconstructions: the Pairwise Sequentially Markovian Coalescent.	190
45	§-98. Reconstruction validation through simulation.	192
46	S-2: Field observation data.	

1	§-99. Capture-mark-recapture experiments.	196
2	S-3: Palaeoclimate of the Southern Ocean.	
3	§-100. Definition and constraints of the Antarctic Polar Front.	197
4	§-101. Current state of knowledge.	198
5	S-4: Atmosphere-Ocean General Circulation Models (AOGCMs)	
6	§-102. AOGCMs choice and multi-model ensemble approach.	201
7	§-103. Sea Surface Temperature (SST).	202
8	§-104. Winter sea-ice concentration (SIC).	204
9	§-105. Uncertainties assessment.	205
10	Acknowledgements	209
11		
12	Chapter 5. The Emperor synnome	211
13	Context	211
14	§-106. Can a species be studied from a single colony?.....	211
15	§-107. Linking generation-scale and coalescent-scale demography.....	212
16	Abstract	213
17	Results	214
18	§-108. Studying demography in the context of climate change.	214
19	§-109. Instability of Emperor penguin colonies.	215
20	§-110. Study design.	216
21	§-111. A panmictic species.	217
22	§-112. Demographic reconstruction.	218
23	§-113. Continental-scale migration.	220
24	§-114. Consequences of high dispersal.	221
25	§-115. Future prospects.	221
26	Methods Summary	222
27	Modelling the changes in Emperor penguin habitat.	223
28	§-116. Aim of this study.....	223
29	§-117. How to model the Emperor's habitat?.....	223
30	§-118. Sea ice in the CMIP5 experiments.	225
31	§-119. Model and variable selection.	226
32	§-120. Is coastal sea ice really changing?.....	228
33	§-121. Uncertainties in the CMIP5 models.	232
34	§-122. Can we predict Emperor penguin demography based on sea ice?.....	235
35	Supporting information	236
36	S-0: Main concepts in use	
37	§-123. Choice and definition of a « synnome ».	236
38	§-124. Dispersal and migration.	236
39	S-1: Supplementary methods and analysis	
40	§-125. Sample collection and DNA extraction.	237
41	§-126. Genome-wide Single Nucleotide Polymorphism (SNP) typing.	238
42	§-127. Sequence alignment and genotyping.	239
43	§-128. RAD data description.	241
44	§-129. Pairwise FST, AMOVA, and summary statistics.	242
45	§-130. Identity-by-state (IBS) and identity-by-descent (IBD).	244
46	§-131. Principal Component Analysis (PCA).	245

1	§-132. Population splits and migration topology.....	246
2	§-133. Clustering analysis.....	247
3	§-134. Coalescent-based analysis: Migrate-n.	247
4	§-135. fastsimcoal2 analysis.	250
5	§-136. Coalescent-based analysis: BEAST2.	252
6	§-137. mtDNA analysis.	253
7		
8	Chapter 6. Empirical evidence of heterogeneous dispersal	259
9	Context.....	259
10	§-138. How do coalescent-scale migration translate at the ecological scale?.....	259
11	Abstract	260
12	Introduction	261
13	Materials and Methods	263
14	§-139. Animal ethics.....	263
15	§-140. Aerial photographs.....	264
16	Results	265
17	§-141. Habitat and Chick Census.....	265
18	§-142. Extrapolation to Total Colony Size.	267
19	Discussion.....	268
20	§-143. Strength and Shortcomings of the Remote-sensing Technique.....	268
21	§-144. Influence of the Breeding Location at Fine-spatial Scale.	271
22	§-145. Status of the Emperor Penguins Census.	272
23	Acknowledgements	274
24		
25	Chapter 7. Unexpected philopatry in an insular seabird, the Peruvian diving-petrel.....	275
26	Context.....	275
27	§-146. The Peruvian diving-petrel as a reference species.	275
28	§-147. The Humboldt Current System.....	276
29	Abstract	277
30	Introduction	278
31	§-148. Seabirds as terrestrial animals.	278
32	§-149. Recent history of the Peruvian diving-petrel.....	279
33	§-150. Present-day status of the species.....	280
34	§-151. Panmixia in the Humboldt Current System.	281
35	§-152. Aims of the study.	282
36	Results	282
37	§-153. Sequencing data.	282
38	§-154. Population structure.	283
39	§-155. Isolation and migration analysis.	285
40	Discussion.....	286
41	§-156. A highly structured population.	286
42	§-157. Population size and gene flow.....	287
43	§-158. Evolution of philopatry.	288
44	§-159. Implications for conservation.....	289
45	Material and Methods.....	290
46	§-160. Sample collection and DNA extraction.	290
47	§-161. Sampling and DNA extraction.....	290

1	§-162. Mitochondrial marker sequencing.	291
2	§-163. RAD library preparation and SNP typing.	291
3	§-164. Analysis of population structure.	293
4	§-165. Model-based estimation of population history.	293
5	Acknowledgements	296
6		
7	Chapter 8. Discussion	297
8	Population structure in colonial seabirds	297
9	§-166. <i>Aptenodytes penguins</i> are synonyme species.	297
10	§-167. Genetic landscapes in the Southern Ocean.....	299
11	§-168. Synonymes and speciation.....	304
12	Seabirds on the evolutionary scale.	305
13	§-169. Inferring demography in synonyme species.....	305
14	§-170. Pleistocene history of the <i>Aptenodytes penguins</i>	309
15	Seabird response to Anthropocene environmental change.	312
16	§-171. Seabird response to climate change.....	312
17	§-172. Seabird response to environmental change.	314
18	Further work on the penguin synonymes.	315
19	§-173. A final examination of penguin population structures.	315
20	§-174. Extending palæodemographic reconstructions.	316
21	§-175. From neutral to adaptive genomics.	318
22	§-176. Synonymes and gene swamping.	320
23	Conclusion.....	323
24		
25	Annexes	327
26	King penguin demography since the last glaciation inferred from genome-wide data	327
27	Abstract	327
28	Introduction	328
29	Results	332
30	Discussion.....	337
31	§-177. Using RAD sequencing data in a coalescent-based framework.....	337
32	§-178. King penguin population history during the Last Glacial Maximum.....	340
33	Material and methods	342
34	§-179. Sampling, DNA extraction and quality assessment.....	342
35	§-180. Restriction site-associated DNA sequencing.	342
36	§-181. Genome-wide demographic analysis and calibration.	343
37	§-182. Mitochondrial DNA control region analysis.	344
38	§-183. Acknowledgements.....	346
39	Supporting information	347
40	§-184. Sampling, DNA extraction, quality assessment.	347
41	§-185. MtDNA marker analysis.	347
42	§-186. RAD sequencing and genome-wide demographic inference.	348
43	RAD-seq bench protocol optimisation.	353
44	Updated RAD-sequencing protocol	357
45	References.....	366
46	Introduction (traduction française).....	408

1 **Index of tables**

2 Table 1: Microsatellite analysis: multiplexing parameters and summary statistics.144
3 Table 2: Pairwise cluster comparison of inbreeding and relatedness distributions153
4 Table 3: Ensemble members used in habitat predictions.201
5 Table 4: Correlation of observed and modelled SST in the Southern Ocean.203
6 Table 5: Ensemble members used in sea ice modelling.227
7 Table 6: Pairwise Fst, according to Reich's estimator.244
8 Table 7: Migration rates and sizes as estimated from the joint allele frequency spectrum.252
9 Table 8: 2012 census of the Mertz emperor penguin colonies.268

10 **Index of figures**

11 Figure 1: Demography in the Ecological and the Genetic paradigms.24
12 Figure 2: Structure of water masses in the Southern Ocean.37
13 Figure 3: Primary productivity concentration on the Antarctic Polar Front.38
14 Figure 4: The formation of sea ice in coastal polynyas.42
15 Figure 5: Sea Surface Temperature changes in the Southern Ocean.45
16 Figure 6: The Food web of the Southern Ocean.53
17 Figure 7: Distribution of the two Aptenodytes species.77
18 Figure 8: The RAD sequencing approach.87
19 Figure 9: Stochasticity and demography in the coalescence process.102
20 Figure 10: Coalescence and population size.107
21 Figure 11: Influence of demographic processes on the joint AFS.120
22 Figure 12: The Pairwise Sequentially Markovian Coalescent model.122
23 Figure 13: Sampling distribution.143
24 Figure 14: Observed individual inbreeding distribution150
25 Figure 15: Individual inbreeding and nearest-neighbours-relatedness151
26 Figure 16: Spatial correlation and Intrinsic clustering tests do not support strong structure.152
27 Figure 17: Ecological descriptors of breeding-site quality155
28 Figure 18: First axis of principal component analysis as a summary habitat-quality index.156
29 Figure 19: Penguin paleodemography in response to Quaternary climate change.170
30 Figure 20: Past and future breeding range of the King penguin.172
31 Figure 21: Projected foraging distance under three RCP scenarios.175
32 Figure 22: Sampling design.177
33 Figure 23: Principal component analysis.182
34 Figure 24: Neighbour-net.183
35 Figure 25: Robustness of the Stairway plot method.185
36 Figure 26: Extended Bayesian Skyline Plot.190
37 Figure 27: Pairwise Sequentially Markovian Coalescent.192
38 Figure 28: Validation of the demographic reconstructions through simulation.195
39 Figure 29: (NEXT PAGE) Foraging distance from single models.206
40 Figure 30: Proportion of models predicting extinction of King penguin colonies.209
41 Figure 31: Drift and mixing in the Emperor penguin.217
42 Figure 32: Demography is a matter of scales.219
43 Figure 33: Changes in sea ice configuration along coastal Antarctica.229
44 Figure 34: Uncertainty in the representation of sea ice in the CMIP5 models.232
45 Figure 35: Consensus SNP calling.241

1	Figure 36: The imprint of geography on the Emperor synnome.....	243
2	Figure 37: Pairwise genetic distance matrix.....	245
3	Figure 38: Inference of population splits.....	247
4	Figure 39: Haplotype network for mitochondrial sequences.	256
5	Figure 40: Western Emperor penguin colony.....	266
6	Figure 41: Eastern Emperor penguin colony.....	267
7	Figure 42: Satellite images of the Mertz glacier.....	271
8	Figure 43: Full distribution of <i>Pelecanoides garnotii</i>	280
9	Figure 44: Consistent genetic structure within the Chilean range.	284
10	Figure 45: Winter sea ice extent.....	330
11	Figure 46: Demographic inference based on the AFS.....	334
12	Figure 47: Demographic reconstructions of the Crozet king penguin colony.....	335
13	Figure 48: Past demographic trend of the king penguin colony of 'La Baie du Marin'	339
14	Figure 49: Demographic reconstructions of the Crozet king penguin colony.....	345
15	Figure 50: Bayesian Skyride Plot inferred from mitochondrial Control Region.....	346

1 *Articles and communications.*

2 *The following articles are included in this thesis:*

3 - Cristofari R, Trucchi E, Whittington JD, Vigetta S, Gachot-Neveu H, Stenseth NC, Le Maho
4 Y, Le Bohec C (2015) **Spatial heterogeneity as a genetic mixing mechanism in highly**
5 **philopatric colonial seabirds.** PloS one 10: e0117981.

6 - Cristofari R, Liu X, Bonadonna F, Cherel Y, Le Maho Y, Pistorius P, Raybaud V, Stenseth NC,
7 Le Bohec C & Trucchi E, **Stepping-stone range shifts in response to climate change in the**
8 **Southern Ocean.** (*in prep.*)

9 - Cristofari R, Bertorelle G, Ancel A, Benazzo A, Le Maho Y, Ponganis PJ, Stenseth NC, Trathan
10 PT, Whittington JD, Zanetti E, Zitterbart DP, Le Bohec C & Trucchi E, **Full circumpolar mi-**
11 **gration ensures evolutionary unity in the Emperor penguin.** (*Nature Communications - in prep*)

12 - Ancel A & Cristofari R, Fretwell PT, Trathan PN, Wienecke B, Boureau M, Morinay J, Blanc
13 S, Le Maho Y, Le Bohec C (2014) **Emperors in Hiding: When Ice-Breakers and Satellites Com-**
14 **plement Each Other in Antarctic Exploration.** PloS one 9:e100404

15 - Cristofari R, Fernandez-Zamora F, Gouin N, Le Bohec C, Plaza P, Trucchi E, Zavalaga C, Luna
16 Jorquera G, **Unexpected population isolation in a critically endangered insular seabird, the Pe-**
17 **ruvian Diving Petrel (*Pelecanoides garnotii*).** (*in prep.*)

18 *The following communications also served as a basis for this work:*

19 - Le Bohec C, Whittington JD, Ancel A, Chatelain N, Cornet C, Courtecuisse J, Crenner F,
20 Cristofari R, Marpaux S, Allemand D & Le Maho Y (2014, August) **Predict changes in polar**
21 **ecosystems : biological adaptation and technological innovation.** Oral presentation at the
22 XXXIIIth SCAR Open Science Conference, Auckland, New Zealand.

- 1 - Cornet C, Amélineau F, Babel D, Boureau M, Courtecuisse J, Cristofari R, Descamps S,
2 Marpaux S, Morinay J, Whittington JD, Le Maho Y. & Le Bohec C. (2014, July). **Personality
3 and environmental heterogeneity in the Adélie penguin**. Oral presentation at the 15th Confer-
4 ence of the International Society for Behavioural Ecology, New York, USA.
- 5 - Cornet C, Amélineau F, Babel D, Boureau M, Courtecuisse J, Cristofari R, Descamps S,
6 Marpaux S, Morinay J, Saraux C, Whittington JD, Le Maho Y. & Le Bohec C. (2013, Septem-
7 ber). **The adaptive capacities of Adélie penguins to face environmental variability : the role of
8 heterogeneity within populations**. Keynote presentation at the 8th International Penguin Con-
9 ference, Bristol, UK.
- 10 - Cristofari R, Trucchi E, Whittington JD, Vigetta S, Gachot-Neveu H, Stenseth NC & Le Bo-
11 hec C (2015) **Spatial heterogeneity as a genetic mixing mechanism in highly philopatric colo-
12 nial seabirds**. Poster presented at the XIth SCAR Biology Symposium : « Life in Antarctica :
13 Boundaries and Gradients in a Changing Environment », Barcelona, Spain.
- 14 - Cornet C, Amelineau F, Babel D, Boureau M, Courtecuisse J, Cristofari R, Descamps S,
15 Marpaux S, Morinay J, Whittington JD, Le Maho Y. & Le Bohec C. (2013, July). **Personality
16 and its effect on fitness in the Adélie penguin**. Poster presented at the XIth SCAR Biology Sym-
17 posium : « Life in Antarctica : Boundaries and Gradients in a Changing Environment », Barcelo-
18 na, Spain.
- 19 - Le Bohec C, Cornet C, Cristofari R, Whittington JD, Courtecuisse J, Chatelain N, Crenner F,
20 Allemand D & Le Maho Y. (2013, July). **Adaptive strategies and population trends of penguins
21 to predict changes in polar marine ecosystems**. Oral presentation at the XIth SCAR Biology
22 Symposium : « Life in Antarctica : Boundaries and Gradients in a Changing Environment »,
23 Barcelona, Spain.

1 *Foreword*

2 The work we present here lies in-between two poles of biological sciences - ecology and evolu-
3 tion. Increasingly well explored from a theoretical and laboratory perspective, the practical inter-
4 face of these two disciplines, when applied to wild species and natural ecosystems, is at the core
5 of intense contemporary investigation: for this reason, as will appear to the reader, the methods
6 we use matter nearly as much as the conclusions we reach. To that aim, we will examine the par-
7 ticular stories of three seabirds of the Southern Seas, the Emperor penguin (*Aptenodytes forsteri*),
8 the King penguin (*Aptenodytes patagonicus*), and the Peruvian diving-petrel (*Pelecanoides garnotii*),
9 each of them confronted to environmental change in its own manner. Our primary line of inves-
10 tigation will be the way genetic data - an evolutionary exploration tool by excellence - can be in-
11 tegrated in the framework of ecological models, and in particular in numeric representations of
12 climate.

13 This thesis is divided into eight main sections. A first introductory chapter will present the main
14 concepts around which our work revolves, as well as its climatic and oceanographic framework. A
15 second chapter will give an overview the methodological concepts that will be deployed in this
16 particular study. The third and fourth chapters will focus on the King penguin, and its interac-
17 tions with its habitat both at local, and global scale. The fifth and sixth chapters will repeat this
18 exercise for the Emperor penguin. The seventh chapter, that presents the particular case of the lit-
19 tle-known Peruvian diving-petrel, will showcase how these methods can be used as an exploratory
20 tools in remote species where field data is scarce. Finally, an eighth and last chapter will sum-

1 marise the main findings of this work, and propose some conclusions, both for our three focal
2 species, and for the development of more accurate investigation methods.

3 As already appears from its organisation, the question of scales - both spatial, and temporal - will
4 be a structuring one throughout this manuscript. The complexity of temporal scales in speciation
5 processes is still at the core of evolutionary investigations (see *e.g.* Gould & Eldredge 1972) - and
6 this complexity only increases when considering ecological processes together with evolutionary
7 ones. When focusing on a metazoan organism (a simple case compared to the bewildering organ-
8 isation of plants), a point of relatively atomic singularity is reached with the *individual*¹. Above
9 that scale, the longer-lasting populations and species are increasingly complex ensembles obeying
10 to stochastic processes; while below that scale, the rapid and minute scales of physiology bring us
11 into yet another stochastic realm of populations, of cells or molecules this time: thus, going be-
12 low or above the individual indistinctly leads us back from the most singular towards an increas-
13 ing form of universality.

14 The nested temporal scales of *individuals*, *populations* and *species* also need to be thought of in the
15 complex temporality of their environment. Metazoan life is directly characterised by the instanta-
16 neous interface between the individual, and the immediate conditions of its environment. At an
17 instantaneous scale, this interface is highly variable, both externally (through the presence or ab-
18 sence of food, conspecifics, competitors, or the state of the weather, *etc*) and internally (*i.e.*
19 through the physiological state of the individual), and leads to a wide variation of « *instantaneous*
20 *fitness* » - or, more simply, immediate well-being. Yet this stochastic variation is averaged through-
21 out the lifespan of the individual to a range of external and internal conditions, that results in a
22 individual general *fitness*². Further on, the large-scale fluctuations in average weather, or *climate*,

1. The « *un-divisible* », by etymology.

2. Fitness is first defined by Darwin 1859, together with the concept of *natural selection*, as the idea that « *individuals hav-
ing any advantage, however slight, over others, [...] have the best chance of surviving and of procreating their kind. On the other hand,
we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favourable variations,
and the destruction of injurious variations, I call Natural Selection, or the Survival of the Fittest* » - fitness being here loosely conceived
as the advantage conferred to an individual by its particular characteristics. As further noted by Barker 2008, « *fitness may refer to a*

1 together with the associated changes in biological communities¹, impose long-term patterns on
2 the average relationship of large numbers of individuals to their changing environment - and, ul-
3 timately, *evolution*. Finally, in-between the immediate contact of the *individual* and the *instant*,
4 and the slow change of the relationship of the *species* and the state of the world, lie the multiple
5 facets of intermediate variations. The oscillating climatic phenomena that operate on a few years
6 scale, the centennial variations of the Southern Annular Mode, or the slower pulse of Milanković
7 cycles and Pleistocene glaciations, all blur any direct, linear extension of short-term phenomena
8 to long-time scales. This fractal conception of the scales on which chronological processes operate
9 is mirrored in space, through the local heterogeneity both of individuals and of environment: lo-
10 cal adaptation or local extinction often occur at cross-current with the general stream of
11 evolution.

12 The theory of population genetics seems to take all its sense at the light of this complex structure
13 of time and space - with the curious meaning that *time* and *space* take there. A gene (this very
14 material, real and transient fragment of a DNA molecule bounded both in space and in time) is
15 progressively *abstracted* in the population genetics framework until it becomes a continuous time-
16 object - a *sequence*. In this framework, the diversity of DNA molecules is relieved both from its
17 synchronic and diachronic boundaries, as formalised through the approximations of the *continu-*
18 *ous-time coalescent* and the *diffusion equations of drift* (see later on in §39 p. 99 and §40 p. 101).
19 The resulting generalised object now transforms slowly and continuously over the evolution of
20 the species, in an essentially Braudelian *longue durée* conception². Yet this approximation requires
21 the sacrifice of the most obviously discrete aspects of metazoan life: the individual and the
22 generation.

genotype, an individual, a population or a species», but may be broadly defined as « *the ability of organisms to pass on the genes they carry*».

1. By «biological community», or *biocœnosis*, we refer to the ensemble of species that share a single habitat and interact within it (see Möbius 1877).

2. See Braudel 1958, and §16 p. 60.

1 In parallel, however, a fundamental postulate of the modern synthesis is that evolution operates
2 precisely at the level of the *individual* (Mayr & Provine 1998; Seehausen *et al.* 2014). At this lev-
3 el, genetic material is once again highly bounded: within the space of the individual's body, it is
4 classically (although simplistically) considered *one*, and collectively subjected to fitness selection
5 at the interface of the body and the environment. The sanction of this fitness selection on the
6 space of the body is the shift of its boundaries in time: the frequency of *beginning* of a new organ-
7 ism carrying these genes, or *breeding*, and the delayed *demise* of that ensemble (or *death*): more
8 beginnings and later demises are the way evolutionary success is sanctioned. Breeding and death
9 have a position in time, yet they are also *spatial* boundaries, as they describe the beginning and
10 end of the *body*, that *embodiment* of the abstract sequence along its continuous tree-like
11 evolution.

12 And so, a *mise en abyme* occurs for the gene too: the processes that apply on the individual (a col-
13 lection of identical DNA molecules) as *fitness selection* apply on the family group (an ensemble of
14 highly similar molecules) as *kinship selection*, and on the species as *evolution*. And we are left won-
15 dering whether the variations in fitness that we document at the level of the individual (the varia-
16 tion in individual «quality» and life-history choices, and their direct consequences on that indi-
17 vidual's life) are of the same nature as the broad trends in the adaptation of the species to an
18 environment and a state of the climate, or whether the fitness of an individual is an essentially
19 different concept from the fitness of a population, the one involving a stochastic, particular asso-
20 ciation of alleles, and the other a collective distribution of allele frequencies and expected fitness
21 selected through repeated, but inaccurate, individual trajectories and accidents.

22 *In fine*, this amounts to asking how prevalent the individual should be in the study of the collec-
23 tive, and whether there is any sense at all in disregarding, even for theoretical purposes, that intu-
24 itively central figure when considering the fate of species. Although here purely methodological,
25 this question ultimately converges towards ethical perspectives: indeed, the relative importance of
26 the individual and the species is a difficult intuition, or decision, in our relationship to nature as

1 a whole. Conservation biology is mostly concerned with species, at the cost of disregarding indi-
2 vidual trajectories: yet it may sometimes appear that this conception fails to capture the true
3 complexity of our surroundings, by reducing to a handful of taxonomic categories what is really a
4 rich spectrum of individual consciousnesses¹. And paradoxically, despite its near-complete disap-
5 pearance throughout the present manuscript, I finish with a clear intuition that only the indi-
6 vidual makes absolute sense, and that the ultimate goal of my study was to better understand
7 what exactly is the *experience* of the singular seabird who, against larger population trajectories
8 and environmental changes, « *ever stands forth his own inexorable self* »². However, these considera-
9 tions must remain the salt of this work, and should not interfere with the detachment that is re-
10 quired by all sound scientific endeavours.

11 Thumbing through this manuscript, it will be evident that the diversity of methods is beyond a
12 single person's work. The contributions of Dr Céline Le Bohec on the ecological side, and of Dr
13 Emiliano Trucchi on the evolutionary side, have been determinant at every single step of this
14 study. No less important was the expertise of Dr Giorgio Bertorelle, Dr Xiaoming Liu, and Dr
15 Virginie Raybaud, in giving shape and soundness to the different parts of this manuscript: I want
16 to thank them here once again.

1. « *Yet what is consciousness? You can imagine that I will not define something so concrete, to constantly present to the experience of each of us. But without giving of consciousness a definition that would be less clear than consciousness itself, I can characterise it through its most prominent feature: consciousness means memory* » (Bergson 1911).

2. To quote from Melville's *Moby Dick*.

1 Chapter 1: Introduction

2 *Populations, philopatry and dispersal*

3 *§-1 Species: type and repetition.* The systematic description of species can be seen as an extreme
4 mediation of the *individual* and the *collective*. A single sample, the *holotype*, is taken as the explic-
5 it reference for bestowing the species' name - and the rest is essentially a *repetition*, in time and in
6 space, of that individual type. That repetition, however, cannot be perfect (Kierkegaard 1843):
7 variation, or *diversity*, plays a central role, both in time and in space (Mayr 1982). Intra-specific
8 phenotypic variation only became the focus of biological sciences as «*individual differences*» with
9 the works of Darwin, who famously understood them to «*afford materials for natural selection to*
10 *act on and accumulate*» (Darwin 1859): the genetic basis of this diversity is now at the centre of
11 the modern evolutionary synthesis (Mayr & Provine 1998; Seehausen *et al.* 2014). While not
12 serving any pre-defined purpose, variation may *a posteriori* be organised along a spectrum ranging
13 from *adaptive* to *neutral* diversity (Kimura 1983; Wagner 2008). At the molecular level, adaptive
14 variation can be defined as the appearance of mutations with a beneficial phenotypic effect, that
15 are positively selected in the population (Mayr 1963), while neutral variation is the appearance of
16 mutations offering virtually no hold to natural selection (Kimura 1983). Without taking position
17 as to the relative importance of neutral and adaptive variation in molecular evolution (a still de-
18 bated topic, see *e.g.* Wagner 2008), the present work will mostly focus on *neutral variation*, as an
19 insight into non-molecular ecological processes.

Introduction - §1

1 The organisation of molecular variation in space and time is the foundation of population genet-
2 ics (Wright 1978; Kingman 2000), which can be defined as the study of genetic diversity within
3 the conventional spatio-temporal bounds of a species - those «*merely artificial combinations made*
4 *for convenience*» in the midst of «*intermediate gradations*» (Darwin 1859). The extent of this di-
5 versity and its organisation reflect basic characteristics of the species: genetic diversity itself has
6 been linked to a variety of parameters, ranging from organismal complexity (Lynch & Conery
7 2003) to population size and history (Amos & Harwood 1998) or life-history strategies
8 (Romiguier *et al.* 2014), while its organisation in space and time is largely a consequence of its
9 demographic and migratory history (a point introduced in §41 p. 103). Therefore, studying
10 species as a population-scaled system allows us to capture not only their intrinsic taxonomic di-
11 versity, but also the variety their ecological and evolutionary dynamics.

12 A most remarkable feature of the population genetic approach to within-species diversity, that
13 has important methodological consequences, is (as far as metazoans are concerned) its necessarily
14 ambivalent conception of the *individual*. Population genetics are, in a sense, pervaded with an
15 overwhelming notion of singularity: as Ernst Mayr put it, «*wherever we look, we find uniqueness,*
16 *and uniqueness spells diversity*» (Mayr 1982). The individual, as a singular (and likely unique)
17 combination of alleles, is the only possible form upon which fitness selection may operate, and
18 thus the only effective shape of genetic material - in that sense, a population, or a gene pool, is
19 nothing else than an abstract representation of what is really only an aggregation of singular indi-
20 viduals. Although the concepts of species or populations (see §2 p. 25) are essential tools for un-
21 derstanding behavioural, demographic or evolutionary processes, the only observable, uninter-
22 preted atomic unit of metazoan life is the individual. Hence, any empirical study that relies on
23 the sampling of groups or populations really relies, in fact, on the sampling of a collection of
24 unique individuals, the assignment to a group being a posterior decision of the observer (which
25 poses very concrete questions of sampling design, see *Fine-scale colony structure*, p. 135).

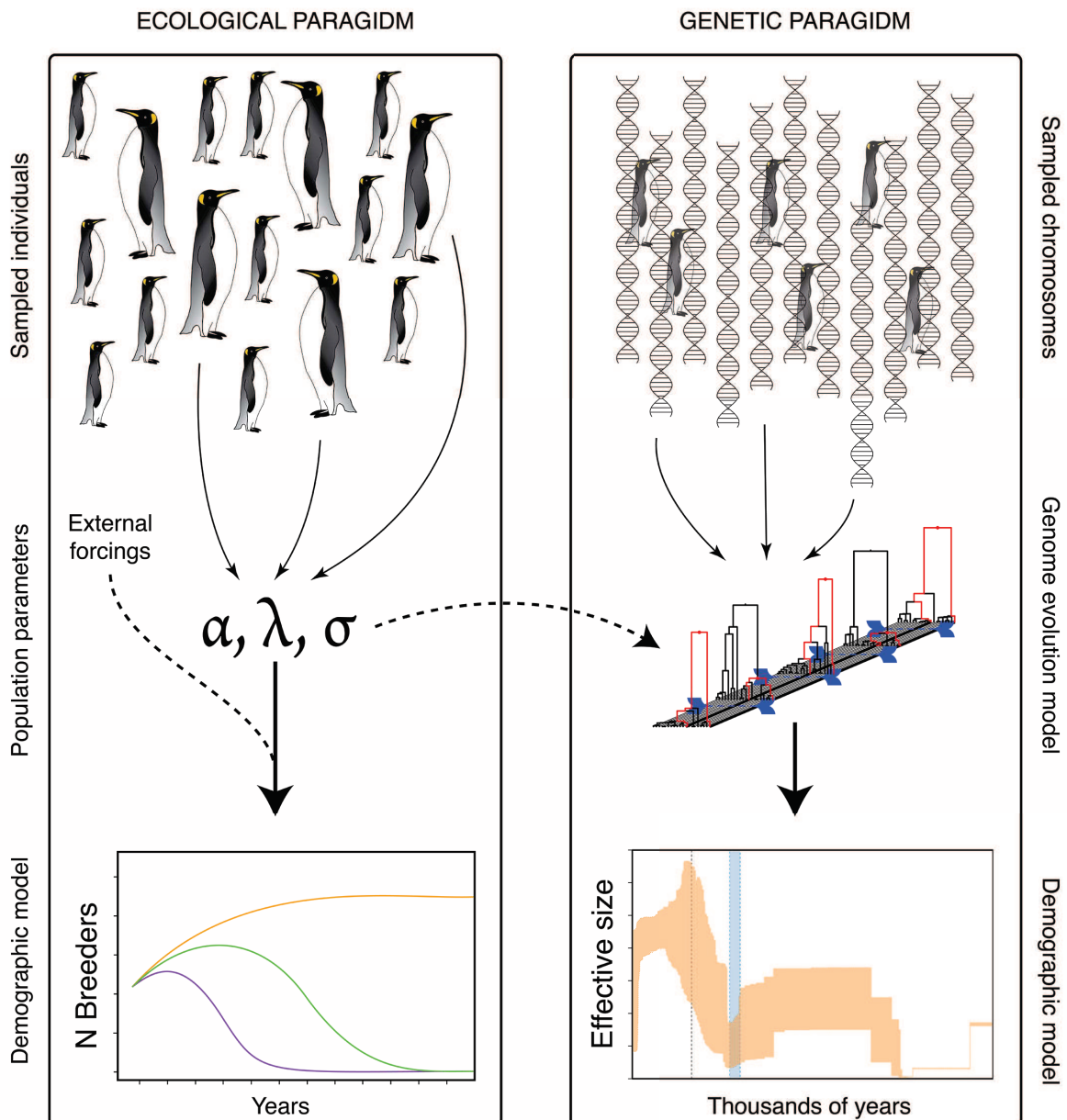
Introduction - §1

1 And yet the *individual* is dissolved either *vertically*, or *transversally*, throughout the theoretical
2 framework of population genetics. The central notion of *allele frequency* (see §45 p. 117), for
3 example, exists only through the transversal analysis (*i.e.* at a single point of time) of a large num-
4 ber of chromosomes at a single genomic position, without retaining any information as to the
5 particular associations of alleles within individuals - it can thus be conceived as a removal of the
6 *synchronic*¹ boundaries between individuals - an idea that reaches its mathematical consecration in
7 the diffusion equations approximating genetic drift (see §39 p. 99). The whole structure of the
8 *coalescent* (see §40 p. 101 *sqq.*), on the other hand, relies on the vertical (*i.e.* cross-generation)
9 analysis of sequences, throughout which the advent and demise of the individual (*i.e.* the *genera-*
10 *tion*) is an accident, and a mere unit of time: thus, the coalescent concept itself involves the disso-
11 lution of the *diachronic* separation between individuals - a conception that is formalised in the
12 continuous-time representation of the coalescent, that completely discards the very notion of
13 generation or individual (see §45 p. 117).

14 This paradoxical representation of singularity in the population genetic framework is in stark
15 contrast with the prevalent concept of the individual both in the systematic framework (where
16 singularity is reduced to dispersion around the holotype in morphospace), and in the ecological
17 framework - where diversity is either replaced by homogeneous population parameters such as
18 density (Bolnick *et al.* 2011), or on the contrary emphasised through attention to singular life-
19 history choices (*e.g.* Le Bohec 2008; Weimerskirch 2013). One of the challenges of the modern
20 evolutionary synthesis therefore remains the reconciliation of these different paradigms despite
21 widely divergent tenets. In particular, the difference between the concepts of *population size* or
22 *migration rate* in the ecological and the population genetic (henceforth «evolutionary») para-
23 digms (see §41 p. 103 and Fig. 1 p. 24) makes the integration of cross-disciplinary sources diffi-
24 cult (see *The Emperor synnome*, p. 211) - and so do the very different time scales on which these

1. *i.e.* boundaries between individuals occurring *at the same time*, or «physical» boundaries - as opposed to *diachronic* boundaries, that occur between *successive generations*.

1 different frameworks are based: in many respects, the population genetics framework occupies an
 2 intermediate position between systematics (that operates on the speciation time scale) and ecolo-
 3 gy (that focuses on the discrete lifespans of individuals). While the synthesis of these different
 4 modes is, of course, beyond the scope of this work, the questions addressed here, being at the
 5 crossroads of ecology and population genetics, will be best understood through this unavoidable
 6 ambivalence.



7 **Figure 1 | Demography in the Ecological and the Genetic paradigms.** In the ecological paradigm, a
 8 subset of individuals is used to infer population-wide demographic parameters, and these define a general population
 9 model, normally in units of number of breeders and at the decadal scale. In the genetic paradigm, a subset of

Introduction - §2

1 chromosomes is sampled, that integrates diachronic aspects of the population: this subset is used to fit a genome evolution model, that further defines a general demographic model, in units of *effective breeders* (see §41 p. 103) and usually at the scales of several thousand years.

4 *§-2 Population, deme, colony.* Just as the distribution of individual positions in the morphological, behavioural or genetic space defines a species' patterns of spatial and temporal diversity, the extent of the overlap between an individual's movement range on the one hand and the species' range on the other determines a spectrum of spatial structures, with homogeneous, *panmictic* ensembles on the one end, and highly heterogeneous and fragmented systems on the other. The particular architecture of a species integrates processes occurring on very different scales: and its description will change according to the framework in which analysis is conducted (Esler *et al.* 2006). The definition of a *population* - a central concept in a field such as population genomics - thus differs widely according to whether we place ourselves in a *genetic*, or an *ecological* paradigm (Waples & Gaggiotti 2006). In the ecological paradigm, the criteria retained are usually geography (individuals occurring *at the same place*, *e.g.* Krebs 2013), interactions between individuals (such as sharing food resources, *e.g.* Huffaker & Gutierrez 1984), or, importantly, demographic independence (*i.e.* migration between groups is not sufficient to force them into similar demographic trajectories, or influence extinction risk of one group, *e.g.* McElhany *et al.* 2000). Waples and Gaggiotti (Waples & Gaggiotti 2006) summarise the ecological definition as « *a group of individuals of the same species that co-occur in space and time and have an opportunity to interact with each other* » - a definition that relies on the examination of a species distribution, demography and behaviour at an observable timescale, *i.e.* in the order of magnitude of the observer's own lifetime.

23 In the genetic or «evolutionary» paradigm (Waples & Gaggiotti 2006), on the other hand, a population is rather defined as a breeding community, *i.e.* a group of individuals that have a higher probability of breeding with each other, than with an individual from another group (Hartl *et al.* 1988), and consequently as a group of individuals with correlated genotypes. The

Introduction - §2

1 quantitative aspect of the minimum correlation needed to achieve population status is debated:
2 but whenever a standard criterion is proposed, it relies on a comparison of the relative influences
3 of generic drift and migration in shaping a group's genetic diversity (see §39 p. 99 and §41 p.
4 103). For example, a threshold of one effective migrant per generation has been used in empirical
5 studies (Mills & Allendorf 1996; Vucetich & Waite 2000; Wang 2004): if two groups exchange
6 more than one effective migrant per generation, they are not considered separate. Higher thresh-
7 olds have been proposed - *e.g.* a maximum of 5, or 25 effective migrants per generation (see
8 Waples & Gaggiotti 2006 for a review). As these criteria are concerned with per-generation para-
9 meters, they can only apply when averaged over several generations - *i.e.* on a much longer time
10 frame than the ecological definition.

11 The relationship between the concepts of population in the ecological and the evolutionary para-
12 digms is not a trivial one. Although it would seem that behavioural and geographical isolation of
13 a group of individuals naturally leads to reduced gene flow and genetic isolation, it is not clear
14 how far the patterns observed on the decadal scale also apply over longer periods, as group cohe-
15 sion may be heterogeneous over time or space, and lead to the apparent paradox of behaviourally
16 very structured, yet genetically fully admixed species (*e.g.* Harlequin ducks, Esler *et al.* 2006, or
17 Wandering albatrosses, Milot *et al.* 2008). In order to clearly distinguish between these different
18 paradigms, in the present work, we will mainly use three related - but not equivalent - terms to
19 describe the architecture of species: a *deme*, a *population*, and a *colony*.

20 A *deme* is a fully genetic definition of a group of individuals. It may be summarised as «*the*
21 *largest area or collection of individuals where mating is (on average) random*» (Hamilton 2011). This
22 definition does not necessarily include a spatial component, as several groups of random-mating
23 individuals may occur in sympatry (*e.g.* Killer whales, Riesch *et al.* 2012). In particular, demes do
24 not need to have boundaries (neither geographical nor behavioural), and may be defined in a
25 «sliding window» fashion, *e.g.* in a classical isolation-by-distance system, where each individual
26 may be taken as the focal point of a group or «genetic neighbourhood» (Crawford 1984) within

Introduction - §2

1 which alleles are highly correlated, and which decays gradually with distance. In that case the
2 *deme* might be seen as the local approximation of a broader spatial autocorrelation system (Dou-
3 ble *et al.* 2005; Epperson 2005). A threshold of 5% probability of mating has been proposed to
4 define the limits of a *deme* (Hamilton 2011) - but this is obviously arbitrary, as the concept
5 makes most intuitive sense in discretely distributed species, with strong discontinuities in mating
6 probability. When defining a *deme*, we therefore assume that within the ensemble under scrutiny,
7 mutation and drift, as opposed to migration, are the major contributors to genetic diversity.

8 A *colony* - especially for seabirds, see §18 p. 65 - is the conceptual counterpart of a *deme*: it is a
9 completely geographical and behavioural concept, describing a spatial aggregation of breeding in-
10 dividuals over a period of time, and does not make any assumption as to the underlying genetic
11 structure or evolutionary processes. The term is widely used in a variety of organisms, ranging
12 from polyps (*e.g.* corals or siphonophores) to insects (*e.g.* ants, Giraud *et al.* 2002), each time
13 with a different set of characteristics (in particular as to functional integration, which is extreme
14 in siphonophores, see *e.g.* Dunn 2005, but minimal in corals, see *e.g.* Soong & Lang 1992).
15 Coloniality is also well-spread in terrestrial vertebrates, whether in mammals (such as sper-
16 mophiles, Armitage 1981; Johnson 1981), or in birds (Rolland *et al.* 1998). While some bird
17 species' colonial behaviour extends to all aspects of their life (*e.g.* the well organised European
18 starling *Sturnus vulgaris* colonies), most only exhibit colonial behaviour during reproduction
19 (Lack 1968). This is especially the case in seabirds (see §18 p. 65), that may spend most of the
20 year foraging alone or in small groups, out at sea, but gather in large colonies on land at the start
21 of the breeding season. Paradoxically, philopatry is high in almost all seabirds (see §5 p. 32): al-
22 though individuals roam far and wide during their foraging season, and have virtually no need
23 for the shore, they keep a strong inner link to a particular patch of land, and return there regular-
24 ly to breed. For some species with a limited foraging range, this may be the only available breed-
25 ing grounds in the area (*e.g.* for the Black guillemot *Cepphus grylle*, see Ewins 1986). Yet in the
26 widest-ranging seabird species, such as the great albatrosses (*Diomedea sp.*), or the *Aptenodytes*

Introduction - §3

1 penguins, colonial behaviour becomes a truly abstract conception: non-breeding foraging range
2 overlaps widely between breeding colonies (and may in reality extent all around the world, see
3 Croxall *et al.* 2005), while individuals will return to their birth-place with astounding regularity.
4 Individual learning, and horizontal, or «cultural» (*sensu* Whitehead & Rendell 2014) transfer of
5 information between birds breeding at the same place, may even lead to colony-specific idiosyn-
6 crasies, for example different foraging grounds for nearby colonies of the same species (Weimer-
7 skirch 2013). Thus, the *colony* can be seen as the most extreme representation of the population
8 concept in the ecological paradigm: a spatially and temporally stable aggregation of individuals,
9 that can be easily observed, defined and bounded (although not without conceptual choices, see
10 §41 p. 103), but without assumed genetic characteristics.

11 Our use of *population* is at the intersection of *deme* and *colony*. In this work, a population will be
12 defined as an ensemble that is bounded at the same time in the geographical, the demographical
13 and the evolutionary spaces - in other words, a colony, or group of colonies, that also happens to
14 be a deme. Rather than an anecdotal coincidence, speaking of a *population* implies that the
15 processes that can be observed at the ecological scale (*i.e.* geographical clustering of individuals,
16 use of common resources, or philopatric behaviour) are strong and stable enough through time to
17 have an influence on evolutionary mechanisms, so that both the ecological and the genetic para-
18 digms overlap. Thus, as opposed to a *deme*, our definition of population requires a *geographical*
19 component - and as opposed to a *colony*, it also requires an *evolutionary* criterion.

20 Two antagonistic forces thus interact in shaping the wide variety of possible species architectures:
21 *dispersal*, or the active or passive ability of individuals to move across their habitat, and *philopatry*,
22 or the tendency for individuals to stay, or return, to their birth place (when fully passive, this may
23 more accurately be called *inertia*). Taken together, these two forces may respectively be consid-
24 ered as the specific *fluidity* and *viscosity* of a species.

25 §-3 *What is a Synnome?* The particular architecture of the two *Aptenodytes* penguin species falls
26 short of any existing concept. As will be exposed in detail in the following chapters (especially

Introduction - §3

1 *The King synnome*, p. 163 and *The Emperor synnome*, p. 211), *Aptenodytes* penguins display a star-
2 tling tension between the ecological and the evolutionary definitions of a population. Colony
3 structure is brought to an extreme, with highly discrete aggregations of individuals that appear
4 fairly stable through time, at least at the centennial scale. Philopatry is high (see §19 p. 68), and
5 is thought to extend not only to the colony, but also regularly to the very nesting site within that
6 colony (see *Fine-scale colony structure*, p. 135 for details). Colonies have a high degree of demo-
7 graphic independence, as shown for example by the contrasting trajectories of the five rookeries
8 on Possession Island, in Crozet Archipelago (Delord *et al.* 2004). However, none of the genetic
9 criteria are met, and the long-term fluidity of the species appears to fully counteract any incipient
10 genetic structuring, with migration asserting itself as a major evolutionary force in these species.
11 Such a structure is distinct from a metapopulation system (see *e.g.* Hanski 1998), in which sub-
12 units are, to some extent, genetically separated - but it is also distinct from a panmictic system,
13 since the extremely clustered spatial distribution of individuals, and high philopatry, contradicts
14 the assumption of random mating at the generation-scale. Although extensive research remains to
15 be done as to the underlying mechanisms, it is possible that heterogeneous demographic process-
16 es are involved - for example *pulsatile* dispersal, in which catastrophic local events would prompt
17 massive dispersal during some generations, and no dispersal at all some others (see *Empirical evi-*
18 *dence of heterogeneous dispersal*, p. 259 for a possible instance of this phenomenon). In order to
19 describe this paradoxical structure, in this work, we use the term *synnome*, that we derive from
20 the Greek σύννομος, «*a common grazing of flocks*», that has in particular been used for gathering
21 flocks of birds¹, and which was commonly used by extension for «*reunions*», «*gatherings*», and
22 even «*kindred*» (see §123 p. 236). The pivotal concept of *grazing together* (or συν - νομός, *syn-*
23 *nomos*, «*shared pasture*») accurately renders the importance of central place foraging in the

1. See for example Aristophanes' *Birds* (v. 1755 sqq.): «*ἔπεσθε νῦν γάμοισιν ὦ | φῦλα πάντα
συννόμων | πτεροφόρ' ἐπὶ δάπεδον Διὸς | καὶ λέχος γαμήλιον.*» - «*let them now gather, all the
feathered flocks [synnómōn], and follow the brides to the houses of Zeus and to their wedding bed*».

1 colony structure of seabird species, while distinguishing it from more deeply fragmented
2 metapopulation systems. These different meanings together convey the particularity of the ob-
3 served structure, that we here define as « *one single, nearly homogeneous pool of individuals is that*
4 *distributed in a highly discrete way throughout its range. Local concentrations, or colonies, are highly*
5 *consistent on the scale of a few generations, at which scale philopatry may be the norm. Yet, migration*
6 *between these areas is high enough to maintain total homogeneity of the species' gene pool, so that,*
7 *viewed on a micro-evolutionary time-scale, the only relevant unit is the species as a whole.* »

8 *§-4 Dispersal and migration.* Populations may be defined, as we exposed earlier, under two diver-
9 gent paradigms - the *ecological* and the *evolutionary* ones (see §2 p. 25). Across both paradigms, a
10 population is a *bounded* entity: either geographically or genetically, it is characterised by its *autar-*
11 *kic* aspect. Any kind of movement amongst groups works against the constitution of population
12 units. Demographically, movement promotes the coupling of population trajectories, and
13 counters the evolution of demographic independence. Genetically, movement provokes the trans-
14 fer of alleles between demes, and works towards homogenising genetic variation and allele fre-
15 quencies. However, the distinction between the ecological and the evolutionary paradigms also
16 apply to inter-group movements, and, here again, prompted the choice of different terms. An un-
17 fortunate problem arises, however, because of the convergent choice of « migration » to describe
18 very different phenomena, that we will clarify here.

19 The most common sense of *migration*, especially in a bird-biology context, is the seasonal move-
20 ments of groups of birds (often *colonies*) between distinct breeding and overwintering grounds.
21 Migratory species, in that sense, have evolved particular adaptations (see *e.g.* Berthold 1991) that
22 allow them to achieve well-timed departure and arrivals to track the most beneficial environ-
23 ments year-round. Interestingly, migratory behaviour is a challenge to the definition of popula-
24 tions in the ecological paradigm, since groups may have different geographical and demographic
25 boundaries in their summer, and winter ranges. However, migratory behaviour in that sense will
26 not be examined in this work, mainly because it is not classically observable in our focal system

1 (although the inter-breeding foraging trips of penguins may arguably be related to migratory
2 behaviour).

3 A more simple, one-way movement is the *dispersal* of individuals out of their original group. Dis-
4 persal was originally described by Howard 1960 as «*the movement the animal makes from its point*
5 *of origin to the place where it reproduces*». From each individual's perspective, it is «*the greatest dis-*
6 *tance its genetic characteristics are transmitted, rather than the greatest distance the animal may have*
7 *migrated or otherwise travelled away from the place it was conceived, hatched or born*». Dispersal can
8 be defined both in the ecological and in the genetic paradigms. Ecologically, dispersal is the force
9 working towards demographic linkage of populations, as an excess of juveniles in one location
10 can compensate a low recruitment in another through direct displacement of individuals. Geneti-
11 cally, a dispersal event is the «unit» of gene flow, with each dispersing individual bringing a set of
12 alleles from one location to another.

13 The second sense of *migration*, and the one we will use throughout this work, has been described
14 by Dingle and Drake (Dingle & Drake 2007) in a biogeographical context as «*range expansions of*
15 *faunas or individual species*», such as «*the northward extension of ranges following the retreat of glaci-*
16 *ers at the end of the ice ages*». More specifically, in a population genetic context, the migration pa-
17 rameter M has been defined by the same authors as «*the exchange of genes among populations by*
18 *whatever means, including but not limited to migration as we consider it here*». It is used in that
19 sense in the coalescent framework, in particular by Beerli and colleagues (Beerli 2006; Beerli &
20 Palczewski 2010, see more details in §41 p. 103). In this sense, *migration* is the averaged extent,
21 on the long term, of individual dispersal events. Whereas dispersal is an individual- and genera-
22 tion-centred phenomenon that may be observed directly, migration is a time-averaged, popula-
23 tion-centred event that is only detectable through indirect methods, such as gene flow recon-
24 struction. Thus, while dispersal properly belongs to the life-history theory¹, and is expected to be

1. In a word, the life history theory reconsiders traits of anatomy, physiology and behaviour in the light of «life history traits», *i.e.* reproductive, demographic or foraging strategies (see Stearns 1992 for details).

1 highly variable depending on individual characteristics and temporary environmental conditions,
2 migration is a structural characteristic of the system, and is only expected to vary slowly accord-
3 ing to broad changes in the organisation of a species.

4 *§-5 Philopatry and site fidelity.* Philopatry has first been described as *Ortstreue* («site fidelity») by
5 von Haartman (Von Haartman 1949), as the tendency of the pied fly-catcher *Ficedula hypoleuca*
6 to breed near its birth place, and has been generalised under the term *philopatry* by Huntington
7 (Huntington 1951) as «*the tendency of an animal to return to its birthplace or breeding place, par-*
8 *ticularly for breeding*». Thus, although the term has occasionally been used to describe adult fi-
9 delity to its breeding spot (Anderson *et al.* 1992), it describes, in the strict sense, the fidelity of
10 adults to their birth place, or «natal philopatry», and is only properly used in that sense (Pearce
11 2007). Philopatry *sensu stricto* has important implications for the genetic processes that shape a
12 species' architecture. The tendency to breed near one's birth place implies a significant departure
13 from the panmixia hypothesis, since the probability of mating, for two randomly chosen individ-
14 uals, becomes inversely proportional to their distance at birth. At a low level, this generates pat-
15 terns of isolation by distance, where relatedness between individuals is a function of their geo-
16 graphical separation (Wright 1943; Wright 1946); at higher intensity, this may lead inbreeding
17 and population fragmentation (Mayr 1963; Avise *et al.* 2000). However, the benefits of philopa-
18 try may be considerable. They may be behavioural, such as increased knowledge of the higher
19 quality breeding spots or partners (Wheelwright & Mauck 1998; Heg *et al.* 2011; Arnaud *et al.*
20 2012), or enhanced selective value of proximal defensive behaviour (Dunford 1977) and allofeed-
21 ing (Lecomte *et al.* 2006) through kinship selection; or they may be genetic, *e.g.* through the pro-
22 motion of local micro-adaptation (Richardson *et al.* 2014) - which may apply either on land, or
23 on the associated foraging areas, which are often specific to a colony (see Weimerskirch 2013).
24 Philopatric behaviour is considered to be the basis of coloniality (Bowler & Benton 2005). How-
25 ever, there is an important discrepancy between the genetic and ecological time scales. On an
26 ecological scale, the mean return rate of juveniles to their natal colony over a few generations is

Introduction - §6

1 generally taken as an estimator for the strength of philopatry. In several species, in particular
2 seabirds, very high observed return rates seem to be inconsistent with the lack of genetic structure
3 between colonies - the so-called *seabird paradox* identified by Milot *et al.* (Milot *et al.* 2008).
4 However, based on simulation approaches, comparatively low (5% to 10%) migration rates have
5 been shown to be sufficient to counterbalance the effects of local genetic drift on a longer time
6 scale (Waples & Gaggiotti 2006). Moreover, the assumption that philopatry is a stationary
7 process in colonial species, and that the mean return rate should converge over time to the species
8 philopatry, does not take into account the possibility of pulsatile mass-dispersal events, that may
9 reshuffle gene pools entirely amongst several populations. Philopatry, as opposed to the instanta-
10 neous probability of fidelity (F) or dispersal ($1 - F$), is not a directly estimable parameter
11 (Kendall & Nichols 2004; Pearce 2007), but rather a long-term behavioural trait of the species,
12 that may or may not lead to a choice of site fidelity, depending on the actual environmental
13 conditions.

14 *§-6 Philopatry and dispersal in oceanic systems.* The dynamics of oceanic structures differ dramati-
15 cally from terrestrial ones when it comes to dispersal (Steele 1985). Although at the lower levels
16 of the trophic system, passive long-distance dispersal is equally present in terrestrial and marine
17 environments, as atmospheric currents perform the same role as sea currents for seed or propaga-
18 tion dispersal (Cain *et al.* 2000; Nathan & Muller-Landau 2000; Cowie & Holland 2006; Nathan
19 2006; Nathan *et al.* 2008; Nikula *et al.* 2013), the difference in spatial organisation and structure
20 becomes more important for larger organisms, especially vertebrates (Carr *et al.* 2003). Terrestrial
21 ecosystems have typically higher habitat fragmentation, less dispersal and gene flow, and globally
22 more closed systems (Waser & Jones 1983; Turchin 1998; Carr *et al.* 2003): the interaction be-
23 tween latitudinal and altitudinal gradients results in very complex habitat distributions that con-
24 trast with the mostly zonal pelagic habitats (Burrows *et al.* 2014). On the other hand, oceans may
25 be considered as a *fluid landscape* in which both active and passive dispersal occurs on very large
26 scales (Queiroz 2005; Cowie & Holland 2006; Nikula *et al.* 2013). Habitat fragmentation, al-

1 though documented (Acosta 1999; González-Wevar *et al.* 2010), is usually restricted to coastal or
2 benthic species with extremely specialised niches (Rex *et al.* 1993; Whitlatch *et al.* 1998; Poulin
3 *et al.* 2014). Long distance passive (*i.e.* current-driven) dispersal has been observed in organisms
4 ranging from planktonic species (*e.g.* vagile stages of kelp, or invertebrates, see Fraser *et al.* 2011;
5 Nikula *et al.* 2013) to larger vertebrates, especially at early life stages (*e.g.* turtles, see Gaspar *et al.*
6 2012). Active dispersal by vertebrate species is common at transoceanic scales (Bowen & Siniff
7 1999; Le Boeuf *et al.* 2000).

8 In oceanic systems dominated by strong currents (such as, in this study, the Antarctic Circumpo-
9 lar Current system (ACC - see §7 p. 35, or the Humboldt Current system - see §147 p. 276),
10 passive and active dispersal are general features of communities. In the ACC, it has been evi-
11 denced for species ranging from algæ and small invertebrates (Nikula *et al.* 2010), to benthic
12 species (Arango *et al.* 2011) and fish (Matschiner *et al.* 2009; Damerou *et al.* 2012). However, the
13 extent of genetic diversity along circumpolar ranges remains variable amongst taxa, and appears
14 to depend largely on specific dispersal potential and life-history traits (Rogers 2007). In many
15 cases, discontinuities such as the Antarctic Peninsula region (which extends North to almost
16 62°S) break these homogeneous dispersal systems, and stand out as distinct provinces, both in
17 terms of climate (Mulvaney *et al.* 2012) and of biogeography¹ (Terauds *et al.* 2012).

18 The active dispersal power of vertebrates has, however, one paradoxical consequence: their higher
19 ability to move around oceanic systems allows them to *not* disperse along the currents. While ses-
20 sile species resist to dispersal by sheer inertia, active swimmers and flyers exhibit *philopatric be-*
21 *haviour* (see §5 p. 32 and Frederiksen & Petersen 1999; Steiner 2005; Bicknell *et al.* 2012; Fer-
22 nández-Chacón *et al.* 2013), which can be thought of as counteracting the high fluidity of
23 oceanic environments through opposed active movement. Philopatry thus allows for the perpetu-
24 ation of relatively stable demes (*e.g.* in fur seals, see Bonin *et al.* 2013), family groups (as in killer

1. Broadly considered here as the distribution and co-occurrence of species through space.

1 whales, see Hoelzel 1998), or colonies (as in most seabirds, see Friesen *et al.* 2007). Indeed, in
2 most cases, natal philopatric behaviour is strong enough to promote genetic differentiation even
3 between populations that largely overlap in their foraging or wandering areas (Dearborn *et al.*
4 2003; Friesen *et al.* 2007; Smith *et al.* 2007). Such patterns have been observed in marine mam-
5 mals (both cetaceans and pinnipeds, see Hoelzel 1998; Baker *et al.* 2008; Bonin *et al.* 2013), in
6 turtles (Molfetti *et al.* 2013), and in most seabirds (Friesen *et al.* 2007; Bicknell *et al.* 2012).
7 Thus, the possibility of fast and unhindered movements in oceanic system has the unexpected
8 consequence of allowing high return rates to particular breeding locations in most vertebrates,
9 rather than promoting complete genetic mixing.

10 *Antarctica and the Southern Ocean*

11 *§-7 Oceanography and geography.* The Southern Ocean is the vast water mass encircling the
12 Antarctic continent, and one of the major thermal exchange points between the oceans, the at-
13 mosphere and the cryosphere (Hofmann & Maqueda 2011; Meijers 2014). While its official defi-
14 nition fixes its northern limit at 60°S (International Hydrographic Organization 2000), its func-
15 tional definition is usually extended at least to the Antarctic Polar Front, around 55°S (Moore *et*
16 *al.* 1999; Gersonde *et al.* 2005), where it blends into the Southern Atlantic, Indian and Pacific
17 oceans (see Fig. 2 p. 37 and Fig. 7 p. 77).

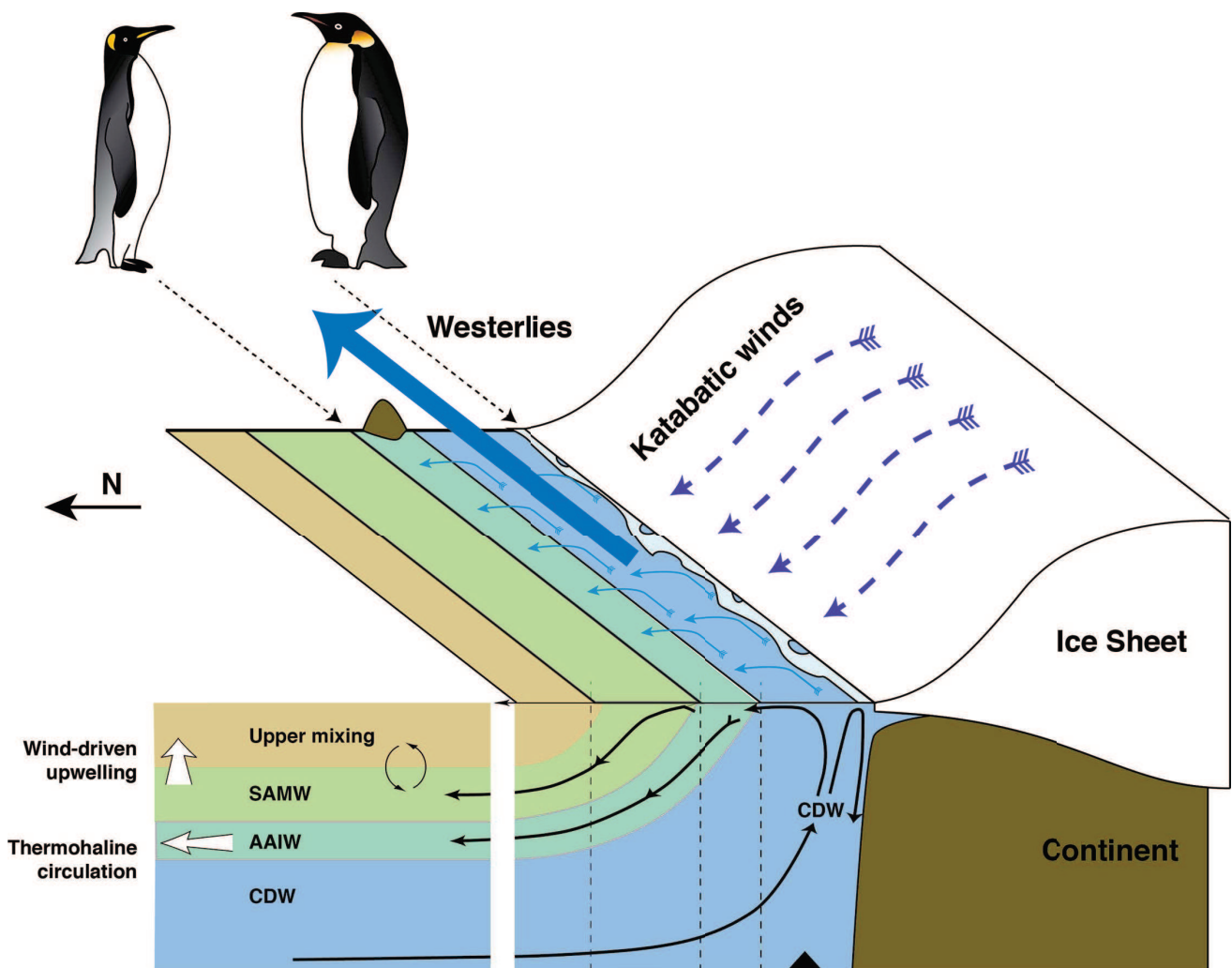
18 Southern Ocean oceanography is dominated by the influence of the strong planetary Westerly
19 winds in the southern Ferrell cell (De Boer *et al.* 2008; Kohfeld *et al.* 2013) that blow continu-
20 ously over the middle latitudes, unhindered by land barriers in the Southern Hemisphere, bring-
21 ing the surface waters into motion. The resulting West Wind Drift, or Antarctic Circumpolar
22 Current (ACC) is one of the largest ocean currents on earth, with a mean transport of 105 to 140
23 Sverdrups, and the only one to directly connect the three major ocean basins (Rintoul *et al.*
24 2001). Westerly wind stress' main effect is the eastward flow of the ACC, but it also causes im-

1 portant Ekman transport (in the Southern hemisphere, a counterclockwise deviation of up to 90°
2 of the surface drift compared to the wind stress, see Toggweiler & Samuels 1998) that brings the
3 surface waters equatorwards. In the Antarctic region, wind and evaporative cooling under kata-
4 batic wind stress, and salt concentration through brine rejection during sea ice formation around
5 polynyas (Thomas & Dieckmann 2008, see §18 p. 65), make the surface waters especially cold
6 and salty - and consequently dense. During their equatorwards transport, these water masses con-
7 verge with the warmer and less dense northern water masses: the resulting steeper isopycnals¹
8 mean that the Antarctic surface waters ultimately sink to the sea floor at the Antarctic Conver-
9 gence or Antarctic Polar Front (APF), forming the Antarctic Bottom Waters (AABW, see Rintoul
10 *et al.* 2001). South of the APF, on the other hand, Ekman transport is divergent, and causes the
11 upwelling of warm circumpolar deep waters (Warm Deep Water, WDW and Circumpolar Deep
12 Water, CDW, see Sarmiento *et al.* 2004; Anderson *et al.* 2009 for details) south of the APF.
13 North of the APF lies the Subantarctic region, bounded to the North by the Subtropical conver-
14 gence (STC) where Antarctic and Subantarctic waters become stratified under the tropical surface
15 waters, around 40°S. While nowhere near as sharp as the APF, the STC is best defined by the dis-
16 appearance of the permanent thermocline² (Tomczak & Godfrey 2003): north of the STC, tem-
17 perature drops rapidly from the surface towards the bottom until ~1000 m. South of the STC,
18 on the other hand, the temperature gradient between the surface and deep waters becomes less
19 and less pronounced, with sometimes less than 1°C difference in the Antarctic region. The APF,
20 however, remains characterised by minor seasonal thermoclines at the lower limit of the surface
21 mixed layer (SMC, see *e.g.* Charrassin & Bost 2001). South of the ACC, and close to the Antarc-
22 tic continent, on the other hand, the Antarctic Coastal Current, or East Wind drift, flows in the
23 opposite direction of the ACC over the narrow continental shelf (Tchernia & Jeannin 1980;

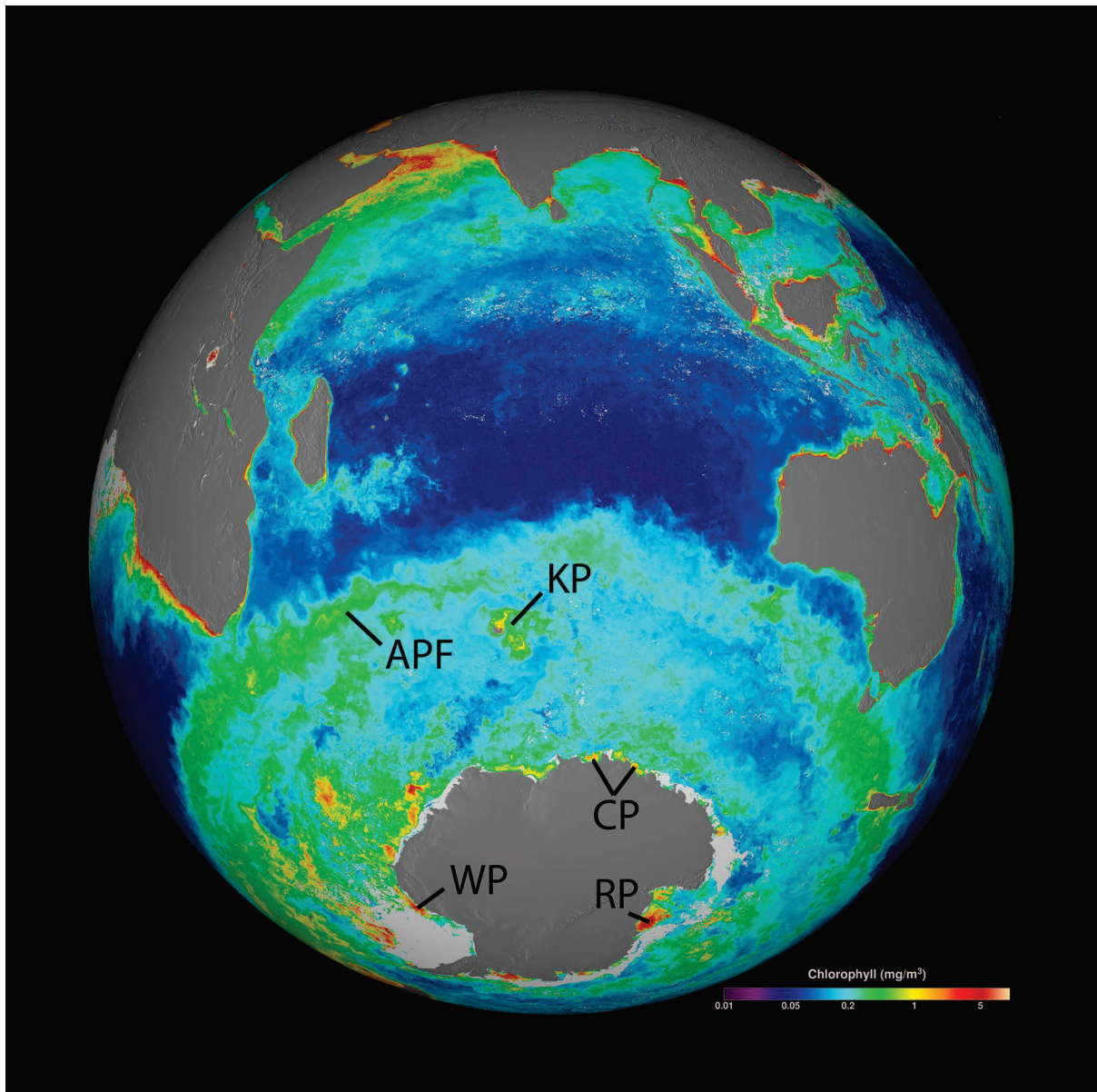
1. Or planes of equal water density. As opposed to pressure which is mostly determined by depth, density is determined primarily by temperature and salinity.

2. The *thermocline* is the transition zone between the surface mixed layer, where diurnal solar radiation and wind- and tide-driven water mixing result in rather homogeneously warmer temperatures, and deeper and colder waters. It is characterised by a steep temperature gradient in the water column.

1 Fahrbach *et al.* 1994). Both currents form the two main gyres of the Southern Ocean, in the Ross
 2 and Weddell seas (Gouretski 1999; Meredith *et al.* 2000), that have distinct oceanographic fea-
 3 tures - for example the warmer deep water upwelling and resulting stable sensible heat polynya in
 4 the Weddell sea (*i.e.* a polynya caused by water heat rather than wind stress, see Thomas &
 5 Dieckmann 2008).



6 **Figure 2 | Structure of water masses in the Southern Ocean.** Main layers of the Southern Ocean:
 7 Circumpolar Deep Water (CDW), Antarctic Intermediate Water (AAIW), Subantarctic Middle Water (SAMW) and
 8 upper mixing layer. Thick blue arrow represents Westerly winds and the main flow of the Circumpolar current.
 9 Lighter arrows in the Antarctic Circumpolar Current represent Ekman transport (*Based on Sarmiento et al. 2004*).



1 **Figure 3 | Primary productivity concentration on the Antarctic Polar Front.** Surface chlorophyll
2 concentration, 21 dec. 2011 to 20 mar. 2012. High chlorophyll concentration is visible all along the APF, with a local
3 peak on the Kerguelen plateau (KP), as well as in latent heat coastal Antarctic polynyas (CP), and in the sensible
4 heat Weddell polynyas (WP) and latent and sensible heat Ross Sea Polynya (RP) - see §18 p. 65 for details). Image
5 courtesy of NASA, Suomi NPP VIIRS satellite program.

6 The importance of the Southern Ocean and of the ACC, however, goes well beyond their regional
7 effects. Together with the North Pacific, the ACC is the major upwelling area for ancient bottom
8 waters (Primeau 2005; Marshall & Speer 2012), and therefore a central contributor in the
9 thermohaline circulation, or Meridional Overturning Circulation (MOC, see Marshall & Speer

1 2012), which makes its importance central for global heat exchange and circulation. As it is one
2 of the few areas where deep waters outcrop directly at the surface, it is also central for gas exchange
3 between the deep ocean and the atmosphere, and enhanced upwelling activity has been linked to
4 a rise in atmospheric CO₂ in the Southern Ocean (Anderson *et al.* 2009). But the Southern
5 Ocean upwelling has also been identified, together with the North Pacific, as the major return
6 path of nutrients towards the surface after downwelling at lower latitudes, which makes it a key
7 component of the global biological pump¹, accounting for up to three quarters of the productivi-
8 ty north of 30°S (Sarmiento *et al.* 2004). Finally, being the only zonally unbounded water mass
9 in the world, and blending into the three main ocean basins, the Southern Ocean functions as a
10 central coupling mechanism between otherwise independent water masses, and allows for tele-
11 connections between basins (White & Peterson 1996; Rintoul *et al.* 2001). It also spills out in
12 the major northward Eastern Boundary Currents of the three Southern Hemisphere ocean basins
13 (the Humboldt Current on the western shore of South America - see §147 p. 276, the Benguela
14 Current on the western shore of Africa, visible on Fig. 3 p. 38, and the lesser West Australia
15 Current).

16 The local influence of the ACC and of the APF on biological processes is tremendous. The up-
17 welling of nutrient-rich deep waters to the surface, and their increased oxygen concentration in
18 the surface mixed layer, is responsible for a massive concentration of primary productivity in the
19 APF area (up to 20% of the world's marine primary productivity, see Laubscher *et al.* 1993;
20 Bathmann *et al.* 1997; Carr *et al.* 2006, and Fig. 3 p. 38). Enhanced primary productivity in
21 turns sustains stocks of krill (*Euphausia superbia*, see Murphy *et al.* 2007a) and fish (Pakhomov *et*
22 *al.* 1996) - these being dominated by the myctophids (mostly *Electrona sp.*, *Protomyctophum sp.*,
23 *Gymnoscopelus sp.* and *Krefftichthys sp.*, see Collins *et al.* 2008 for details). Several top predators
24 take advantage of prey aggregation on the APF - for example elephant seals *Mirounga sp.* (Boyd

1. Or, put simply, the global circulation of nutrients, that integrate both active components (*e.g.* animals themselves), and passive components (*e.g.* drift with water masses).

Introduction - §8

1 & Arnbohm 1991) or king penguins *Aptenodytes patagonicus* (Charrassin & Bost 2001). At the
2 southern boundary of the ACC, the extended winter sea ice cover also promotes blooms of ma-
3 rine productivity: years of increased sea ice have the highest krill biomass (Nicol *et al.* 2000), with
4 direct consequences on the populations of top predators (*e.g.* the Adélie penguin *Pygoscelis adeli-*
5 *ae*, Nicol *et al.* 2008), making the coastal pack-ice zone a distinct high-productivity area (Tynan
6 1998).

7 *§-8 Climate variability in the Southern Ocean.* The complex structure and global connections of
8 the Southern Ocean make it respond strongly to the major modes of climate variability in the
9 Southern Hemisphere: the circumpolar Southern Annular Mode (SAM), and the subtropical di-
10 pole events, of which the most important is the El Niño Southern Oscillation (ENSO), but
11 which also include the Southern Indian Ocean Dipole (SIOD) and the South Atlantic Subtropi-
12 cal Dipole (SASD).

13 The Southern Annular Mode is the average pressure difference between the mid-latitudes and
14 Antarctica. The resulting atmospheric circulation is the major cause of climate variability in the
15 Southern Ocean (Marshall 2003; Abram *et al.* 2014), with important decadal-scale cyclicity,
16 and broader centennial-scale oscillations (Abram *et al.* 2014). SAM has opposite effects in the po-
17 lar and subtropical zones. During positive SAM events, the poleward shift of the westerlies
18 strengthens the Ekman transport in the ACC, resulting in an equatorward shift of the APF, in-
19 creased upwelling activity south of the APF, and increased marine productivity in the polar
20 frontal zone (Lovenduski & Gruber 2005) - while in the subtropical zone, warm SST anomalies
21 result in decreased productivity. Conversely, negative SAM events (*i.e.* northwards expansion of
22 the Westerlies) have a strong effect on the whole food chain, up to top predators (Forcada &
23 Trathan 2009; Bost *et al.* 2015).

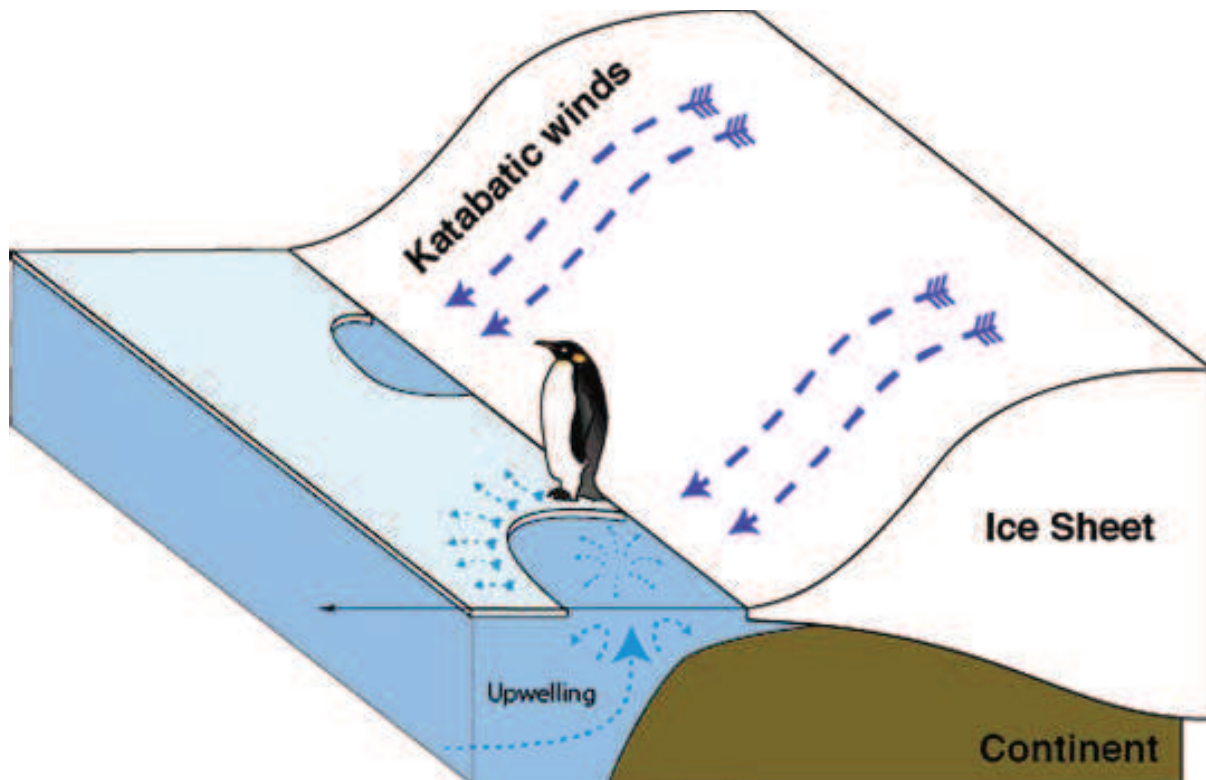
24 While the impact of the SAM is essentially felt all around the Southern Ocean, each ocean basin
25 is also subject to the more local (although tele-connected) effects of the dipole events. While the
26 El Niño phenomenon differs significantly from the South Indian and Atlantic dipoles, its

1 general structure remains the oscillation of temperature differences between the East and the
2 West of the subtropical zone. A steep zonal temperature gradient guarantees strong trade winds
3 and intense Walker circulation¹. The periodic diminution of winds off the West coast of South
4 America diminishes cold deep water upwelling, reduces the east-west temperature difference
5 across the Pacific, and, as a result, also diminishes the Walker circulation: this brings high precipi-
6 tations so South America, and high sea surface temperatures across the whole Pacific (Tomczak &
7 Godfrey 2003; Stuecker *et al.* 2013). More recently, similar cyclical events arising from zonal
8 temperature gradients have been identified in the Southern Indian Ocean (the SIOD, Saji *et al.*
9 1999) and in the South Atlantic (the SASD, Taschetto & Wainer 2008; Wainer *et al.* 2014).
10 Their link with the ENSO, and their global connections, however, still remain to be fully investi-
11 gated (Saji & Yamagata 2003; Ashok *et al.* 2004; Abram *et al.* 2008). Together, the SAM, ENSO
12 and dipole events also contribute to the variability in sea ice extent around Antarctica (Simmonds
13 & Jacka 1995; Lefebvre 2004; Stammerjohn *et al.* 2008).

14 The integrated effect of the SAM, the ENSO, and the dipole events result in combined modes of
15 variability in the Southern Ocean (Fogt & Bromwich 2006; Ciasto & Thompson 2008; Fogt *et*
16 *al.* 2011), that have deep impact on biological communities. Primary productivity in the South-
17 ern Ocean responds both to the SAM (Lovenduski & Gruber 2005) and to ENSO (Beaufort *et*
18 *al.* 2001; Behrenfeld *et al.* 2001), mostly mediated through changes in surface water temperature,
19 sea ice, and nutrient availability. Both changes in the physical properties of surface waters and in
20 trophic interactions also have an impact at trophic levels above phytoplankton, *e.g.* on the
21 Antarctic krill (Murphy *et al.* 2007b): in turn, these effects are passed on to top-level predator
22 species such as right whales (Leaper *et al.* 2006), elephant seals (McMahon & Burton 2005) or
23 penguins (Bost *et al.* 2015). Although climate variability is a perfectly natural phenomenon, and
24 should not threaten species resilience, the highly integrated character of the Southern Ocean cli-
25 mate makes it liable to entering large feedback mechanisms, thereby amplifying the effects of cli-

1. *I.e.* a zonal circulation of air masses in the tropics, driven mostly by east-west temperature and pressure differences.

1 mate change (see §12 p. 52). It is therefore of the utmost importance to gain a good understand-
2 ing both of the climatic processes themselves and of their effects on biological communities.



3 **Figure 4 | The formation of sea ice in coastal polynyas.** A schematic representation of coastal sea ice for-
4 mation (light blue arrow on the ice) and local upwelling around polynya area (see §18 p. 65 for details).

5 *§-9 Climatic history of the Southern Ocean.* The Southern Ocean is considered to have taken its
6 present-day shape by the onset of Pliocene (-5.3 Myr). The physical separation between Antarcti-
7 ca and South America is thought to have occurred between the mid-Eocene (~41 Myr ago, Liver-
8 more *et al.* 2005) and the Oligocene-Miocene transition (-23 Myr ago, Pfuhl & McCave 2005),
9 together with the slow onset of icehouse-earth conditions - the opening of the Drake passage it-
10 self having possibly contributed to the global temperature decrease (Nong *et al.* 2000; Livermore
11 *et al.* 2004). The Antarctic Circumpolar Current (see §7 p. 35) is thought to have settled around
12 the early Miocene (Pfuhl & McCave 2005), but the present-day water mass boundaries appear to
13 have been established later, perhaps around the Middle Miocene Climatic Transition (MMCT,

Introduction - §9

1 ~16-11.5 Myr, see Haywood *et al.* 2008; Knorr & Lohmann 2014). This idea is confirmed by the
2 DNA-inferred radiation date of several marine taxa, for which Antarctic and Subantarctic clades
3 have been isolated from each other around the MMCT (Poulin *et al.* 2014), which is also coinci-
4 dent with the radiation of crown penguins (see §21 p. 72).

5 Miocene marks the transition from greenhouse-earth conditions to present-day-like icehouse
6 conditions. The Antarctic Ice Sheet is thought to have taken over the prevailing tundra environ-
7 ment by the early Miocene (Raine & Askin 2001), and developed to reach its current extent by
8 the end of the period (Westerhold *et al.* 2005). Despite high (and still not fully understood) vari-
9 ability, prevalent marine conditions were warmer during the early Pliocene than today (Haywood
10 *et al.* 2008) - although with high marine productivity. The SST was about 5°C warmer than to-
11 day throughout the Southern Ocean (Whitehead 2003), and the APF lay about 6° South of its
12 present position in the mid-Pliocene (Barron 1996), while the global thermohaline circulation
13 may have been stronger than observed now (Haywood *et al.* 2008). By the late Pliocene (-2.5
14 Myr), however, Antarctic cooling strengthened, with increased Westerlies and equatorward mi-
15 gration of the APF, and possibly a reduction in the thermohaline circulation (McKay *et al.* 2012).

16 Most of the Subantarctic islands came to existence during that period. Apart from the continen-
17 tal-shelf South Georgia and Falklands (that are remnants of Gondwana, and as old as neighbour-
18 ing South America), and sub-continental Kerguelen islands (that were formed during the
19 Oligocene), the subantarctic islands are mostly young, volcanic mounds raised from the sea-floor.
20 Crozet, Heard and Macquarie were thus formed during the late Miocene, and Prince Edward,
21 Bouvet, Gough, etc. appeared between the Pliocene and the Pleistocene - and as late as 0.45 Myr
22 ago for Prince Edward (see a complete review in Quilty 2007). Thus, both in terms of geography
23 and of oceanography, we may consider that the Southern Ocean did not take its present-day
24 shape before the Pleistocene.

25 Although better known, Pleistocene (-2.6 Myr to 11.7 Kyr BP) and Holocene (~11.7 Kyr BP to
26 the present) history in the Southern Ocean is still subject to considerable uncertainties. General-

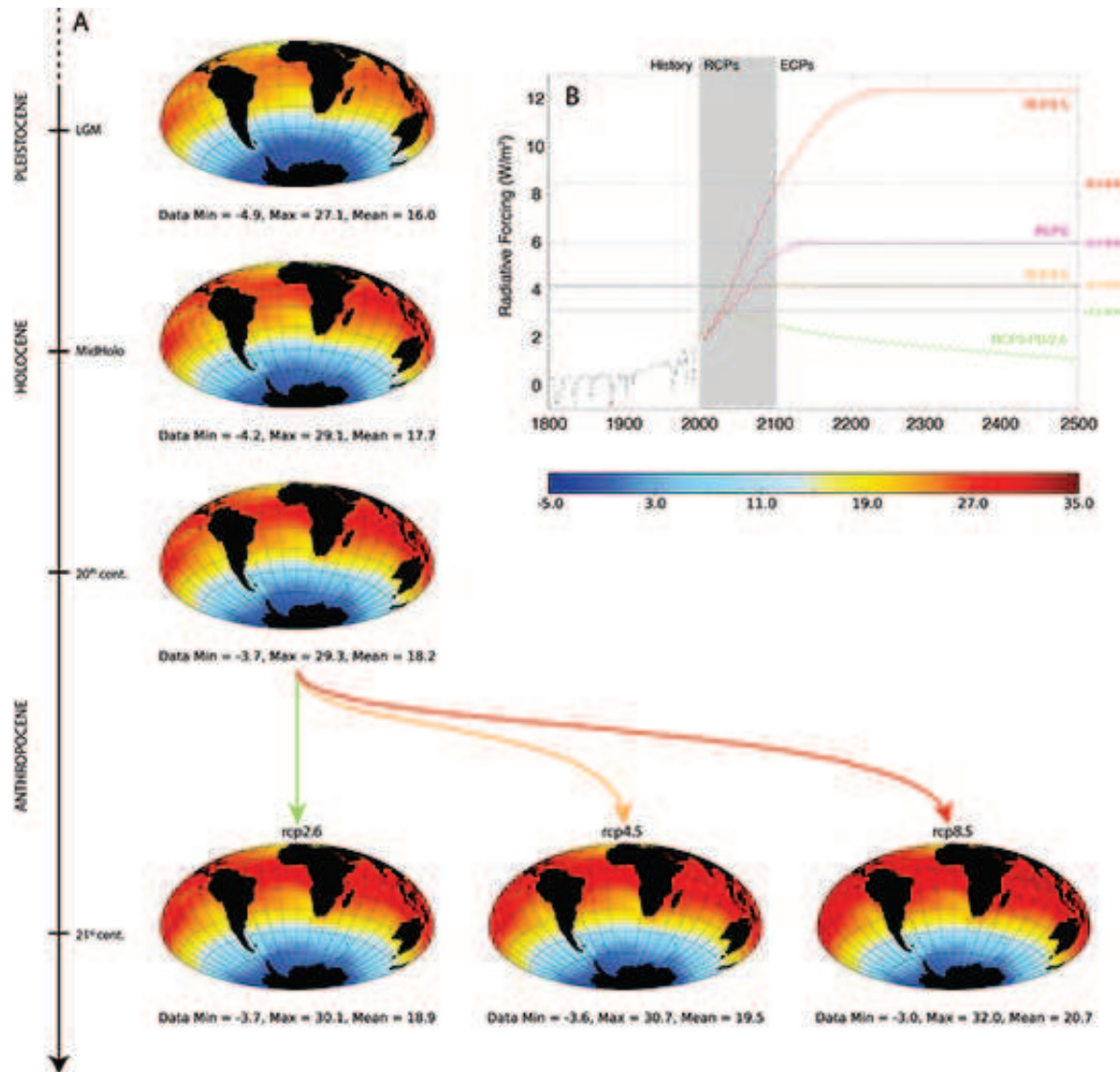
1 ly, reconstructing oceanographic and atmospheric processes at these relatively recent periods relies
 2 on to different approaches. First, direct or indirect evidence may be obtained from cores, whether
 3 of marine sediment throughout the Southern Ocean, of ice on the Antarctic ice sheet, or of peat
 4 on the subantarctic islands. Alternatively, General Circulation Model (GCM) developed for the
 5 historical period may perform rather well in that recent past, since the major features of Ocean-
 6 Atmosphere circulation are assumed stable over that period: the Coupled Model Intercomparison
 7 Project phase 5 (CMIP5, see §50 p. 130) panel thus includes Pleistocene Last Glacial Maximum
 8 and mid-Holocene experiments (see Fig. 5 p. 45 and §50 p. 130 for details on these experi-
 9 ments). Core evidence provides chemical, physical and biostratigraphic markers that may be used
 10 as proxies for a variety of variables Gersonde *et al.* 2005; Armand & Leventer 2010; Hodgson *et*
 11 *al.* 2014, either directly at the location of the core (which is how marine sediment and terrestrial
 12 peat cores are usually interpreted, see Gersonde *et al.* 2005; Martínez Garcia *et al.* 2009; Armand
 13 & Leventer 2010), or as a mixture of local and oceanic source area signal (see Wolff *et al.* 2003;
 14 Wolff *et al.* 2006). Foraminiferan and diatom assemblages in marine sediment cores offer an ac-
 15 curate index of the state of the surface water masses, that can be used to locate the position of the
 16 Southern Ocean fronts through changes in surface temperature (*e.g.* the complex frontal structure
 17 around the Campbell plateau, in Neil *et al.* 2004, or the APF in the Indian and Atlantic oceans,
 18 in Kemp *et al.* 2010) or the extent of sea ice (*e.g.* Hodell *et al.* 2001; Gersonde *et al.* 2005). Fossil
 19 pollen in terrestrial or marine cores provides direct evidence as to the prevailing climatic condi-
 20 tions (*e.g.* McGlone *et al.* 2010). Chemical indices can be extracted from ice cores¹ and marine
 21 sediment cores². These multiple proxies are used either directly, or as an input in Atmosphere-

1. Considering they have been deposited at the core site as wind-borne aerosols, this requires hypotheses as to the atmospheric circulation at the core site (see Wolff *et al.* 2006): these normally include methanesulphonic acid (MSA, as a proxy for marine dimethyl sulphide, a biomarker that reflects algal growth and thence marine productivity - see Legrand *et al.* 1991), sea salt sodium (freed as high-salt brine at the surface of sea ice, and thus indicating sea ice extent, see Wolff *et al.* 2003), iron and calcium (transported as dust, and revealing atmospheric circulation, see Wolff *et al.* 2006). Deuterium/hydrogen ratio ($\delta D = {}_2H/{}_1H$) is commonly used as a proxy for temperature anomaly (Augustin *et al.* 2004; Wolff *et al.* 2006), and can be combined with the record of greenhouse gases (CO₂ and CH₄ mostly, see Fischer *et al.* 1999; Petit *et al.* 1999; Augustin *et al.* 2004).

2. Marine sediment cores, besides isotopic signatures ($\delta_{18}O$, see Lisiecki & Raymo 2005) and chemical signatures such as iron flux (from continental input, as a marker of marine productivity, see Martínez Garcia *et al.* 2009), also provide biomarker ev-

Introduction - §9

1 Ocean General Circulation Models, to reconstruct the prevailing oceanographic and climatic
2 conditions. However, reconciling these diverse sources is still problematic, with mixed signals be-
3 tween the glacial, oceanic, and terrestrial environments (Armand & Leventer 2010; McGlone *et*
4 *al.* 2010).



5 **Figure 5 | Sea Surface Temperature changes in the Southern Ocean.** A. 20-years mean sea surface
6 temperature at four different time-points (from a 14-member ensemble, see §102 p. 201), from the CMIP5 panel
7 (see §50 p. 130 for details), and under three different forcing scenarios (in B, adapted from Meinshausen *et al.*
8 2011).

idence such as alkenones or Mg/Ca ratio (as a proxy for sea surface temperature, see Kucera *et al.* 2005; MARGO 2009).

1 Despite local variations and alternating conditions in the different oceanic basins, we can separate
2 four major periods in the Southern Ocean recent history, that appear consistent at the global
3 scale: (i) Quaternary conditions (mid-Pleistocene to 22 Kyr BP), (ii) LGM conditions (21-18
4 Kyr BP), (iii) Pleistocene glacial retreat and early holocene optimum (17-9 Kyr BP), (iv)
5 Holocene hypsithermal and neoglacial conditions (8-0 Kyr BP).

6 (i) *Quaternary conditions (1.3 Myr - 22 Kyr BP)* are characterised by the slow onset of glacial con-
7 ditions, in the continuity of the Miocene-Pliocene icehouse trend (Haywood *et al.* 2008), al-
8 though with marked glacial-interglacial cycles with a phase of ~100 Kyr (Petit *et al.* 1999; Au-
9 gustin *et al.* 2004; Wolff *et al.* 2006), during which the APF exhibits important latitudinal
10 variations (see Kemp *et al.* 2010). Winter sea ice cover already reached as far as ~56°S in the Pa-
11 cific during the period (Gersonde *et al.* 2005). Little is known of land ice throughout the period,
12 as the following LGM glacier growth obliterated most of the direct evidence (Hodgson *et al.*
13 2014).

14 (ii) *Last Glacial Maximum (LGM, 21-18 Kyr BP)* has been the focus of much research, in particu-
15 lar regarding the Antarctic ice sheet (*e.g.* Anderson *et al.* 2002; Denton & Hughes 2002), the
16 Southern Ocean frontal structure and sea ice extent (*e.g.* Gersonde *et al.* 2005), land ice cover
17 (*e.g.* Hodgson *et al.* 2014) and atmospheric circulation (*e.g.* Kohfeld *et al.* 2013). Late Pleistocene
18 glacial conditions culminate for the last time during the LGM, with a ~5°C cooling in summer
19 SST throughout the Southern Ocean (Gersonde *et al.* 2005; MARGO 2009) and associated
20 equatorward movement of the APF that may have reached 40°S at some locations (Kohfeld *et al.*
21 2013). Precise dynamics of the APF still remain to be understood, however, as the strong bathy-
22 metric constraints at several locations (the Campbell Plateau, the Kerguelen Plateau, or the
23 Drake passage) are thought to have prevented equatorward frontal displacement (Rosenthal *et al.*
24 1997; Neil *et al.* 2004; Kohfeld *et al.* 2013). Westerly wind change reconstruction is still uncer-
25 tain, but available data point to strengthened circulation and equatorward widening of the west-
26 erlies (Kohfeld *et al.* 2013). Winter sea ice expanded greatly during the period, reaching at least

Introduction - §9

1 50°S in winter (CLIMAP 1981; Gersonde *et al.* 2005 - with variation between the icy Atlantic
2 and Indian oceans, and the more ice-free Pacific). A shift in marine productivity has been in-
3 ferred from the Antarctic to the Subantarctic region (Hodgson *et al.* 2014), without however a
4 global reduction in biomass (Wolff *et al.* 2006). The islands of Heard, Crozet, Marion, and the
5 Drake Arc were entirely covered by ice, while Kerguelen and South Georgia may have had some
6 ice-free areas, and the Falklands and Macquarie underwent periglacial conditions, with a harsh
7 tundra environment (Hodgson *et al.* 2014). Like in the Northern Hemisphere, the prevailing
8 glacial conditions pushed the higher latitude terrestrial or semi-terrestrial biodiversity into a
9 handful of refugia, mostly located in the now subtropical (but then subantarctic) islands of
10 Gough and StHelena in the Atlantic, Amsterdam in the Indian Ocean, and Auckland and Camp-
11 bell in the Pacific. The Falklands and the Patagonian Shelf area (where swathes of land emerged
12 from the lower sea level) may also have granted ice-free refugia (see Fraser *et al.* 2009; Fraser *et al.*
13 2012). Marine diversity, on the other hand, was not severely affected (see *e.g.* Morin *et al.* 2015).

14 *(iii) Pleistocene glacial retreat and early Holocene optimum (17-9 Kyr BP)* is a period of global
15 thawing and warming following the LGM. In the Southern Ocean, the deglaciation has very
16 contrasting chronologies, in particular in relation to the local topography and bathymetry: while
17 the dry land expanses of Kerguelen and South Georgia, and the shelf-less archipelagos of the Sco-
18 tia Arc, saw ice retreating early, the shelves of Crozet and Prince Edward retained extensive land
19 ice caps as late as the early Holocene (Hodgson *et al.* 2014). Sea ice retreated slowly during the
20 late Pleistocene, with an important regain during the Antarctic Cold Reversal, around 14.7 Kyr
21 BP (likely due to the massive influx of meltwater from the Antarctic Ice Sheet, see Weber *et al.*
22 2014). The Younger Dryas Northern hemisphere cold reversal corresponds, on the other hand, to
23 a warm period in the Southern Ocean (Kaplan *et al.* 2010). Oceanic fronts moved poleward and
24 reached their current position by the beginning of the Holocene, during the so-called Holocene
25 Climatic Optimum (Kohfeld *et al.* 2013). Around ~9 Kyr BP, a second cold reversal in the

Introduction - §10

1 Antarctic waters led to a slight regain of sea ice (Nielsen *et al.* 2004) - although the state of the
2 Southern Ocean may not have been homogeneous between ocean basins.

3 *(iv) Holocene hypsithermal and neoglacial conditions (8-0 Kyr BP)* mark the advent of the modern
4 (pre-industrial) conditions throughout the Southern Ocean. Sea surface temperature reaches a
5 maximum around 7.5 Kyr BP (Calvo *et al.* 2007), with subantarctic ice-free conditions at $\sim 50^{\circ}\text{S}$
6 around 6-5 Kyr BP (Hodell *et al.* 2001; Nielsen *et al.* 2004). Land ice conditions are then similar
7 to the present, with complete deglaciation on Crozet and Prince Edward (Hodgson *et al.* 2014).
8 There warmer hypsithermal conditions are followed by colder conditions as early as 5 Kyr BP in
9 East Antarctica, and 3 Kyr BP in West Antarctica (Armand & Leventer 2010). However, the
10 oceanographic landscape is not thought to have changed much, let aside minor variation in sea
11 ice, and species distribution is thought to have been very similar to pre-industrial ranges.

12 *§-10 Biogeography of the Southern Ocean.* The present-day Southern Ocean has a rather simple
13 biogeography, made of broad, often zonally bounded systems, but that host a considerable diver-
14 sity of marine life (Smetacek & Nicol 2005; Chown *et al.* 2015). Four distinct areas may be dis-
15 tinguished: *(i)* the *Subantarctic zone*, which is outside the Southern Ocean proper, but biogeo-
16 graphically bound to it, is bounded by the Subtropical convergence to the North, and the APF to
17 the South, *(ii)* the *Antarctic Ocean zone*, south of the Subantarctic zone, and bounded to the
18 South by the Northern Limit of the Circumpolar Current, *(iii)* the *Coastal Antarctic zone*, which
19 includes the Antarctic coastal countercurrent and shelf area, as well as the coastal ice area, and
20 *(iv)* the *Antarctic peninsula* (Convey *et al.* 2012; Terauds *et al.* 2012). In this study, we are mostly
21 concerned with the first three concentric zones. The Subantarctic area is characterised by mild air
22 temperatures, high precipitation, absent sea ice cover, and prevailing Westerlies. Islands are dom-
23 inated by peat and low vegetation, or periglacial habitat. Subantarctic islands are few and far in-
24 between (from the Pacific to the Atlantic: Macquarie, Heard-and-McDonald, Kerguelen, Crozet,
25 Marion-and-Prince Edward, Bouvet, the South Sandwich, South Georgia, Tierra del Fuego and
26 the Falklands), and concentrate both avian and pinniped fauna. The Antarctic Ocean zone, South

Introduction - §10

1 of the Polar Front, has much colder surface waters and extensive winter sea ice cover, islands are
2 rare and ice-covered (from the Pacific to the Atlantic: Peter I, Scott, Balleny, and the South
3 Orkneys). Vegetation is scarce and mostly limited to lichens, liverworts and mosses (Convey *et al.*
4 2012). The Antarctic coastal zone (Tynan 1998), finally, is a shelf environment, characterised by
5 almost perennial pack-ice, and a strong influence of the Antarctic Ice Shelf. Vegetation is almost
6 absent, and displays extreme adaptations, and breeding vertebrate fauna is limited to a handful of
7 species: Adélie and emperor penguins, snow, cape, antarctic and giant petrels, Wilson's storm-pe-
8 trels, southern fulmars and southern polar skuas, as well as a few pinniped species. Of these, only
9 the Emperor penguin breeds during the Southern winter.

10 Foraging opportunities stem mostly from two sources. As mentioned earlier (see §7 p. 35), the
11 tremendous upwelling areas associated with the APF are a major contributor to the world's ma-
12 rine productivity (Laubscher *et al.* 1993; Bathmann *et al.* 1997; Carr *et al.* 2006), and attract
13 long-distance central-place foragers such as the King penguin (Péron *et al.* 2012). The pack-ice
14 area in the Antarctic and Coastal Antarctic areas also concentrate massive primary productivity
15 (Arrigo *et al.* 1997; Arrigo *et al.* 1998), and are an important food resource both for Antarctic
16 Coastal foragers such as the Snow Petrel *Pagodroma nivea* (Barbraud & Weimerskirch 2001b) or
17 the Adélie penguin (Wienecke *et al.* 2000). In particular, open-water and upwelling conditions
18 around polynyas along the coast (see §18 p. 65 and Fig. 4 p. 42) concentrate sea-ice-linked
19 productivity, and are critical foraging areas for several species (Ancel *et al.* 1992; Ancel *et al.*
20 1999). The most salient feature of the Southern Ocean biogeography is therefore its almost direct
21 link to very broad climatic features. Unlike terrestrial environments, where the complex interac-
22 tion of latitude and altitude may blur surface wind currents to a considerable extent and result in
23 very heterogeneous patterns (Burrows *et al.* 2014), the zonal structure of the Southern Ocean
24 places biogeographic boundaries in regular accordance with annular, circumpolar circulation pat-
25 terns. As a result, the extent and location of these biozones are directly influenced by global circu-
26 lation patterns (see §18 p. 65).

1 *Species' responses to anthropogenic climate change in the Southern Ocean.*

2 §-11 *Climate is changing.* Palæoclimate reconstructions teach us that climate change is, by itself, a
3 natural and unavoidable phenomenon at the millennial scale (see §9 p. 42), and may be linked to
4 external forcings (*e.g.* changes in solar activity or in Milankovitch orbital cycles, with positive
5 feedbacks such as the «snowball earth» loop, see *e.g.* Hyde *et al.* 2000; Berger 2013), or to inter-
6 nal biogenic modifications of atmosphere composition (*e.g.* Frolking & Roulet 2007). However,
7 the pace of these changes is slow. Even the LGM did not alter significantly the composition or
8 structure communities in Europe on the long term (see *e.g.* Yeakel *et al.* 2013). Current climate
9 change, on the other hand, sees modifications of the same magnitude as the Holocene transition
10 happening at the decadal - and not millennial - scale. Climate change in general has been unam-
11 biguously and rather intuitively defined by the Intergovernmental Panel on Climate Change
12 (IPCC) as a «*change in the state of the climate that can be identified (e.g. by using statistical tests) by*
13 *changes in the mean and/or the variability of its properties, and that persists for an extended period,*
14 *typically decades or longer. Climate change may be due to natural internal processes or external forcings*
15 *such as modulations of the solar cycles, volcanic eruptions and persistent anthropogenic changes in the*
16 *composition of the atmosphere or in land use» (Solomon *et al.* 2007). During the last decades of the
17 20th century, accumulated evidence increasingly pointed towards unprecedented rapid changes
18 in the climate of the earth, starting shortly after the 19th century Industrial Revolution. The
19 combination of deforestation and large-scale use of fossil fuels led to higher production, and low-
20 er fixation, of greenhouse gases (GHG), in particular carbon dioxide and methane, to the point
21 that «*atmospheric concentrations of carbon dioxide, methane and nitrous oxide [...] are unprecedented*
22 *in at least the last 800,000 years» (Pachauri *et al.* 2014). Increased GHG results in an increase in
23 greenhouse effect, or effective radiative forcing on global climate. Thus, all available evidence
24 make it «*extremely likely that human influence has been the dominant cause of the observed warming*
25 *since the mid-20th century» (Stocker *et al.* 2014).***

Introduction - §11

1 Intense and coordinated effort is currently deployed to understand the underlying mechanisms,
2 and possible outcomes, of ongoing climate changes (see §50 p. 130 for an overview of the In-
3 tergovernmental Panel on Climate Change) - the same can unfortunately not be said of political
4 endeavours. The result of this scientific effort is a growing understanding of the articulation be-
5 tween three different domains: *climate change*, *environmental change*, and *meteorology*. Although
6 deeply intertwined, these three paradigms should not be confused. *Climate change* alone is the
7 object of the IPCC definition, and is limited to the atmosphere-ocean circulation effects - *i.e.*
8 long-term changes in temperature, precipitation, wind regime, etc. These may be appropriately
9 described and captured through physical investigation methods, such as Atmosphere-Ocean Gen-
10 eral Circulation Models (see §50 p. 130 for details), and are mostly bounded by uncertainties
11 about the nature of physical processes at work, and the socio-economic hypotheses underlying
12 our projections of greenhouse gas emissions. Yet while numerical climate models are increasingly
13 successful at predicting the physical environment of the Earth in the 21st century, the response of
14 biological communities, or «*environmental change*», is a more elusive topic (see §13 p. 55 *sqq.* for
15 an overview), because of the lack of a prior example (although early Holocene climate change is
16 often used as an example, its pace differs significantly from ongoing changes - see Petit *et al.*
17 2008; Blois *et al.* 2013; Moritz & Agudo 2013). The diverse mechanisms through which these
18 changes occur in ecosystems make them particularly difficult to predict. They range from the dis-
19 ruption of developmental processes in lower-trophic-level organisms by higher sea surface tem-
20 peratures (Gregg *et al.* 2003; Behrenfeld *et al.* 2006) or water acidification (Harvey *et al.* 2013),
21 with cascading trophic-web effects (*e.g.* McMahon & Burton 2005), to shifts in species distribu-
22 tion and associated biodiversity loss (*e.g.* in Arctic copepods, see Beaugrand *et al.* 2002). Marine
23 ecosystems, just as terrestrial ones, are already severely affected by human exploitation, which di-
24 minishes their resilience to climate change (see Halpern *et al.* 2008; Watson *et al.* 2013; Mc-
25 Cauley *et al.* 2015). Thus, expected biodiversity collapse under the darker scenarios (see §50 p.

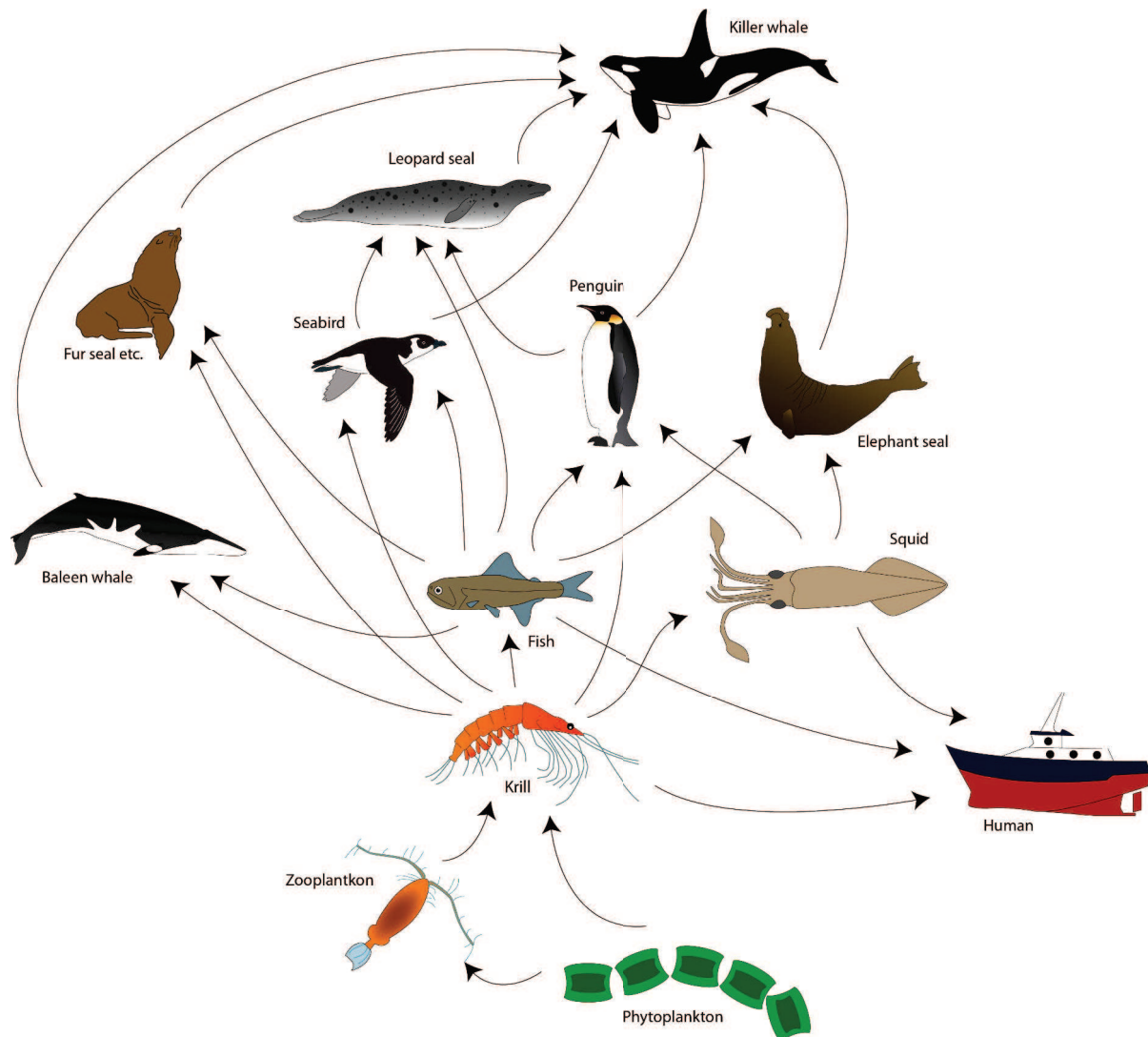
1 130) is projected to affect 50% to 70% of the oceans to a greater extent than all the changes that
2 took place in the Pleistocene and Holocene (Beaugrand *et al.* 2015).

3 Concurrently, a several studies have investigated the effects of climate change on species through
4 the cyclic variability of climate, which is especially prevalent in oceanic systems (see for example
5 §8 p. 40). These studies have often exploited the high contrast between positive and negative
6 phases of the El Niño Southern Oscillation (ENSO, *e.g.* Le Bohec *et al.* 2008), the Southern An-
7 nular Mode (SAM, *e.g.* Weimerskirch *et al.* 2012), or the North Atlantic Oscillation (NAO, *e.g.*
8 Frederiksen *et al.* 2004). While this is often a reasonable approximation - and importantly the
9 only proxy available anyway - these *meteorological changes* should not be fully equated to *climate*
10 *change*. *Weather*, which is mostly at stake there, is only one interface between climate change and
11 species. Warm phases of the ENSO may accurately reproduce warm-earth conditions in the
12 Southern Ocean, but not additional coupling mechanisms with, *e.g.*, the Northern Hemisphere -
13 despite the fact that Arctic-Antarctic couplings have been shown to be central to climate change
14 *stricto sensu* (Barbante *et al.* 2006). While the approximation is often a methodological necessity,
15 *meteorological change* should therefore not be confused with *climate change*.

16 *§-12 Particularities of climate change in the polar regions.* The effects of climate change are especial-
17 ly visible in the polar regions, a fact due to their climatological characteristics, as well as to the
18 structure of their biological communities. From a purely climatological (or «*abiotic*», see §49 p.
19 127) perspective, the polar regions stand out as distinct from the rest of the planet. The singulari-
20 ties of Ocean-Atmosphere circulation in the Arctic, in particular the threshold change in albedo
21 with the presence or absence of sea ice in the Arctic Ocean, leads to an «Arctic amplification »
22 phenomenon of accelerated climate change (Serreze & Francis 2006; Pithan & Mauritsen 2014).
23 Antarctica, on the other hand, exhibits rather mixed patterns: while Eastern Antarctica seems to
24 «resist» climate change (possibly due to the buffering effect of the Southern Ocean, see Mulvaney
25 *et al.* 2012 - although observation uncertainties may be hiding the true trend, see Hanna *et al.*
26 2013), West Antarctica and the Antarctic Peninsula are amongst the fastest-warming areas of the

Introduction - §12

1 Earth (Mulvaney *et al.* 2012; Bromwich *et al.* 2013). In the Southern Ocean, accelerated
2 warming both of the surface waters (Liu & Curry 2010) and of the deep water masses (Purkey &
3 Johnson 2010) has been observed, and the effects of the influx of fresh meltwater on upwelling
4 activity have already been documented (*e.g.* de Lavergne *et al.* 2014). The fundamental global
5 connections of the Southern Ocean also make it especially sensitive to changes in tropical circula-
6 tion patterns, such as the ENSO (see §8 p. 40 and Collins *et al.* 2010). Therefore, a generally in-
7 creased effect of climate warming is expected.



8 **Figure 6 | The Food web of the Southern Ocean.** A simplified structure of the food web of the Southern
9 Ocean. Arrows figure p redation l links. Th e di versity of zo oplankton an d fis h str ategies is col lapsed in our
10 representation.

Introduction - §12

1 However, the biogeography of the Southern Ocean (see §10 p. 48) also contributes to aggravated
2 effects of climate change. While habitats are generally expected to shift poleward in response to
3 climate warming (see Williams *et al.* 2007; Fraser *et al.* 2012, and §15 p. 57), the polar ecosys-
4 tems have little margin for further displacement, and the high latitudes are expected to become
5 «climate sink areas» (Burrows *et al.* 2014) where habitats contract and ultimately disappear,
6 rather than shift (see §15 p. 57). Further complexity arises from the fact that human activity has
7 heavily influenced the Southern Ocean from the bottom of the food chain (*i.e.* through anthro-
8 pogenic climate change and reduction in primary production) as well as from the top (*i.e.*
9 through the massacre of top-level predators such as whales, seals and penguins - see Ainley *et al.*
10 2007), which results in an intricate combination of effects (Smetacek & Nicol 2005 - see also a
11 spectacular and concrete example of this intricate system in Trivelpiece *et al.* 2011).

12 As a result, the seemingly pristine polar regions have largely come to concentrate the effects of cli-
13 mate change. For a long time, it has been thought that polar ecosystems were simpler, and with
14 shorter trophic chains, than lower-latitude systems, thus leading to more easily understandable
15 shifts. However, it is now recognised that both taxonomic diversity and trophic network com-
16 plexity is just as high in the high latitudes than in the temperate regions (Smetacek & Nicol
17 2005; Chown *et al.* 2015 and Fig. 6 p. 53), and it has been suggested that the apparent specifici-
18 ties of the polar oceans are due more to the delayed human impact due to sheer inaccessibility,
19 than to a unique character favouring, *e.g.*, marine macrofauna: indeed, Smetacek & Nicol 2005
20 have proposed that these regions rather represent a vestigium of the pre-historical global marine
21 ecosystems - in their words, a «*Serengetis of the sea*». Thus, we may be currently witnessing in the
22 high latitudes what happened already long ago closer to the equator and what is also happening
23 in continental Africa, with the demise of mammalian macrofauna and its replacement by ubiqui-
24 tous human exploitation. Whether climate change is really particular *per se* in the polar regions
25 beyond its physically increased pace, or whether the nature of its effects on biotic communities is

1 rather a consequence of the relative integrity of these communities, is a question that may be ur-
2 gent to answer.

3 *§-13 Species' responses to climate change.* The importance of the impact of current global
4 changes on the world's biological communities is now recognised beyond question by the sci-
5 entific community (see § 11 p. 50 and Thomas *et al.* 2004; Warren *et al.* 2013; Pacifici *et al.*
6 2015), although the detail of its impact is complex and debated (Pereira *et al.* 2010; He &
7 Hubbell 2011; Pacifici *et al.* 2015). The different levels of impact (ranging from physiology to
8 community level, see §11 p. 50), the complexity of biotic interactions (Davis *et al.* 1998; Nor-
9 berg *et al.* 2012; Midgley & Bond 2015), and the idiosyncratic character of each species' sensitiv-
10 ity and exposure to climate change (Moritz & Agudo 2013; Comte *et al.* 2014; Dickinson *et al.*
11 2014) contribute to apparently heterogeneous responses at the ecosystem level (Parmesan & Yohe
12 2003). Most studies, however, agree on the idea that three fundamental mechanisms interact in
13 determining response to climate change at the species level: (i) *adaptive microevolution* through
14 fitness selection (Bradshaw & Holzapfel 2006; Hoffmann & Sgrò 2011), (ii) *phenotypic plasticity*
15 and in particular changes in phenology (Charmantier *et al.* 2008; Hoffmann & Sgrò 2011), and
16 (iii) *range shift* to track habitat isolines (Walther *et al.* 2002; Chen *et al.* 2011; VanDerWal *et al.*
17 2013).

18 *§-14 Micro-evolution and phenotypic plasticity.* Micro-evolution and phenotypic plasticity share a
19 common structure: in order to accommodate its changing environment, a species changes its
20 physiological, phenological, or generally behavioural characteristics (Hoffmann & Sgrò 2011).
21 The mechanisms involved, however, are very different. Micro-evolutionary response happens on a
22 multi-generation scale, through selection of the fittest heritable features (either genetic or epige-
23 netic): thus, it is a population-scale process. Phenotypic plasticity, on the other hand, is an indi-
24 vidual-based phenomenon. Either through developmental, or through purely behavioural modifi-
25 cations, each individual optimises its fitness under the prevailing conditions - but its new state is

1 not heritable, if we let aside the comparatively rare cultural transmission of behaviour (White-
2 head & Rendell 2014).

3 Evolutionary responses to anthropogenic climate change have already been evidence in several
4 very short-lived species, such as mosquitoes (Gienapp *et al.* 2008; Hoffmann & Sgrò 2011).
5 However, all studies conducted in vertebrates have up to now failed to detect any clearly evolution-
6 nary process (see Møller *et al.* 2010 for a complete review of the subject in birds). The mismatch
7 between the very fast pace of climate change, the longer generation time of vertebrates, and their
8 generally K-like reproductive strategies¹, makes true evolution difficult to achieve.

9 Plasticity, on the other hand, has been shown to play an important part in several species' short-
10 term response (Charmantier *et al.* 2008; Hoffmann & Sgrò 2011). It is primarily known to affect
11 phenology in order to match the meteorological conditions and most importantly the peak of
12 nutrient availability, as has been evidenced in birds such as the Great tit (*Parus major* Char-
13 mantier *et al.* 2008; Husby *et al.* 2010) or the Collared flycatcher (*Ficedula hypoleuca*, Dunn *et al.*
14 2010), but also in plants (Anderson *et al.* 2012) - but it may also apply to metabolic traits (McK-
15 echnie *et al.* 2006; Nicotra *et al.* 2010) or morphology (Przybylo *et al.* 2000). The highly dynam-
16 ic nature of the Southern Ocean system makes phenology a crucial condition of survival for most
17 species. The seasonal temperature cycle brings about a latitudinal variation in the position of the
18 Polar Front (see §7 p. 35), thus potentially changing the distance between foraging grounds and
19 breeding locations along the breeding season for several species (*e.g.* Massom *et al.* 2009; Bost *et*
20 *al.* 2015), or drives the strength of the upwelling activity, and thus the productivity, in the Hum-
21 boldt Current System (Thiel *et al.* 2007). Thus, just as the timing of the insect abundance peak is
22 a critical factor for several bird species (Dunn *et al.* 2010), the seasonal variation of foraging areas
23 in oceanic systems may be a major constraint on their phenology, and its evolution is expected to

1. Classically, K and r reproductive strategies are opposed: K-strategists have fewer offsprings, but a very high parental investment, and consequently high offspring survival, while r-strategists have numerous offspring with minimal investment and low survival.

1 force a plastic shift in the phenology of several species. Such shifts, however, have a high fitness
2 cost (Van Buskirk & Steiner 2009), as they increase the likelihood of a mismatch between differ-
3 ent ecological constraints (Both *et al.* 2006), and may therefore rapidly reach their limits (Visser
4 & Both 2005; Both *et al.* 2006; Møller *et al.* 2008).

5 Microevolutionary response, since it relies on inter-generational selection in pre-existing varia-
6 tion, is a longer-time process. It has been evidenced for several short-lived organisms (Van Heer-
7 waarden & Hoffmann 2007; Hoffmann & Willi 2008; Hoffmann & Sgrò 2011; Krehenwinkel
8 *et al.* 2015), but the extreme velocity of current changes (Burrows *et al.* 2011; Mahlstein *et al.*
9 2013; Poloczanska *et al.* 2013) is a particular challenge for species that have long generation
10 times (Hoffmann & Sgrò 2011) and low genetic diversity (Hoffmann & Sgrò 2011; Norberg *et*
11 *al.* 2012), all the more as these two traits tend to be highly correlated (Romiguier *et al.* 2014).

12 Micro-evolution and plasticity are, however, closely interwoven: only the strength of heritability
13 may allow to distinguish between both (Réale *et al.* 2003), and more importantly, plasticity itself
14 is increasingly recognised as an evolving trait (Nussey *et al.* 2005; Pelletier *et al.* 2007).

15 In this study, however, both the pace of current climate change and the life-history of long-lived
16 seabirds exclude, *a priori*, any significant evolutionary response. A plastic response, on the other
17 hand, is possible, and even to be hoped for. It is however beyond the scope of our work, as cross-
18 generational data is required in order to build a clear picture of such a process. Up to now, evi-
19 dence has been presented in several short-generation-time species (see *e.g.* Charmantier *et al.*
20 2008, etc.) - but several decades of phenological and behavioural observations would be necessary
21 to build a clear picture given the long generation time of most seabirds.

22 *§-15 Habitat displacement and range shift.* Range shift is increasingly recognised as a major feature
23 of short-term response to climate change (Davis & Shaw 2001; Chen *et al.* 2011; Burrows *et al.*
24 2014, etc.). While the main expression of plastic response concentrates on the phenology (see
25 §14 p. 55), that is to say on a *temporal* frame shift in the breeding cycle of the species or popula-
26 tion, range shift is a *spatial* frame shift in the distribution of the organism. As opposed to micro-

1 evolution and plasticity, it exclusively involves a change in geographical, and not temporal, physi-
2 ological or morphological space - although cases of «pure» range shifts, not associated with phe-
3 nological shifts, are scarce (Møller *et al.* 2010). Range shift can be conceived as the expression of
4 different individual- and species-scale factors. On a functional level, *niche conservatism* can be de-
5 fined as the tendency, for an individual, to remain within its physiologically and ecologically de-
6 fined comfort zone (Wiens *et al.* 2010). Depending on the width of that niche, and on the evol-
7 ability of the species, niche conservatism may apply on very long time scales. For example, New
8 World bird taxa exhibit a high degree of evolutionary-level niche conservatism, with the more de-
9 rived species occupying the most specialised environments (Hawkins *et al.* 2006), while similar
10 structures have been identified in stream fish (Comte *et al.* 2014) or land mammals (Martínez
11 Meyer *et al.* 2004). This slow-time niche conservatism means that habitat preference is not a fast-
12 evolving trait, and may be phylogenetically inherited within taxa, slowing down the colonisation
13 of new habitats.

14 On shorter ecological time scales, on the contrary, niche conservatism is the main force that
15 prompts individuals to follow their habitat rather than adapt to new conditions. As such, it
16 results in *habitat tracking*, a movement by which a population will be displaced together with en-
17 vironmental conditions. Habitat tracking, when observed from a human perspective, results in
18 *range shift*, a global modification of the species geographical distribution in response to trends in
19 habitat change. Global warming, for example, typically (and schematically) results in a displace-
20 ment of habitat and species towards the poles, or the higher altitude regions (Jump *et al.* 2009;
21 Fraser *et al.* 2012).

22 In an ideal system, species may be expected to follow habitat displacement smoothly (Scheffer *et*
23 *al.* 2001). However, the anisotropic character of biozones implies that habitats do not transform
24 in a homogeneous manner (Difffenbaugh & Field 2013), and that changes are likely to be doubly
25 heterogeneous - between ecosystems and between species within ecosystems. A recent study on
26 Australian birds (VanDerWal *et al.* 2013), for instance, has shown that during the past 60 years,

1 species' range shifts had followed mostly idiosyncratic trajectories on the whole continent, only
2 partly explained by habitat displacement. Such an example stresses the importance of biotic inter-
3 actions in the determination of a species' range and range displacement, as several interspecific re-
4 lationships can be affected simultaneously by climate change (*e.g.* feeding behaviour, see Char-
5 mantier *et al.* 2008, or and interspecific competition, see Stenseth *et al.* 2015), resulting in non-
6 linear, and sometimes downright paradoxical habitat tracking (*e.g.* in the Barnacle goose *Branta*
7 *leucopsis*, see Bauer *et al.* 2008; Eichhorn *et al.* 2009). Thus, understanding current and future
8 species' distributions may not be properly achieved based solely on single-species (*e.g.* «climate
9 envelope», see Pearson & Dawson 2003) models, but need to be extended to include essential bi-
10 otic interactions (a point further developed in §49 p. 127).

11 The heterogeneity and idiosyncratic character of range shifts is exacerbated whenever discontinu-
12 ities prompt *habitat convergences* (Burrows *et al.* 2014) or «subduction», for instance when possi-
13 ble climatic corridors, where range shift could theoretically happen linearly, are interrupted either
14 by natural, or anthropogenic barriers. Habitat fragmentation, which is often a result of human
15 influence (although it may be natural, in particular in insular environments, see *The King syn-*
16 *nome*, p. 163 and *Unexpected philopatry in an insular seabird, the Peruvian diving-petrel*, p. 275)
17 thus interacts with climate change and has rather unpredictable consequences. This is especially
18 true around “climate sink areas” where local conditions disappear instead of being displaced (Bur-
19 rows *et al.* 2014). Moreover, individual species are in reality not expected to respond linearly to
20 habitat shift: insufficient dispersal possibilities may lead to a lagged range shift (Bertrand *et al.*
21 2011; Schloss *et al.* 2012), increased competition around sink areas may prevent range extension
22 (Norberg *et al.* 2012), and compensatory metabolic processes may mask the effects of climate
23 change until a tipping-point is reached (Doak & Morris 2010). These delayed effects can lead us
24 to underestimate the depth of current changes, while placing species in a state of *climatic debt*, ul-
25 timately giving way to sudden ecosystem collapses (Scheffer *et al.* 2001; Hoegh-Guldberg &
26 Bruno 2010; Barnosky *et al.* 2012).

1 Thus, the polar regions, which undergo accelerated environmental change (Serreze & Francis
 2 2006; Turner *et al.* 2014; Chown *et al.* 2015) and have an extremely limited margin for habitat
 3 shift (Burrows *et al.* 2014), are an extreme example of problematic habitat displacements
 4 (Williams *et al.* 2007). Detecting lagged response to ongoing environmental changes before the
 5 tipping-point is therefore critical if we are to develop appropriate and efficient management and
 6 conservation strategies and to accurately forecast the actual distribution of future critical refugial
 7 areas (Keppel *et al.* 2012; Hope *et al.* 2013; Watson *et al.* 2013), in particular for top-level preda-
 8 tors that largely determine ecosystem resilience (Soulé *et al.* 2003; Heithaus *et al.* 2008) in order
 9 to develop appropriate and efficient management and conservation strategies. Thus, range shift
 10 predictions are becoming a major challenge in ecological modelling.

11 *Aims of this work*

12 *§-16 The scales of climate change.* One of the most challenging aspects of the biology of climate
 13 change is the maze of scales on which it operates. Extant studies often oscillate between two ex-
 14 tremes: the very *slow* time, in which geological climate change (such as the Pleistocene-Holocene
 15 transition) is taken as a proxy for anthropogenic events, or the very *short* time, in which meteoro-
 16 logical stochasticity and cyclic phenomena (for example the North Atlantic Oscillation) are used
 17 to the same purpose¹. In that context, one of the difficulties of understanding contemporary cli-

1. An opposition theorised by Fernand Braudel in his study of the *History of the Mediterranean*, as the distinction between the «*longue durée*» (in which the interaction of humans with their geography and habitat prevail) and the «*short, fast and nervous oscillations*» of «*event history*» (Braudel 1949; Braudel 1969). Along this spectrum, Braudel stresses the «*increasingly precise notion of the multiplicity of time, and of the exceptional value of long time*» (Braudel 1958). Thus, «*every historical work decomposes past time, chooses between its chronological realities, according to more or less conscious preferences and exclusions. Traditional history, which is focused on the swift time, the individual, the event, has long since accustomed us to its precipitated, dramatic, gasping narrative. The newer economic and social history emphasises its search for cyclic oscillations and has tuned itself to them: it became entangled in the illusion, but also in the reality of the cyclical raises and falls of prices. Thus there is, next to the narrative (...), a recitation of the conjuncture that examines the past by thick slices: decades, scores and fifties. But well beyond that second recitation lies a history of a steadier breath still, and now of centennial scale: the long, longue durée history*» (Braudel 1958). This essential decomposition of history into layers of causes and events - structural, conjunctural and proximal - is now a tenet of modern history. Yet, it is hardly theorised in biological sciences, where it only lives in the often unreconciled opposition of evolutionary and ecological paradigms (§1 p. 21).

Introduction - §16

1 mate change is that it operates on a scale for which we have little or no concepts, because it has
2 no known precedent in the history of the earth, and that requires us to find a middle ground be-
3 tween ecological and evolutionary times. And so, a central question is whether the observations
4 we extract from very long term (*i.e. palaeodemographic*) studies are of the same *nature* as the ones
5 we extract from short term (typically *monitoring*) studies - and whether any of these may explain
6 and predict the events to be expected under anthropogenic climate change. In many respects, it
7 may appear that the nearly instantaneous observations that we make on monitored populations,
8 and that span a few generations at best, are dominated by several layers of stochasticity and mixed
9 forcings - the stochasticity and cycles of meteorology around the trends of climate and its nested
10 cycles, the stochasticity of demography around the cycles and trajectories of populations and
11 species, the slow but chaotic restructuring of ecosystems when its parts move in different direc-
12 tions - and that these contradictory signals may hide for some time the inevitably directional
13 trends of global change. At the other end of the spectrum, however, palæodemography, by its
14 evening out of this very stochasticity and these high-frequency variations, retains only broad pat-
15 terns of change, such as the slow pulse of Pleistocene glaciations, that leave time for evolution to
16 occur. It may, however, totally fail to capture the now comparatively swift movement of man-
17 made change.

18 For these reasons, what really occurs in a situation such as contemporary anthropogenic climate
19 change can be conceived as a disruption of the pace and manner in which these nested scales are
20 integrated, and an increasingly likely mismatch between layers of spatial and temporal variation.
21 The duration of the generation, for example, may be simplistically considered as the time lapse
22 over which an individual integrates environmental conditions and undergoes selection: it is long
23 enough to average over the seasons, and sometimes the oceanic modes, but fast enough to let a
24 consequential series of generations to adapt to geological changes. A new, intermediate pace of
25 change may disrupt this timing, and be too fast for evolution to keep up. The organisation in
26 space of individuals and species reflects the extent over which conditions make them most fit (see

1 §49 p. 127), and the size and mobility of individuals evolved to match this space: there again, the
2 convergence and divergence of habitats disrupts this equilibrium. Thus, in many respects, it ap-
3 pears that the key to understanding contemporary climate change is understanding how the
4 different chronological and spatial paradigms articulate themselves for ecosystems, for species,
5 and for individuals.

6 *§-17 Understanding the scales of organisation in seabirds.* The primary focus of this work is the un-
7 derstanding of the different spatial and temporal scales in the demographic organisation and dy-
8 namics of seabirds - a group at the forefront of climate change, as presented in §18 p. 65 - using
9 three contrasting species as models: the King penguin *Aptenodytes patagonicus*, the Emperor pen-
10 guin *Aptenodytes forsteri*, and the Peruvian diving-petrel *Pelecanoides garnotii*. While set primarily
11 from an evolutionary point of view, we explore the results provided by a broad set of population
12 genetic methods in the light of the insights offered into different scales by alternative approaches,
13 including bio-logging, habitat modelling, and direct field monitoring techniques. For each of the
14 two *Aptenodytes* penguin species, we present two complementary studies: one focusing on world-
15 wide, evolutionary-scaled patterns and based on genome-wide datasets (*The King synnome*, p. 163
16 and *The Emperor synnome*, p. 211), and one focusing on local, «*chronicle-scaled*»¹ phenomena
17 (*Fine-scale colony structure*, p. 135 and *Empirical evidence of heterogeneous dispersal*, p. 259). These
18 two approaches are combined in the more exploratory study of the Peruvian diving-petrel (*Unex-
19 pected philopatry in an insular seabird, the Peruvian diving-petrel*, p. 275).

20 (i) The comparative biogeography of these three species is the foundation of this study. A first
21 task will be to understand how range-wide distributions are organised, at scales ranging from in-
22 tra-colonial diversity to island and range-wide levels, and to establish what seems to be the rele-
23 vant demographic unit for studying past, current and projected demographic trends. This will

1. We borrow this term from Fernand Braudel (see above, §16 p. 60), as it accurately reflects the framework in which our monitoring data operates: like medieval *chronicles*, monitoring approaches are based primarily on recording individual birth, deaths, lineages, and memorable events, and are by definition set to the generation time of the focal model, by opposition to *histories*, scaled to a people - that is, a *population*.

1 also involve reviewing current estimates of population size, structure, and, in the case of the King
2 penguin, taxonomic diversity.

3 (ii) Predictions regarding the fate of these three species in the face of environmental change are
4 rather grim. The Emperor penguin has recently been declared doomed throughout its range (Je-
5 nouvrier *et al.* 2014), the King penguin received dire forecasts for its best-studied colony (Le Bo-
6 hec *et al.* 2008; Péron *et al.* 2012), and the Peruvian diving-petrel achieved the *Endangered* status
7 since 1994 update of the International Union for Conservation of Nature (IUCN) Red List¹.
8 Nevertheless, achieving a better understanding of seabird population dynamics requires a critical
9 examination of these predictions, and an integration of the methods that prompted them. We ad-
10 dress this issue using the total available evidence for the best known of these three species, the
11 King penguin, and propose a possible re-interpretation of the Emperor penguin's response to
12 Southern Ocean warming. In both cases, we try to propose a likely consensus projection based
13 both on *long-* and *short-time* observations, as a tentative framework to solve the difficult prob-
14 lem of ecological and evolutionary integration in the context of climate change.

1. As will be exposed in details in **Concepts and methods** (p. 65), different reasons prompted these three projections: the observed dependency on sea ice, and the modelled evolution of that same sea ice for the Emperor penguin, the correlation between foraging area and sea surface temperature for the King penguin, and the observed and partly unexplained (but likely only marginally climate-related) population decline for the Peruvian diving-petrel.

1 Chapter 2: Concepts and methods

2 *A brief introduction to Seabirds.*

3 *§-18 Seabirds, and Penguins in particular, as sentinels for polar ecosystems.* The effects of climate
4 change (§13 p. 55) are but incipient, and, although they likely already involve all species and
5 ecosystems on the planet (Barnosky *et al.* 2012), they are most clearly observable only in a subset
6 of organisms that integrate and amplify these changes. Since several confounding factors may
7 blur our understanding of the ongoing changes (see §13 p. 55), an accurate assessment of the
8 true processes at work requires the selection of study models that integrate those changes across
9 trophic levels and geographical areas in a significant way. Due to the particular speed and mo-
10 ment of climate change in the polar regions (see §12 p. 52), polar or sub-polar ecosystems are of-
11 ten used to that purpose: and within these, top-predator species are commonly chosen. Indeed,
12 the structure of the trophic networks places apex predators in a particular position. Being down-
13 stream of the nutrient and energy flow, external forcing on the biomass of every trophic level
14 (from the primary phytoplankton production to the upper fish or krill biomass) will ultimately
15 affect the availability of their resource (Bakun 2006; Frederiksen *et al.* 2006; Hunt & McKinnell
16 2006; Ainley *et al.* 2007). For example, a decrease in upwelling activity south of the Polar Front
17 (see §7 p. 35) will result in a reduced phytoplankton growth, and ultimately in a decrease of
18 available prey biomass for top predators (in a classical «bottom-up» forcing scenario, Frederiksen
19 *et al.* 2006). Yet, considering a constant primary productivity, a specific forcing, such as increased

1 ocean acidity, may reduce the growth of an intermediate organism, such as krill - also resulting in
2 a decreased prey abundance for top predators (in a so-called «wasp-waist» forcing, see Bakun
3 2006). In parallel, other constraints may converge on top-predators: bio-accumulation, for exam-
4 ple, increases along the trophic levels, with anthropogenic contaminants reaching the highest
5 concentrations in long-lived apex predators.

6 Amongst top predators, most seabird species share some additional characteristics that make
7 them particularly suitable to monitor environmental changes. As opposed to top-predator fish or
8 shark species, and to cetaceans, that never touch land, seabirds and pinnipeds require firm
9 ground (either land or ice) to breed, and usually exhibit a high degree of philopatry (see §6 p. 33
10 and §19 p. 68), which makes it possible to study individuals over long time periods, and to
11 measure critical parameters, such as individual breeding success - a possibility made even more
12 interesting by the long life span of most species (up to ~45 years for large penguins, see Bost *et al.*
13 2013, or more for several procellariiformes, and 15 to 40 years for most pinnipeds, see Shirihai &
14 Kirwan 2008). However, seabirds have over pinnipeds the advantage of often very large popula-
15 tion sizes, normally bi-parental offspring care, and typically central-place foraging, which makes
16 them especially suitable for long-term monitoring programs (Cairns 1988; Croxall *et al.* 2002;
17 Parsons *et al.* 2008; Grémillet & Charmantier 2010; Cury *et al.* 2011).

18 Amongst seabirds, the evolutionary history of spheniscids (see §21 p. 72) makes them a perfect
19 model for understanding the changes in the Southern Ocean. Constrained to tight latitudinal
20 zones, most penguin species are specialised in a narrow set of preys (*e.g.* squid for the Emperor
21 penguin *Aptenodytes forsteri* - see §23 p. 78, or myctophid fish for the King penguin *Aptenodytes*
22 *patagonicus*, see §25 p. 80) which reduces their short-term plastic response to modifications in
23 the trophic network. But specific characteristics of the King and Emperor penguins are of partic-
24 ular interest. The Emperor penguin breeds during winter on the sea ice (see §23 p. 78 for details),
25 and, during that period, it relies on coastal *polynyas* for foraging. These coastal open-water areas
26 may occasionally be caused by warmer deep-water upwellings (they are then called *sensible heat*

1 *polynyas*, e.g. the Weddell sea polynya, see Thomas & Dieckmann 2008 and §7 p. 35), but are
2 mostly the result of continental wind circulation patterns. The very large temperature and
3 pressure difference between the extremely cold and high-altitude central Antarctic ice sheet, and
4 the comparatively mild coastal zone, result in strong coastwards *katabatic* winds, that follow the
5 topography of the ice sheet. These winds cause offshore drift of the coastal sea ice, and maintain
6 ice-free areas at the breaking point between the grounded ice and the floating ice, before dying
7 off a few miles from the coast - at the same time, the offshore drift of the ice draws a local up-
8 welling of deeper shelf waters, resulting in increased marine productivity (see Fig. 4, and Ancel *et*
9 *al.* 1999; Thomas & Dieckmann 2008). These *latent heat* polynyas are essential for the survival
10 and breeding success of Emperor penguins, and can be affected either locally by icescape modifi-
11 cations (Ancel *et al.* 2014; Kooyman & Ponganis 2014), or globally by changes in the atmospher-
12 ic circulation patterns, and reduction in katabatic wind activity.

13 The King penguin, on the other hand, relies for breeding-season foraging on the myctophid fish
14 stock of the Antarctic Polar front (see §7 p. 35, §25 p. 80, and *The King synnome*, p. 163). The
15 Antarctic Polar front position is directly determined by the strength of Westerly wind circulation
16 and resulting Ekman transport, and by sea surface temperature in the Antarctic and Subantarctic
17 zones (see §7 p. 35 and Fig. 2): higher temperatures and weaker Westerlies result in a increased
18 mean latitude for the Polar Front (as is the case during negative Southern Annular Mode events,
19 see §8 p. 40). Since the insular breeding locations of the King penguin, on the other hand, are
20 fixed (see §25 p. 80, and *The King synnome*, p. 163), the foraging distance of the species varies
21 with the latitude of the frontal zone. Thus, King penguins thrive when Westerlies are strong and
22 surface temperature cold south of the Polar Front.

23 And so, *Aptenodytes* penguins share this fascinating characteristic that their survival, breeding suc-
24 cess, and overall fitness are directly tied to the planetary atmospheric and oceanic circulation,
25 rather than to local, finer scale proximal causes. While the sensitivity of e.g. Great tits to climate
26 change happens through several mediations - the local climate change forced by global evolu-

1 tions, the combined effects of temperature and precipitation of the abundance of insect larvæ, etc
2 (Charmantier *et al.* 2008), the relationship of large penguins to climate change is, so to speak,
3 immediate: increased radiative forcing results in changes in the planetary wind circulation, and
4 drives the King penguin foraging areas further South, or decreases polynya availability for the
5 Emperor penguin. This characteristic has been exploited in both species to predict demographic
6 evolution in relation to global changes (Le Bohec *et al.* 2008; Jenouvrier *et al.* 2014; Bost *et al.*
7 2015).

8 Other seabird species have been monitored in a similar way, for example the Wandering albatross
9 *Diomedea exulans* (e.g. Weimerskirch *et al.* 2012), the Adélie penguin *Pygoscelis adeliae* (e.g. Bal-
10 lard *et al.* 2010), or the Kittiwake *Rissa tridactyla* (e.g. Moe *et al.* 2009). Generally, we may con-
11 sider that the value of a species as a bio-indicator is linked to the extent of its foraging areas (that
12 may be the whole Southern Ocean, as for the Wandering Albatros, see Croxall *et al.* 2005) and
13 breadth of its trophic basis: whereas colonies of large penguins integrate large swathes of South-
14 ern Ocean, some locally-foraging, highly philopatric species may rather be indicative of finer-
15 scale changes. This is for example the case of the Peruvian Diving-petrel *Pelecanoides garnotii*, a
16 small insular procellariid dependent on the coastal upwellings of the Humboldt Current System
17 along the Pacific coast of South America, and mostly subject to the local variations in the El
18 Niño Southern Oscillation (see §8 p. 40, and *Unexpected philopatry in an insular seabird, the Pe-*
19 *ruvian diving-petrel*, p. 275). Thus, whether they integrate large-scale, planetary processes, or
20 more local tendencies of marine ecosystems, the sentinel role of seabirds makes the study of their
21 demography, life history, and population dynamics a priority for understanding the ongoing
22 changes in the earth's climate.

23 *§-19 Philopatry and dispersal amongst seabirds.* Seabirds are, in many respects, intermediate species
24 between oceanic and terrestrial ecosystems (see §18 p. 65). While their foraging grounds are ma-
25 rine, and are subject to all the particularities of oceanic systems, their breeding grounds are
26 terrestrial, and usually extremely fragmented (being insular, or coastal with very selective charac-

1 teristics). Thus, their ability to cover very large areas during their foraging trips is in contrast with
 2 the scarcity of available breeding habitat, and a strong tendency to return to the same location to
 3 breed year after year - although most of species have the ability to disperse well beyond the bor-
 4 ders of their local grounds, and most have been observed to do so (e.g. Croxall *et al.* 2005; Milot
 5 *et al.* 2008, etc.). As a result, genetic population structure has been evidenced in most species (see
 6 Friesen *et al.* 2007), and often results in speciation (see Friesen 2015).

7 Thus, patterns of population structure are a result of the equilibrium established between disper-
 8 sal and philopatry - and penguins are a particularly striking example of this equilibrium. Several
 9 studies have concluded in a high degree of philopatry in different species: in the King penguin,
 10 juvenile return rate is typically very high (up to 87% on Crozet archipelago, Saraux *et al.* 2011a).
 11 In the Emperor penguin, although estimated from flipper-banded individuals and therefore po-
 12 tentially underestimated (Le Maho *et al.* 2011), juvenile and adult return rate has also been
 13 shown to be important (Jenouvrier *et al.* 2012), and Adélie penguin *Pygoscelis adeliae* juveniles
 14 have a similarly high return rate (45% to 75% recruitment on their natal colony, Cornet *et al. in*
 15 *prep*). Thus, a natural expectation would be for populations to undergo significant reproductive
 16 isolation, and genetic drift. However, surveys have been conducted in most species, either at a lo-
 17 cal or a global scale, with contrasting results. Several species, such as the Magellanic penguin
 18 *Spheniscus magellanicus* (Bouzat *et al.* 2009), the Adélie penguin *Pygoscelis adeliae* (Roeder *et al.*
 19 2001), the Chinstrap penguin *Pygoscelis antarcticus* (Freer *et al.* 2015), have been shown to exhib-
 20 it very weak or absent isolation patterns amongst colonies. On the other hand, the Little penguin
 21 *Eudyptula minor* (Overeem *et al.* 2008; Peucker *et al.* 2009) or the Yellow-eyed penguin
 22 *Megadyptes antipodes* (Boessenkool *et al.* 2009a; Boessenkool *et al.* 2009b) show significant isola-
 23 tion patterns, while the Gentoo penguin *Pygoscelis papua* (De Dinechin *et al.* 2012), and the
 24 Rockhopper penguin *Eudyptes chrysocome* (Banks *et al.* 2006; Jouventin *et al.* 2006; De Dinechin
 25 *et al.* 2008) are thought to be engaged in a speciation process between isolated oceanic provinces.
 26 While morphological evidence points towards a certain degree of homogeneity in the Emperor

1 penguin, the King penguin's alleged subspecies structure (Mathews 1911; Barrat 1976 and §21 p.
2 72), points to much reduced gene flow - a conclusion supported by an early study (Viot *et al.*
3 1993). While the heterogeneity of methods and sampling designs used in these different studies
4 prevents a systematic comparison of population structures, it appears that genetic homogeneity is
5 higher in the pelagic-foraging species such as the Adélie and the Chinstrap, and that population
6 isolation is characteristic of more coastal species such as the Gentoo or the Rockhopper.

7 However, the genetic structure of populations is not a direct reflection of their philopatric behav-
8 iour, as it also integrates other traits such as generation time, substitution rate¹, the extent of ran-
9 dom mating within colonies, etc. And dispersal itself (§4 p. 30) is hardly a directly observable pa-
10 rameter (§5 p. 32) in wild populations. Most surveys of seabird demography are technically
11 limited to one or a few colonies (Frederiksen & Petersen 1999; Barbraud & Weimerskirch 2001a;
12 Gendner *et al.* 2005; Ponchon *et al.* 2015), and monitoring relies on the identification and recap-
13 ture of marked individuals (commonly through the use of rings, or flipper-bands for penguins -
14 or radio-identification tags, see §26 p. 81). Thus, while return rate is a direct observation (once
15 tag loss is accounted for, see *e.g.* Olsson & van der Jeugd 2002), the only alternative observation
16 is non-return - that is, either dispersal, or death. Distinguishing between these would ideally re-
17 quire that the marked individuals are recaptured elsewhere than on the studied colony - a
18 methodological aporia at the present time. Some explicit modelling approaches have been devel-
19 oped for scalable systems where local dispersal can be used as an estimator for long-distance dis-
20 persal (see *e.g.* Cooper *et al.* 2008), but seabirds oppose this essential limitation that dispersal is
21 either long-distance, or null - since intra-colony movement is inherently of a different nature
22 than inter-colony movement (Frederiksen & Petersen 1999 and *Fine-scale colony structure*, p.
23 135). This difficulty is made much worse by the fact that several marking techniques have been
24 shown to influence directly the return rate of individuals: this is notoriously the case in penguins,

1. Which can be defined as the observed number of substitutions (*i.e.* changes of DNA base) per single position in the genome, per million years - an estimator of the probability, for a given monomorphic position, to become polymorphic during the next million years.

1 where flipper-banding has been demonstrated to increase adult mortality (Froget *et al.* 1998; Sa-
2 raux *et al.* 2011a). Beyond the obvious ethical question (Le Maho *et al.* 2011), this raises con-
3 cerns as to the informative content of several studies based on recapture of banded penguins (Sa-
4 raux *et al.* 2011a). Thus, a complementary approach to the study of dispersal, not based on
5 individual recapture probability, is needed (see §41 p. 103).

6 §-20 *Taxonomy of Penguins, and their relationship to other seabirds.* Although treated until here as a
7 consistent group, seabirds are a surprisingly diverse ensemble - both in terms of taxonomy, and of
8 adaptation. In the broader sense, seabirds encompass the whole *Æquornithia* group, including
9 *Procellariimorphæ*¹ (*i.e.* petrels, albatrosses, and penguins), *Pelecanimorphæ* (*i.e.* pelicans, herons
10 and cormorants), and *Gaviimorphæ* (*i.e.* loons), as well as the *Charadriiformes* (*i.e.* seagulls, terns,
11 skuas, and plover-like shorebirds), and in the vernacular sense also include a handful *Anseriformes*
12 (*e.g.* the Common eider *Somateria mollissima*) - see Livezey & Zusi 2007; Hackett *et al.* 2008;
13 Jarvis *et al.* 2014. All the species in these groups share some either plesiomorphic, or convergent
14 traits. They rely on marine productivity for food, either directly preying on fish, crustaceans or
15 mollusks, or foraging in the marine debris on the shore. Their plumage evolved advanced insula-
16 tion and waterproofing properties. Their feet evolved interdigital webbing to facilitate paddling
17 or steering. They are not usually territorial in their foraging areas, as opposed to most raptor or
18 passerine species, and most of them breed colonially. However, their exploitation of the marine
19 environment is highly diverse. Some species are mostly restricted to the shoreline (*e.g.* plovers or
20 herons), and most usually occupy only the coastal waters (*e.g.* seagulls, pelicans, cormorans, etc).
21 Few species are truly pelagic, and most of these belong to the *Procellariimorphæ* or «storm-birds»
22 - often called seabirds *stricto sensu*. Paradoxically, within that group, the colonisation of the high
23 seas followed two opposite paths: (*i*) some species, within the *procellariiforme* family, developed

1. Both recent and questionable compounds of Latin and Greek (*aequor*, the «high seas» and *δρνις*, «bird» for the first, and *procella*, «storm», and *μορφή*, «shape» for the second).

1 extraordinary gliding abilities, and started using the strong winds prevailing at sea to dispense
2 with the metabolic costs of active flight - these are the petrels, fulmars, shearwaters, and above all
3 albatrosses, that may circle the entire planet during a foraging voyage (Croxall *et al.* 2005). (ii)
4 Some others, however, developed swimming abilities instead, at the cost of reduction, or complete
5 loss, of flight: these are first and foremost the *Sphenisciformes* (the penguins), the *Alciformes*
6 (the auks, puffins, mures and guillemots), and the *Pelecanoididae* or diving-petrels (a part of the
7 mostly gliding *Procellariiformes*). Morphological resemblances between these three families are
8 due to true convergence (see Grassé 1950 for details), and the most remarkable is the modification
9 of the wings to facilitate underwater swimming: since the resistance of the water is much stronger
10 than that of the air, wings have become shorter, more rigid, and the pectoral muscles have developed
11 accordingly. The trade-off for this adaptation has been the complete loss of flight in modern
12 penguins - as well as in at least one alcid, the now-extinct Great Auk (*Alca impennis*) - and a
13 costly and laborious flapped flight in the remaining Alcidae, and in Pelecanoididae. Posterior displacement
14 of the hind limbs also occurred in all three groups, with an increasingly upright standing
15 posture from Pelecanoididae to Spheniscidae. Finally, outstanding diving abilities have been
16 observed in the three groups: the massive Emperor penguin may reach depths down to 500m
17 (Kooyman & Ponganis 1998), while the intermediate-sized *Uria* guillemots can reach 180m (Pitt
18 att & Nettleship 1985), and the small Peruvian diving-petrel 80m (Zavalaga & Jahncke 1997)¹.
19 These common characteristics set apart the Northern Hemisphere Alcidae, and the Southern
20 Hemisphere Spheniscidae and Pelecanoididae, from the rest of the seabirds. This work will mainly
21 focus on Spheniscidae, and in particular on the two *Aptenodytes* penguin species. The Pelecanoididae,
22 due to their several convergence points with the penguins, will also be used as a reference
23 outgroup in the seventh chapter, *Unexpected philopatry in an insular seabird, the Peruvian diving-*
24 *petrel*, p. 275.

1. Diving abilities being a direct function of myoglobin availability, and hence body mass.

1 §-21 *Origin, evolution and radiation of sphenisciforms*. Although penguins benefit from an out-
2 standingly rich fossil record (Simpson 1975; Myrcha 2006; Clarke *et al.* 2007), their evolutionary
3 history presents considerable uncertainties. The earliest stem penguin radiation is generally
4 thought to have occurred just after (Baker *et al.* 2006; Clarke *et al.* 2007), or just before (Slack *et*
5 *al.* 2006) the Cretaceous-Tertiary boundary, around ~65 Myr, with early flightless species such as
6 the loon-like *Waimanu*. The dating of the crown group radiation, however, is much more uncer-
7 tain. Based on morphological evidence, it was proposed around ~10 Myr (Clarke *et al.* 2007),
8 but molecular evidence has set it back to ~40-50 Myr (Baker *et al.* 2006), or ~20 Myr for Subra-
9 manian *et al.* 2014, while extant species would have mostly diversified around 16-11 Myr. How-
10 ever, a recent total-evidence study including morphological and molecular data for fossil and ex-
11 tant species in a fossilised-birth-death model (Gavryushkina *et al.* 2015) reconcile these two
12 sources of evidence, and imply a very recent radiation of the crown taxa around ~12.45 Myr, and
13 a divergence of most extant species around ~2 Myr.

14 Eocene and Oligocene penguin diversity is dominated by «giant» taxa such as *Icadyptes* from
15 South America or *Kairuku* from New Zealand, which probably originated in Tertiary temperate
16 Gondwanaland (Baker *et al.* 2008). Early apomorphies that allowed for adaptation to cold wa-
17 ters, and in particular the humeral arterial plexus counter-current heat exchange structure
18 (Thomas *et al.* 2011), are thought to have evolved in the Cenozoic greenhouse-earth, at lower lat-
19 itudes, and independently from the current extreme-cold adaptation. Although theory predicts
20 that larger taxa should be better adapted to cold climate due to a favourable body surface-to-vol-
21 ume ratio, only smaller taxa survived through the late Tertiary icehouse earth transition, which
22 lead to suggest that *i*) giant penguin diversification was linked to increased upwelling activity and
23 marine productivity during the Cenozoic (Clarke *et al.* 2007) and that *ii*) their demise was linked
24 to the appearance of competing fish-eating small cetacean taxa (Fordyce *et al.* 1990; Williams
25 1995).

1 Crown penguins are thought to have originated at higher, subantarctic latitudes, and to have ra-
 2 diated back to lower subtropical and tropical latitudes in the recent past (Baker *et al.* 2006). Sev-
 3 eral adaptations to extreme cold have taken place in the recent evolutionary history, both on a
 4 molecular level (diversified keratin β genes, and modified phototransduction and lipid metabo-
 5 lism pathways, see Li *et al.* 2014), and on a behavioural level (e.g. the Emperor penguin's hud-
 6 dling behaviour, see Stonehouse & Glenister 1953). However the precise timing and structure of
 7 their radiation remains uncertain, and hybridisation is still commonly observed between extant
 8 species (e.g. between Humboldt and Magellanic penguins *Spheniscus humboldti* and *S. magellani-*
 9 *cus* - see Simeone *et al.* 2009, or between Royal and Macaroni penguins *Eudyptes schlegeli* and *E.*
 10 *chrysolophus*, *pers. obs.*).

11 The *Aptenodytes* clade has been alternatively proposed as basal to all other crown penguins (Baker
 12 *et al.* 2006; Subramanian *et al.* 2014), or nested within the crown group (Gavryushkina *et al.*
 13 2015). At any rate, the divergence between the Emperor penguin *Aptenodytes forsteri* and the
 14 King penguin *Aptenodytes patagonicus* is thought to be very recent, during the mid-Pleistocene
 15 (1.52 Myr [0.73, 2.38] according to the total-evidence analysis, see Gavryushkina *et al.* 2015), al-
 16 though the place and conditions of this speciation event are still unknown. The King penguin has
 17 traditionally been separated in two subspecies on a morphometrical basis, with the autonym
 18 *Aptenodytes patagonicus patagonicus* in the southern Atlantic ocean, and *Aptenodytes patagonicus*
 19 *halli* elsewhere (Mathews 1911; Barrat 1976), although it is still unclear how far this classification
 20 reflects the true species architecture (see *The King synnyme*, p. 163).

21 The most remarkable features of the known evolutionary history of the penguins are thus that *i*)
 22 the stem group is ancient and separated early from other *neornithes* species (Gavryushkina *et al.*
 23 2015); *ii*) early penguin diversity is not associated with polar or otherwise cold environment, but
 24 rather with greenhouse-earth temperate habitats (Baker *et al.* 2006; Thomas *et al.* 2011); *iii*) the
 25 giant penguin fossil abundance and the co-occurrence of their demise with the radiation of small-

1 er fish-eating cetaceans makes it likely that they occupied a central top-predator role in the Ceno-
 2 zoic marine food-web (Fordyce *et al.* 1990; Williams 1995; Baker *et al.* 2006); *iv*) the adaptation
 3 of crown penguins to extreme cold conditions is secondary, and builds upon pre-adaptations
 4 (Thomas *et al.* 2011), and the colder-adapted species may have taken advantage of the glacial
 5 conditions onset, and of the lower competition in the polar areas, to diversify in the mid-Pleis-
 6 tocene, and *v*) the recent diversification within the major crown penguin groups (*Spheniscus*, *Eu-*
 7 *dyptes*, *Pygoscelis* and *Aptenodytes*) is shallow, and may still allow for interspecific gene flow.

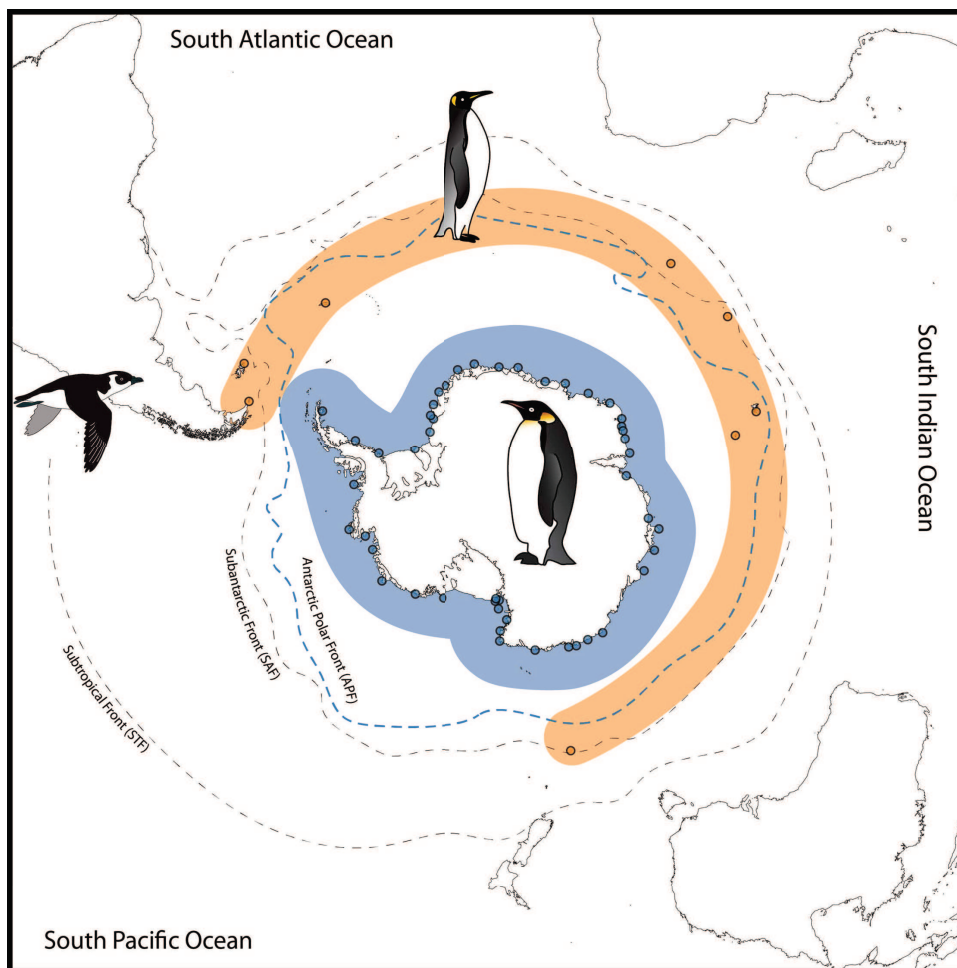
8 §-22 *Aptenodytes penguins in the historical period.* The King penguin (*Aptenodytes patagonicus*)
 9 enters the historical records during Cook's second voyage in the southern seas (1772-75), when it
 10 is sketched by the ship's naturalist Forster in South Georgia, and later described by Miller as
 11 *Aptenodytes patagonicus*. We can hardly estimate how numerous the species was at the time: we
 12 only know that large colonies were described on Macquarie Island, Heard Island, Kerguelen and
 13 Crozet Archipelagos, Marion and Prince Edward Islands, and South Georgia, while smaller pop-
 14 ulations bred in Islas Hornos and Estados in Tierra del Fuego, and possibly in the Falklands (Bar-
 15 rat 1976). Yet, however populous they may have been, King penguins populations were massa-
 16 cred at industrial scales through most of their range by the whaling and sealing expeditions: the
 17 population on Macquarie was thus reduced to *ca* 3,400 birds (~0.7% of its current size) by the
 18 early 20th century (Rounsevell & Copson 1982), while most colonies were driven to near-extinc-
 19 tion throughout the species range (Murphy 1915; Murphy 1936; Barrat 1976; Lewis Smith &
 20 Tallowin 1979; Delord *et al.* 2004; Pistorius *et al.* 2012), and to complete extinction in Heard
 21 and Tierra del Fuego (Barrat 1976; Gales & Pemberton 1988) - although is difficult to ascertain
 22 the global extent of the reduction in population size. Following the end of slaughters in the
 23 mid-20th century, the King penguin population rebounded (Budd 2000; Delord *et al.* 2004; He-
 24 upink *et al.* 2012) at most breeding locations, possibly benefiting from the continued large-scale
 25 hunting of cetaceans and pinnipeds in the Southern Ocean throughout most of the 20th century
 26 (Cressey 2015), and subsequent reduction in competition for food resources (Ainley *et al.* 2010b;

1 Trivelpiece *et al.* 2011). However, populations seemed to have reached the end of their recovery
2 by the early 21st century (Delord *et al.* 2004; Heupink *et al.* 2012), and are now once again chal-
3 lenged by large scale environmental changes in the Southern Ocean (Le Bohec *et al.* 2008; Péron
4 *et al.* 2012; Bost *et al.* 2015).

5 As to the Emperor penguin *Aptenodytes forsteri*, which is endemic to Antarctica, it was first seen
6 by Forster during the same voyage, but was then misidentified as belonging to the freshly dis-
7 covered King penguin species (Wienecke 2010). It was classified as a separate species by the
8 British zoologist G. R. Gray in 1844 (Gray 1844). But it was long before its winter breeding biol-
9 ogy was understood, as the first breeding colony was only discovered in 1902 during the first of
10 Scott's *Discovery* Expeditions (Wienecke 2010), and eggs during his fatal 1911 *Terra Nova* expe-
11 dition, as romantically put by Cherry-Garrard - « *if you march your winter journeys you will have*
12 *your reward, so long as all you want is a penguin's egg* » (Cherry-Garrard 1922). Eggs were also col-
13 lected for study during Shackleton's last expedition (1914-17, as narrated in Shackleton 1919).
14 Discovery of breeding colonies was slow and difficult, owing to the technical difficulties of winter
15 expeditions in the most remote areas of Antarctica (Wienecke 2010), and has been substantially
16 helped by satellite exploration (Fretwell & Trathan 2009; Fretwell *et al.* 2012; LaRue *et al.* 2015).
17 However, it is believed that several colonies are still undiscovered (Ancel *et al.* 2014). In contrast
18 with the large-scale massacres of king penguins, the direct impact of humans on Emperor pen-
19 guins has been relatively small, a fact owing unfortunately more to the remoteness of its habitat
20 than to any particular benevolence towards this remarkable species. However, an impact of re-
21 search activities, in particular flipper banding, has been suggested on studied colonies (Le Maho
22 *et al.* 2011), and, more importantly, the ongoing anthropogenic climate change is likely to have a
23 tremendous impact on the near future of this species (Trathan *et al.* 2011; Jenouvrier *et al.* 2014,
24 see §24 p. 79 for details).

1 *Study sites and species*

2 Although this work encompasses world-wide patterns of population history and geography, we
 3 have taken two particular areas as focal points: the Pointe Géologie Emperor penguin colony,
 4 near Dumont d'Urville research station, in Terre Adélie, East Antarctica (S66°39'46"
 5 E140°0'07"), and the Baie du Marin King penguin colony, on Possession Island, Crozet Archi-
 6 pelago, in the Southern Indian Ocean (S46°24'41" E51°45'22") - see Fig. 7. These two colonies
 7 have been the focus of long-term monitoring programs (Barbraud & Weimerskirch 2001a;
 8 Gendner *et al.* 2005), and their demography, although still not fully understood, is fairly well
 9 known compared to other penguin species.



10 **Figure 7 | Distribution of the two *Aptenodytes* species.** Orange dots: King penguin breeding archipela-
 11 gos, blue dots: Emperor penguin breeding colonies. Orange and Blue areas, general range of the species (schematic).
 12 Antarctic Polar Front, Subantarctic convergence and Subtropical convergence, approximate mean position (Austral-
 13 ian Antarctic Division dataset, after Orsi *et al.* 1995).

1 §-23 *The Emperor penguin*. The Emperor penguin (*Aptenodytes forsteri*) is the largest extant pen-
2 guin species, measuring up to 120 cm and weighting up to 45 kg (Stonehouse & Glenister
3 1953). It breeds during winter along the shoreline of mainland Antarctica, at latitudes comprised
4 between 64°S and 77°S (Wilson 1983). All across this range, it reproduces in well-separated
5 colonies that gather on the sea-ice at the beginning of the breeding season, in the first days of
6 winter, and that dissolve again completely during the austral summer, as the chicks fledge and
7 their breeding ground breaks up, only to gather again at the beginning of the next winter. The
8 size of these colonies usually ranges from a few hundred, to dozens of thousands of breeding pairs
9 (Fretwell *et al.* 2012). Despite their cyclical existence, however, each colony is considered to be a
10 very stable unit, both in terms of location, and of individual identity: the Emperor penguin is
11 thought to be a strongly philopatric species, both as a juvenile, and as an adult (Stonehouse &
12 Glenister 1953; Budd 1961; Prévost 1961; Budd 1962; Isenmann & Jouventin 1970; Isenmann
13 1971; Le Maho 1977; Kooyman 1993; Ancel *et al.* 2013). Recent satellite surveys have shown
14 that its present-day breeding range is made up of 52 colonies, distributed along most of the
15 Antarctic shoreline (Fretwell & Trathan 2009; Fretwell *et al.* 2012; Ancel *et al.* 2014). The only
16 significant interruption in this distribution is the Antarctic peninsula (see §10 p. 48), which ex-
17 tends North to almost 62°S, at the very limit of the specie's breeding habitat: indeed, the first
18 documented loss of an Emperor colony happened at the northern end of the species' range, in
19 Dion Island, as recently as the early 21st century, and has been directly associated with the region's
20 important warming trend (Trathan *et al.* 2011). However, the species' high dependence on sea ice
21 makes it a difficult task to accurately predict its future response to climate change (see §18 p. 65),
22 the dynamics of which we still poorly understand in Antarctica (Turner *et al.* 2013; Shu *et al.*
23 2015).

24 The Emperor penguin's complex breeding system poses particular challenges to population mod-
25 elling. This long-lived bird (the current lifespan estimate is unknown - see Borboroglu & Boers-

1 ma 2013 - but at least equal to 35-40 years, Jenouvrier 2004) has a very late reproductive maturi-
2 ty (ca. 4-8 years for males, and 3-6 years for females, Mougin & Van Beveren 1979, and a long
3 generation time (16 years, Jenouvrier *et al.* 2014), which implies that generations are largely over-
4 lapping, in contradiction with the assumptions of the Wright-Fisher population model (Wright
5 1931 and §39 p. 99). Although divorce rate is high in this serially monogamous bird (ca. 85%,
6 Isenmann 1971), mating is not likely to be fully random (Bried *et al.* 1999). Finally, the species'
7 particular spatial organisation during breeding make capture-mark-recapture technically chal-
8 lenging, so that even instantaneous demographic trends are not trivial to infer.

9 *§-24 The Pointe Géologie colony and its regional and continental ties.* The Pointe Géologie Emperor
10 penguin colony is by far the best-studied one in Antarctica. Its presence was one of the reasons
11 for the choice of location for the Dumont d'Urville research station, in 1952, and it has been
12 monitored continuously ever since (Prévost 1961; Isenmann & Jouventin 1970; Isenmann 1971;
13 Jouventin 1971a; Jouventin 1971b; Jouventin 1972b; Jouventin 1975; Le Maho 1977; Barbraud
14 & Weimerskirch 2001a; Jenouvrier *et al.* 2005; Ancel *et al.* 2009; Jenouvrier *et al.* 2009). As it is
15 also the most accessible colony during the winter breeding season, the breeding behaviour of the
16 Emperor penguin is mostly known through observations made on this colony (see references
17 above). The corollary is that our knowledge of the species is often restricted to local observations,
18 and that we may lack a global perspective on behavioural, phenological or demographic traits (see
19 *The Emperor synnyme*, p. 211).

20 Of particular interest, for example, is the sudden decline in the number of breeders on the Pointe
21 Géologie colony, that abruptly dropped from *ca* 6,000 to *ca* 3,000 breeding pairs between 1975
22 and 1985 (Barbraud & Weimerskirch 2001a). Although it has been alternatively linked to local
23 effects of climate change (Barbraud & Weimerskirch 2001a; Jenouvrier *et al.* 2005; Barbraud *et*
24 *al.* 2011), or to increased predation pressure by killer whale as a consequence of the decline of
25 their usual prey, the minke whale, under whaling industry pressure (Ainley *et al.* 2010b), the pe-
26 riod also corresponds to heavy logistic and scientific activity on the colony, in particular large-

1 scale flipper banding (Jenouvrier 2004) which has been unambiguously shown to negatively
2 affect survival (Gauthier–Clerc *et al.* 2004; Dugger *et al.* 2006; Le Maho *et al.* 2011; Saraux *et al.*
3 2011a). Interestingly, the decline of the Pointe Géologie Emperor penguin study colony has a
4 parallel on Crozet archipelago, where the King penguin study colony in the Baie du Marin (see
5 §26 p. 81), also subject to human disturbance and heavy flipper-banding (Delord *et al.* 2004; Sa-
6 raux *et al.* 2011a), is the only one to have declined, while the four other colonies of the same is-
7 land were undergoing very rapid growth (Delord *et al.* 2004) - thus leaving little doubt as to the
8 effects of human activity on colony demographic trajectory.

9 But the decline of the Pointe Géologie colony also raised the question of the fate of the missing
10 penguins. While their disappearance has usually been explained by increased mortality, we sug-
11 gest that dispersal to other colonies may have played an important role (see *Empirical evidence of*
12 *heterogeneous dispersal*, p. 259 and §135 p. 250) - a possibility that would change our interpreta-
13 tion of the global demographic trajectory of the species. However, the lack of both of direct evi-
14 dence for dispersal, and of a broad comparative perspective in the Emperor penguin makes a
15 global assessment of the species' status difficult, and will require an increased methodological
16 effort (see *The Emperor synnome*, p. 211 and *Empirical evidence of heterogeneous dispersal*, p. 259).
17 In the meantime, it is still unclear whether the Pointe Géologie colony can be taken as a reliable
18 indicator of worldwide population trend or whether its importance is rather on a regional scale.

19 *§-25 The King penguin.* The King penguin (*Aptenodytes patagonicus*) is the second-largest extant
20 penguin species, and sister species to the Emperor penguin. It is somewhat smaller, measuring up
21 to ~90 cm and weighting up to ~15 kg at most (Bost *et al.* 2013), and its colouration is more
22 vivid than the Emperor's. It currently breeds on 7 subantarctic archipelagos, between 45°S and
23 55°S, located indifferently North or South of the Antarctic Polar Front (see §2 p. 35, and details
24 in *The King synnome*, p. 163) - and small, incipient breeding colonies have been documented in
25 Tierra del Fuego, in Bahía Inútil (Kusch & Marín 2012). Morphometric differences have justi-
26 fied the distinction between two subspecies: the autonym, *A. p. patagonicus*, that is found in

Concepts and methods - §26

1 South Georgia, is slightly larger than the Indian and Pacific *A. p. halli* according to Mathews
2 (Mathews 1911) and Barrat (Barrat 1976) - although this distinction has not been revisited by
3 the modern literature (*i.e.* using more recent and systematic morphometric tests, or a genetic ap-
4 proach - see *The King penguin*, p. 163).

5 The King penguin's breeding system, although closely akin to that of the Emperor, offers some
6 noteworthy differences. As opposed to the Emperor's fully vagrant incubation, the King penguin
7 selects a local territory of ca. 1.5m² (Barrat 1976), and behaves there in a remarkably territorial
8 way, although it does not build a nest and incubates its egg on its feet in the same manner as its
9 sister species. The King's breeding system also extends over the winter, but egg-laying, hatching
10 and brooding occur during summer, and chicks fast unguarded in «*crèches*» over the winter
11 (Stonehouse 1960; Barrat 1976). During the early chick rearing period, king penguins forage
12 mostly on the APF (see §18 p. 65), which becomes a critical food resource during the months of
13 January and February (Péron *et al.* 2012; Péron *et al.* 2012; Bost *et al.* 2015).

14 Like the Emperor (see §23 p. 78), the King penguin presents interesting challenges to demo-
15 graphic modelling. This equally long-lived bird has a slightly shorter generation time (~10.5
16 years, see §95 p. 186), and usually only breeds successfully every second or third year (Le Bohec
17 2008). The King penguin also has a high yearly divorce rate (70% to 80% based on flipper-band-
18 ing data - Olsson 1998; Bried *et al.* 1999, or 63% based on more reliable PIT-tagging data,
19 Toscani *et al. in prep.*), yet random mating seems contradicted by several behavioural traits (Ols-
20 son *et al.* 2001, and field observations - see also *Fine-scale colony structure*, p. 135). Finally, like
21 the Emperor penguin, the King penguin is highly sensitive to immediate environmental con-
22 straints (*e.g.* Le Bohec *et al.* 2008) and exhibits a high variability in breeding success (Saraux *et al.*
23 2011a, Le Bohec *et al. in prep.*), adding to the challenges of extracting reliable demographic para-
24 meters on short time series.

25 §-26 *The Baie du Marin colony and long-term monitoring design.* Most of the classical knowledge
26 on the King penguin was gathered on South Georgia by Stonehouse (*e.g.* Stonehouse 1960), or

1 on Possession Island by Barrat (*e.g.* Barrat 1976). More accessible and better studied than the
2 Emperor penguin - yet paradoxically almost absent from popular culture - the King penguin has
3 been the object of several more or less large scale monitoring programs - on South Georgia (see
4 Olsson & Brodin 1997; Olsson & van der Jeugd 2002), on Marion Island (Du Plessis *et al.*
5 1994), and most importantly on Possession Island, in Crozet archipelago (Gendner *et al.* 2005).
6 On this latter island, one colony, situated in the *Baie du Marin* (S46°24' E51°45') has been the
7 focus of intensive research in the past decades - including physiology (*e.g.* Handrich *et al.* 1997;
8 Thouzeau *et al.* 2003), demography (*e.g.* Le Bohec *et al.* 2008), parasitology (*e.g.* Gauthier-Clerc
9 *et al.* 2003), life-history (*e.g.* Le Bohec *et al.* 2007; Saraux *et al.* 2011b; Saraux *et al.* 2012; Celia
10 *et al.* 2014), foraging behaviour (*e.g.* Bost *et al.* 2015; Le Vaillant *et al.* 2015), *etc.* First studied
11 since the seventies using flipper-bands, the discovery of their adverse effects (Froget *et al.* 1998)
12 led to their replacement by subcutaneous radio-frequency identification (RFID) tags (Gendner *et*
13 *al.* 2005). The current monitoring program involves the yearly RFID-tagging of ~ 450 nearly-
14 fledged chicks, and the tracking of their movements in and out of their sub-colony using anten-
15 nas buried under the ground on the major access paths to and from the sea. Informatic logging
16 and treatment of these movements allow for the inference of individual life history patterns, such
17 as recruitment and yearly return rate, or timing and success of the major stages of the breeding
18 cycle: thus, fundamental population parameters such as instantaneous growth rate, age at first
19 breeding, yearly breeding success, *etc.* can be extracted, and integrated in population models (see
20 *e.g.* Le Bohec *et al.* 2008; Saraux *et al.* 2011a; Saraux *et al.* 2011b). This system currently repre-
21 sents the best trade-off between the reduction of disturbance to the birds, and the derivation of
22 reliable information about population dynamics.

23 While relatively well known in comparison with the Emperor penguin, the King penguin's de-
24 mography remains an elusive question. The long life span of the species, and the dramatic impact
25 of human activities on the population (see in particular §22 p. 75) fundamentally limits the
26 scope of human-scaled monitoring programs to almost instantaneous parameters: thus, comple-

1 mentary approaches are required if we are to understand long-term processes at work in the
2 species.

3 §-27 *The Peruvian diving-petrel.* The last study of this work will focus on the case of the Peruvian
4 diving-petrel *Pelecanoides garnotii*, that we briefly introduced in §20 p. 71. This small procellari-
5 iforme weights around 200g (based on our field measurements), and is currently known to breed
6 on 4 Northern Chilean islands (Islas Pan de Azúcar, Isla Choros, Isla Grande, and Isla Pájaros II
7 Mattern *et al.* 2002; Simeone *et al.* 2003; Martinez Palma 2014), and a handful of Peruvian is-
8 lands (Isla San Gallán and La Vieja, Zavalaga *et al.* 2010) - a drastic range reduction for this
9 once-abundant seabird, that used to breed along the totality of the Humboldt Current, from 6°S
10 to 42°S (see §147 p. 276 for references and details). Little is known of the natural history of the
11 species (longevity, return rates, breeding success and strategies are all unexplored yet). It breeds in
12 burrows dug in the soft guano of the islands Hays 1989, and occasionally, in Peru, in scree slopes
13 where guano habitat has been destroyed (Carlos Zavalaga *pers. com.*). Burrow fidelity is unknown
14 in the absence of any capture-mark-recapture study. It feeds mostly on anchovies *Engraulis ringens*
15 and krill *Euphausia mucronata*, but also on fish larvæ and planktonic invertebrates during the
16 breeding season (Jahncke *et al.* 1999; García-Godos & Goya 2006), when it remains in local wa-
17 ters (Zavalaga *et al.* 2010, Fernandez-Zamorra *pers. com.*). Its winter foraging strategy is un-
18 known, but anecdotic sightings and the parallel of two sister-species (the South Georgian diving-
19 petrel *P. georgicus* and the Common diving-petrel *P. urinatrix* - see Navarro *et al.* 2013; Navarro *et*
20 *al.* 2015) suggest that it is probably a year-round coastal water resident.

21 The specialisation of the diving-petrel's nesting behaviour in the highly unique guano substrate
22 makes it a very sensitive indicator of habitat destruction. Guano being a rich fertiliser for inten-
23 sive agriculture (due to its concentration in nitrogen, phosphate and potassium), it has been ex-
24 ploited on a large scale since 19th century¹, and this resulted in the destruction of the majority of

1. Incidentally leading to rather controversial political decisions such as the «US Guano Islands Act» of 18 August 1856, granting US citizens the exclusive property of the guano islands they would chance to discover anywhere in the world - or the

1 insular habitats of the *guanero* birds, such as the Peruvian diving-petrel and the Guanay cor-
2 morant *Phalacrocorax bougainvillii*. The large historical population sizes of the diving-petrel did
3 not survive the exploitation, and the maybe up to 1,000,000 pairs dwindled to a few thousands
4 nowadays (see *Unexpected philopatry in an insular seabird, the Peruvian diving-petrel*, p. 275 for
5 more details). As a result of this chaotic demographic history, it is difficult to assess the true state
6 of conservation of the species, that may be recovering, or, on the other hand, still dwindling be-
7 yond salvation. A more in-depth summary of the current state of knowledge will be presented in
8 §149 p. 279-§150 p. 280.

9 *Laboratory methods*

10 §-28 *Utility of classical markers*. Since the advent in 2005 of «next-generation» or high-through-
11 put sequencing, electrophoresis-based individual marker analysis has been slowly losing ground
12 to massively parallel sequencing technologies. In this work, however, we make use of two such
13 «classical» methods: Sanger-sequencing of mitochondrial markers, and analysis of nuclear vari-
14 able number of tandem repeats (VNTR) or «microsatellite» markers. Microsatellite loci differ
15 from single nucleotide polymorphisms (SNP) by their higher mutation rate, mutation process
16 (usually approximated through a Brownian motion model¹), and consequently higher allelic rich-
17 ness. Thus, a single microsatellite marker is, on average, 4 to 12 times more informative than a
18 single SNP (Liu *et al.* 2005). However, the number of markers available in a typical SNP experi-
19 ment is several hundred times higher than what can reasonably be achieved with microsatellite
20 markers, and it has been shown that an increased number of markers contributed highly to detec-
21 tion power for several genetic patterns (Liu *et al.* 2005; Hoban *et al.* 2013b). Direct comparisons

«Chincha Islands war» of 1864-1866 between Spain and a Peru-Chile alliance for the possession of the Chincha guano islands.

1. Brownian motion of particles is used as an approximation of the stepwise mutation in microsatellites: an n -mer allele can randomly mutate to $n+1$ or to $n-1$ at each mutation even, ultimately leading to an increasingly dispersed distribution of fragment lengths.

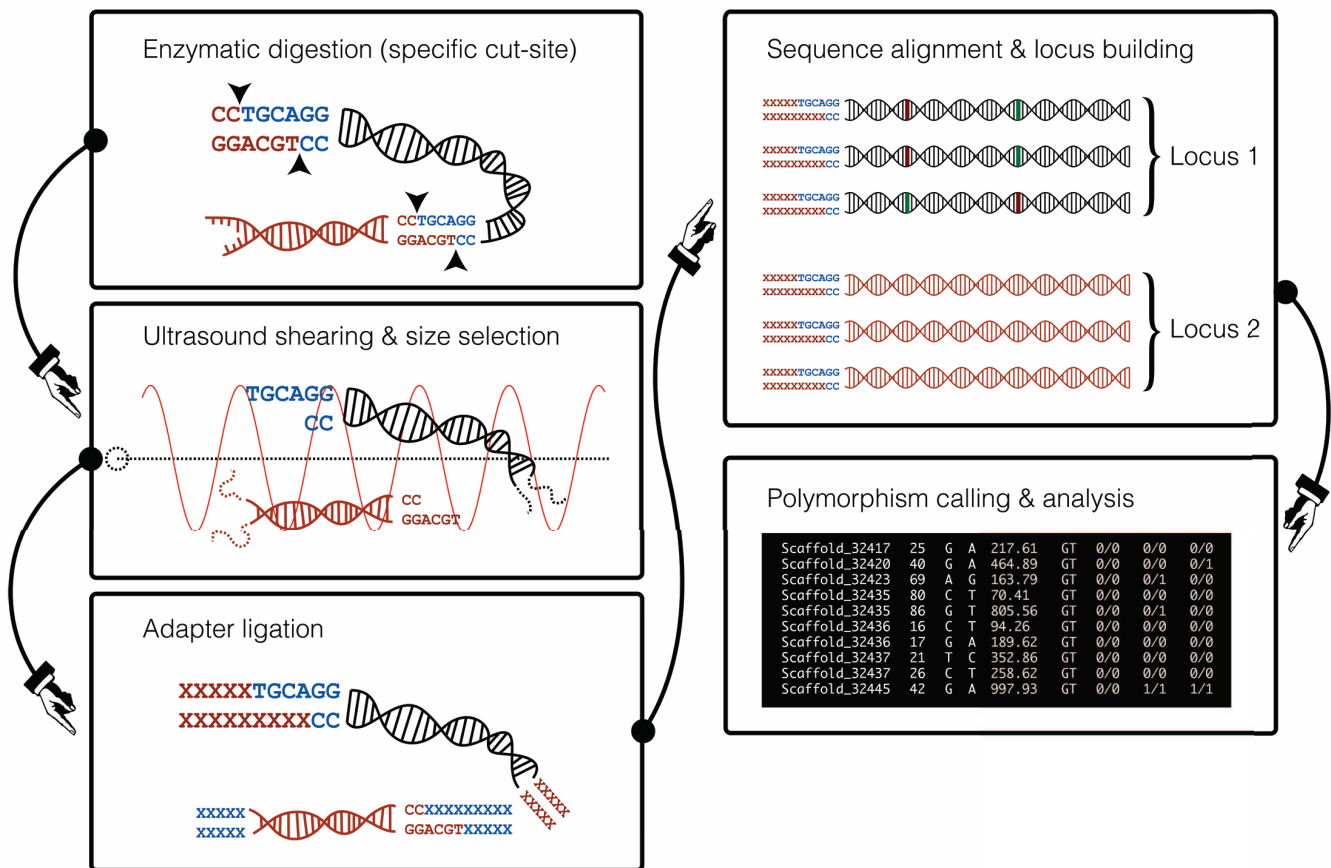
1 informed by empirical data, however, have alternatively concluded in an increased power
 2 (Hoffman *et al.* 2014) or an increased noise (Mesak *et al.* 2014) of SNP compared to mi-
 3 crosatellite markers, so that the latter remain a valid option, at least whenever full transversal
 4 coverage of the genome is not of interest (e.g. for detecting population differentiation, see
 5 Gorospe & Karl 2013). Sanger sequencing, especially of haploid (e.g. mitochondrial) markers, re-
 6 tains one advantage over massively parallel sequencing: the possibility of obtaining a very long
 7 read without a very low error rate, up to 1,000 bp with 99.999% per-base accuracy (Shendure &
 8 Ji 2008), as opposed to 250 bp with a maximum of 99.9% accuracy for NGS (Wall *et al.* 2014),
 9 and mitochondrial DNA barcoding has been shown to retain sufficient resolution for several phy-
 10 logeographical questions (Baker *et al.* 2009; Peucker *et al.* 2009) In the present work, mi-
 11 crosatellite markers were mainly used as a cost-effective way to increase sample size where our pri-
 12 mary interest laid in wide population coverage (see *Fine-scale colony structure*, p. 135). Sanger
 13 sequencing, on the other hand, was used in combination with Restriction-site-Associated DNA
 14 sequencing (RAD-seq data, see §29 p. 85), in order to provide an independent methodological
 15 take on the same dataset, free from potential undetected biases (see *The Emperor synnome*, p.
 16 211).

17 *§-29 The RADseq approach.* Although classical markers remain useful for addressing several popu-
 18 lation genetics questions, the rise of NGS methods has opened new horizons for molecular ecolo-
 19 gy (Allendorf *et al.* 2010). Thus, an important choice when setting up an experiment is the bal-
 20 ance of high-resolution coverage of the genome, and high-resolution coverage of the populations:
 21 in other words, numerous markers in numerous individuals (Hoban *et al.* 2013b). Classical
 22 markers may lack sensitivity to resolve complex population parameters (see §28 p. 84), while
 23 whole-genome resequencing is, as yet, too costly to allow for large-scale surveys of populations. A
 24 trade off is offered by the Restriction-site-Associated DNA sequencing (RAD-seq), developed by
 25 Baird *et al.* 2008. Genomic DNA is cut at specific sites using a restriction enzyme, and flanking
 26 DNA sequences are analysed on a high-throughput platform: thus, a large number of markers

1 can be obtained with a nearly-random representation of the genome. The protocol is as follows:
 2 (i) Genomic DNA is first checked for degradation on a 1.5% agarose gel, and only samples with
 3 consistently high molecular weight are retained, and precisely quantified by fluorometry (Life
 4 technologies™ Qubit®). (ii) approximately 150 ng of genomic DNA per sample are digested with
 5 a restriction enzyme. Each enzyme cutting at a specific restriction site, the enzyme determines the
 6 density of markers along the genome. For this study, we used the Sbf-I HF enzyme (New Eng-
 7 land Biolabs), which is specific for the GC-rich CCTGCAGG pattern. (iii) each sample is then
 8 ligated to a unique barcoded P1 adapter prior to pooling in a single library. The library is then
 9 randomly sheared by sonication; (iv) sonicated libraries are concentrated by DNA capture on
 10 magnetic beads, and the target size range fraction (350-650 bp) is selected by gel electrophoresis;
 11 (v) DNA fragments are prepared for sequencing: single-stranded ends are repaired, a single A base
 12 is added to the repaired ends, and P2 sequencing adapters are ligated to these A-tails; (vi) the li-
 13 brary is enriched by PCR amplification, (vii) the library was then quantified by quantitative
 14 PCR, controlled on a fragment analyser, and sequenced on an ILLUMINA HiSeq 2500 plat-
 15 form, spiked with 20% PhiX control library in order to reduce low-diversity bias introduced by
 16 the aligned restriction sites (see Fig. 29 p. 85).

17 *§-30 Whole-genome resequencing.* In order to complete the RAD sequencing data, we produced
 18 whole-genome resequencing (WGS) data for three King penguins, and three Emperor penguins.
 19 Samples were selected on the basis of DNA quality and efficiency of the corresponding RAD-se-
 20 quencing libraries. For the King penguin, we selected three samples from South Georgia; for the
 21 Emperor penguin, three samples from the Neumayer colony (see Fig. 31 p. 217). Library prepa-
 22 ration followed a standard PCR-free Illumina protocol (Illumina TruSeq low-throughput kit),
 23 and the sequencing was done on two lanes of a HiSeq 2500 V-4 sequencer, at the Norwegian Se-
 24 quencing Centre facility. The two combined lanes yielded a total of 442,637,497 paired-end 125
 25 base-pairs (bp) reads, or a total of 110,659,374,250 bp. Once demultiplexed and combined, we
 26 obtained a total of 73,772,916 (\pm 1,385,583) reads, or 18,443,229,042 (\pm 346,395,842) base

- 1 pairs per sample. Both species were mapped to the published Emperor penguin reference genome
- 2 (Zhang *et al.* 2011b), with a final coverage of ~14X per sample.



3 Figure 8 | The RAD sequencing approach. See details of the protocol in §29 p. 85.

4 From DNA libraries to polymorphism data.

- 5 §-31 Particularity of RAD data for analysis. An important strength of the RAD-sequencing ap-
- 6 proach (see §29 p. 85) is the possibility it offers to work on a genome-wide scale in non-model
- 7 species for which no reference genome is available (Baird *et al.* 2008). As every RAD locus is con-
- 8 strained by the location of cut-sites, first-in-pair read overlap is complete at each particular loca-
- 9 tion. It is thus possible to match the multiple reads covering a locus without any information
- 10 about its position in the genome. But even in the presence of a reference genome, the highly dis-
- 11 cretised character of the RAD-sequencing data changes substantially the analysis paradigm.

1 The « natural » structure of genomic data is shaped by two opposite processes - *coalescence* and *re-*
2 *combination*. The coalescence process can be described as the down-time branching of sequences
3 through the occurrence of new mutations (see §40 p. 101). The recombination process, on the
4 contrary, is the merging of sequences with different coalescent histories, through physical cross-
5 ing-over during meiosis. Thus, a coalescent event can be seen as the fusion of sequences when go-
6 ing *backward* in time, a recombination event is, on the opposite, a fusion of two sequences going
7 *forward* in time (see §42 p. 107). The direct consequence of a shared coalescence history is the
8 linkage between different polymorphic sites: in the absence of recombination, and with an infi-
9 nite-sites model, co-occurrence of derived alleles reflects the coalescent history of the sequences.
10 Recombination, on the other hand, breaks this linkage, by physically linking derived alleles that
11 did not arise in the same lineages in the first place. Since recombination is a relatively random
12 process, the chances that it would happen between two given loci increases with the distance be-
13 tween these loci. Thus, two polymorphic nucleotide positions separated by only a few base-pairs
14 are expected to keep their linkage much longer than two loci separated by several thousands of
15 base-pairs. In continuous genomic data, this means that linkage between markers will decay con-
16 tinuously with increasing distance, until a point where distant loci may be considered effectively
17 independent. Between very close and absolutely linked loci, and distant and effectively indepen-
18 dent loci, virtually all states of linkage disequilibrium can be found.

19 RAD-sequencing data, on the other hand, has a strongly bimodal structure. Since only short
20 fragments (95 to 125 bp) of DNA are sequenced, and since these are assumed to be evenly inter-
21 spersed in the genome, two polymorphic positions that belong to the same locus can be consid-
22 ered non-recombining, and their allele distribution is solely the consequence of the coalescent
23 history of the locus, while polymorphic positions belonging to different loci are effectively inde-
24 pendent, and their co-occurrence on a chromosome is mainly driven by the recombination
25 process. Thus, while the « classical » polymorphism calling framework assumes continuity in the
26 data, and therefore works on a per-base rationale, RAD data can also be handled in a discrete

1 way, first identifying reads belonging to the same RAD locus, and then treating these loci as non-
2 recombining units.

3 *§-32 The Stacks pipeline.* This latter approach is implemented in the popular Stacks pipeline
4 (Catchen *et al.* 2011; Catchen *et al.* 2013). Stacks builds upon the discretised RAD paradigm,
5 and makes the RAD locus a self-standing unit. The analysis pipeline runs in three steps: (i)
6 identifying exactly matching reads, that come from a single genomic location and a single allele
7 (a «stack» in the program's lexicon), and match those stacks within individuals to identify
8 polymorphic loci, and (ii) match loci across individuals to build a common catalog, and (iii) ref-
9 erence all data to this common catalog.

10 The first step can be achieved in two different ways: either (a) by direct local *de novo* assembly,
11 without a reference genome (using the *ustacks* program), by matching the reads against each oth-
12 er, or (b) by mapping the raw reads against a reference genome (using an independent mapping
13 program, in our case *Bowtie2*, Langmead & Salzberg 2012), and collating reads that map at the
14 same genomic position (using the *pstacks* program). While both approach are valid, mapping
15 reads to a reference genome is preferable whenever possible, first because it allows for
16 polymorphism data to be compared between Stacks and other methods through the mapping co-
17 ordinates, and second because it guarantees that sequenced contaminants (bacterial DNA, etc)
18 are not retained in the analysis, as these are not expected to align to the reference genome. It is
19 this second approach that we used in this work, using as a reference the published Emperor pen-
20 guin genome (Zhang *et al.* 2011b), both for the Emperor and for the King penguin sequencing
21 data.

22 The second and third steps are done using the *cstacks* and *sstacks* programs, and allow the build-
23 ing of a MySQL database containing RAD loci (as opposed to polymorphic sites) as individual
24 entries. A particularity of this approach is that polymorphic sites are identified early in the
25 pipeline (at step (i)), within individuals, when matching stacks to form diploid loci. Thus, SNP
26 calling does not integrate information from the whole population as a basis for the likelihood

1 function - therefore there is an increased probability of mistakenly scoring sequencing errors as
 2 legitimate polymorphisms in low-coverage data. A final correction has been implemented in the
 3 rxstacks program, that correct genotype calls based on information from the whole dataset. How-
 4 ever, this two-steps procedure only re-evaluate genotypes at already identified SNP loci, as op-
 5 posed to the SNP positions themselves, and is therefore more error-prone than a native joint
 6 SNP and genotype calling algorithm such as Samtools' mpileup (Li *et al.* 2009) or GATK's Hap-
 7 lotypeCaller (DePristo *et al.* 2011) pipelines.

8 The strength of the Stacks approach, however, remains the rapidity and ease of querying and fil-
 9 tering offered by the MySQL database structure, that allows for efficient data mining, before de-
 10 ploying more robust and computationally-intensive analysis methods.

11 *§-33 SNP-calling approach.* Although the locus-based paradigm offered by the Stacks framework
 12 (see §32 p. 89) is an efficient representation of RAD data structure, it is also possible to handle
 13 RAD-sequencing data as conventional high-throughput sequencing data, provided the bimodal
 14 linkage structure is taken into account in downstream analysis. This approach involves three suc-
 15 cessive steps: *(i)* raw read mapping and filtering, *(ii)* SNP and genotype calling, and *(iii)* data
 16 filtering.

17 For the initial read mapping, we use *Bowtie2* (DePristo *et al.* 2011), allowing only for fully
 18 mapped read pairs, respecting the observed length of our paired-end libraries (250 to 1,000 bp).
 19 While other tools, such as the Burrow-Wheeler transform alignment (BWA, Li & Durbin 2009),
 20 perform a comparable task, not all support end-to-end read mapping: BWA, for example, allows
 21 for missing bases at both ends of the raw reads. This is suitable for most short-read mapping situ-
 22 ations, but not for RAD sequencing, which relies on the presence of a cut-site, and thus puts a
 23 strong constraint on the correct alignment of the first six (in our case) bases of the raw reads. Us-
 24 ing *Bowtie2*, we have typically high alignment rates (75% to 85%). The resulting alignment files
 25 (see §38 p. 98) are filtered for mapping quality, sorted, and indexed using Samtools (Li *et al.*

1 2009) and Picard Tools (picard.sourceforge.net), including a step of duplicate removal (see §35 p.
2 93 and §36 p. 95).

3 Current genotype and SNP calling algorithms can be broadly divided into two categories: «clip-
4 off» or read-count-based calling, and probabilistic calling (Nielsen *et al.* 2011). While count-
5 based calling simply relies on the relative proportion of reference and alternate bases to assign a
6 genotype, probabilistic methods rely on a polymorphism model as a prior, to assign likelihood
7 scores to each possible genotypes - this latter family of methods being generally considered much
8 more accurate, although the details of the prior can vary widely (e.g. assuming or not Hardy-
9 Weinberg equilibrium, etc, see Nielsen *et al.* 2012). In this work, we use two different algorithms
10 for calling SNPs and genotypes from alignment files. A robust and accurate approach is imple-
11 mented in GATK's HaplotypeCaller pipeline (DePristo *et al.* 2011): each individual alignment
12 file is converted to a gVCF format (see §38 p. 98) that includes information about physical phas-
13 ing (see §37 p. 97), using the HaplotypeCaller java program. These individual gVCF files are
14 then scanned all together on a per-site basis by the GenotypeGVCF program, that makes use of
15 individual depth, base quality, variant frequency, and variant phasing information to call
16 genotypes in a maximum-likelihood framework. The main drawback of this approach is its ex-
17 tremely time-consuming character, as computation time grows exponentially with the number of
18 included individuals: thus, on our platform, genotyping ~200 individuals for the whole
19 RADome on 48 threads can take up to 10 days. Another faster, and equally accurate method is
20 the Samtools mpileup/bcftools pipeline (Li *et al.* 2009), which is also deployed in two steps: first
21 the collating of all alignment files into a single «pileup» format, and then a genotype calling and
22 filtering step, also in a maximum-likelihood framework. While the Samtools pipeline can not be
23 multithreaded, it can easily be split in chunks along the genome, and the results merged after
24 genotype calling.

25 The final filtering of called genotypes is mostly done in VCFtools (Danecek *et al.* 2011). The de-
26 tails of filtering greatly depends on the analysis for which data is prepared, although some aspects

1 are invariant. For one, a form of *ensemble* calling may be highly desirable. All maximum-likeli-
 2 hood genotype calling algorithms are model-based: they implement an explicit model describing
 3 the likelihood of a particular genotype given the observed data. Yet these models differ between
 4 algorithms: while they overlap very closely in the «central» part of the genotype calling problem,
 5 they may differ quite widely at its limits - for example, under very high or very low coverage, or
 6 when both alleles are represented with very asymmetric depth. However, while it is possible to as-
 7 sess the accuracy of genotyping using an ascertained SNP panel in model data (*e.g.* human ge-
 8 nomic material), there is no straightforward method for ranking genotype models in more exotic
 9 organisms. A robust method to address the potential lack of agreement between SNP models is
 10 to retain only consensus SNPs, i.e. loci and genotypes called the same way by a panel of models,
 11 as the most reliable set of loci (Greminger *et al.* 2014).

12 Within this consensus SNP set, further filtering may be necessary, relative, this time, not to the
 13 model after which genotypes have been called, but rather to the structure of the available popula-
 14 tion dataset. Some analyses may be overly sensitive to missing data (*e.g.* non-assignment bias in
 15 principal component analysis), in which case only loci genotyped in a high proportion of individ-
 16 uals should be retained. Some analyses, on the other hand, are very sensitive to genotype mis-as-
 17 signment (*e.g.* erroneously calling a homozygous site, where the true underlying genotype is het-
 18 erozygous), in which case it is critical to impose a minimum depth, *e.g.* 10x, over retained
 19 genotypes (see §35 p. 93) in order to obtain reliable results. However, all such filtering always
 20 comes at the cost of reducing the number of loci available for analysis: therefore, a tradeoff often
 21 needs to be made between the quality of the data necessary for the analysis, and the quantity of
 22 data required for robust results (see §36 p. 95).

23 *§-34 Genotype-free approach.* Most current model-based calling algorithms implement a maxi-
 24 mum-likelihood framework (see §33 p. 90). Genotype calling thus commonly happens in two
 25 steps: first, at each position, a likelihood score is calculated for all 10 possible genotypes (AA, AC,
 26 AG, AT, CC, CG, CT, GG, GT, and TT) - this score being defined as the probability of the se-

1 quencing data given each possible genotype (Korneliussen *et al.* 2014). Then, the genotype with
 2 the best likelihood score is retained as the «true» genotype, provided the contrast with the sec-
 3 ond-best likelihood score is steep enough. However, this second step, by choosing one single
 4 genotype, suppresses the information contained in the genotype likelihood distribution (Nielsen
 5 *et al.* 2005; Fumagalli *et al.* 2013). In many cases, this gives a false sense of certainty to data that
 6 did not, in reality, support any positive genotype call: thus, it has been proposed that analysis
 7 may often be more accurately performed by focusing on the genotype likelihood distribution,
 8 rather than on called genotypes (Nielsen *et al.* 2005; Nielsen *et al.* 2011; Nielsen *et al.* 2012; Fu-
 9 magalli *et al.* 2013). This approach has recently been implemented in the ANGSD program (Ko-
 10 rneliussen *et al.* 2014). This approach is particularly powerful for low-coverage data, where the
 11 genotype likelihood distribution tends to make little difference between homozygous and het-
 12 erozygous calls. Therefore, it was preferred for our RAD sequencing data, whenever suitable
 13 analysis algorithms were implemented for downstream data treatment (e.g. principal component
 14 analysis and fixation index in Fumagalli *et al.* 2014, admixture analysis in Skotte *et al.* 2013, or
 15 derived-allele frequency spectrum estimation in Korneliussen *et al.* 2014).

16 §-35 *Exploring the duplicate bias.* Accurate estimation of SNP positions and individual genotype
 17 likelihoods is further complicated by the technical challenges of Illumina platforms, such as we
 18 use in the present study. Two important characteristics of high-throughput sequencing platforms,
 19 in particular of Illumina type, are (i) the random sampling process that occurs during fragment
 20 binding and cluster generation and (ii) the relatively high error rate, with possibly more than
 21 0.1% erroneous base calls. Thus, every analysed fragment must be sequenced several times inde-
 22 pendently, in order to ensure that (i) both alleles have been sampled, and (ii) sequencing errors
 23 can be accurately filtered out.

24 Genotype under-sampling occurs when, at a heterozygote position, only one of the two possible
 25 bases is sampled across the whole sequencing coverage (Nielsen *et al.* 2012). At a sequencing de-
 26 pth of one, this is necessarily the case. At a depth of n , this occurs with a probability of $(1 /$

1 $2^{(n-1)}$). Thus, at a depth of 5, the probability of under-sampling the true genotype is 0.0625 (in
 2 other words, out of a very large number of heterozygous positions, ~6.25% will erroneously be
 3 called as homozygous); at a depth of 10 that probability falls to $9.77e^{-4}$. This, however, refers to
 4 «effective» or «true» sequencing coverage, where 5X means that 5 independent genomic copies
 5 of a DNA region have been sequenced in the same individual. Yet observed coverage may be sig-
 6 nificantly overestimated compared to effective coverage.

7 Indeed, in case of sub-optimal library preparation efficiency, too few individual DNA fragments
 8 are prepared compared to the flow cell capacity. During PCR library enrichment, multiple clonal
 9 copies of these fragments are generated. Normally, the random sampling process occurring dur-
 10 ing flow cell clustering would ensure that ~1 copy of each fragment only is represented in the se-
 11 quenced library. However, in case of a low-diversity library, several PCR-copies of each fragment
 12 will be sequenced. Although these fragments, or «PCR duplicates», contribute to the total se-
 13 quencing depth, they do not stem from independent genomic copies. Thus, if each independent
 14 fragment is, on average, sequenced along with 1 duplicate (50% of the library is thus made up of
 15 duplicates - or 50% duplication rate), an observed coverage of 10X will, in reality, only give in-
 16 formation as to 5 independent genomic copies: its effective depth is 5X. The probability of un-
 17 der-sampling genotypes will therefore be $1 / (2^{(5-1)}) = 0.0625$, and not $1 / (2^{(10-1)}) \approx 9.77e^{-4}$. This
 18 problem can become more important in case of drastic scenarios of PCR duplication, where the
 19 rate can reach up to 90% (a coverage of 10X will represent only one genomic copy, and will not
 20 be enough to identify a heterozygote).

21 The most visible consequence of genotype under-sampling will be the loss of heterozygosity de-
 22 spite acceptable coverage (10 to 20X, see Hoffman *et al.* 2014). Under ideal conditions, observed
 23 heterozygosity increases with depth of sequencing at low coverage, and converges towards true
 24 heterozygosity at 15 to 20X (when virtually every site has a nearly-null probability of having been
 25 under-sampled). However, if a large number of duplicates are present, correlation between het-
 26 erozygosity and depth of coverage is heterogeneous, or absent.

1 If left unfiltered, PCR duplicates are likely to introduce a strong bias in several analyses, by lead-
2 ing to overly confident genotype calls where, in reality, evidence is scarce. In more extreme cases,
3 it will also lead to homozygous calls where both alleles have been sampled, by artificially increas-
4 ing the depth of coverage of one allele only, and leading to discarding the low-coverage alternate
5 allele as a sequencing error.

6 Fortunately, the original RADseq protocol allows for efficient duplicate removal. Since DNA is
7 randomly sheared after restriction, the length of individual fragments is expected to be broadly
8 distributed; at medium to high coverage, no two fragments are expected to have the exact same
9 length in the same individual, at the same locus. Thus, if two fragments of the same individual
10 share the same mapping coordinates both on their front-read and reverse-read, these may be con-
11 sidered to be duplicates, and one of them (the one with the overall lowest quality) may be
12 removed. An approach implemented in the popular pipeline Stacks (Catchen *et al.* 2011;
13 Catchen *et al.* 2013) uses as a filtering criterion the fact that both front-read and reverse-read
14 must match exactly in order to discard the lowest-quality read pair as a duplicate. This approach,
15 however, assumes that sequencing errors are absent, which is a very unrealistic hypothesis in the
16 case of high-throughput sequencing data: the consequence is that all duplicate reads including se-
17 quencing errors will be kept as original genomic copies, together with genuine reads, which may
18 lead to many false-positive SNP calls. Another approach, implemented in PicardTools ([http://pi-
19 card.sourceforge.net](http://picard.sourceforge.net)), relies instead on prior read mapping: two read pairs will be flagged as du-
20 plicate only if their mapping coordinates and mapping descriptors (as symbolised in a CIGAR
21 string¹) are identical. This approach has the advantage of not being sensitive to sequencing errors;
22 however it necessitates a reference genome of some kind in order to perform read mapping. We
23 preferred the latter approach for this study.

1. A character string summarising the structure of the alignment - number of consecutive matching bases, gaps, etc. (see <https://samtools.github.io/hts-specs/SAMv1.pdf>).

1 *§-36 The duplicate bias in our data.* RAD library preparation involves several sensitive steps that
 2 may contribute to decreasing the overall protocol efficiency (see §29 p. 85), and thus lead to in-
 3 creased presence of PCR duplicates in the raw sequencing output (see §35 p. 93). In our data,
 4 the presence of duplicates was first noticeable because of the heterogeneous regression between
 5 heterozygosity and depth of sequencing between libraries, and generally the strong correlation be-
 6 tween individual inbreeding estimators and library identity. Removing PCR duplicates efficiently
 7 solved this problem, however, the resulting effective coverage was consequently decreased. As a
 8 result, the number of sites that could be reliably called by conventional (e.g. maximum-likeli-
 9 hood) SNP and genotype calling algorithms (see §33 p. 90) was much reduced. Thus, in order to
 10 keep a SNP-based analysis framework, we must balance two antagonistic parameters, (i) the min-
 11 imum depth of coverage to retain a genotype and (ii) the number of loci necessary for the analy-
 12 sis. While low coverage is reducing the certainty of genotypes (see §35 p. 93), and thus potential-
 13 ly introducing noise in the data, increasing the minimum genotype depth effectively amounts to
 14 increasing the amount of missing data, and thus decreasing the number of exploitable loci (see
 15 §33 p. 90). We tested this tradeoff on a dataset containing RADseq data from two well-separated
 16 populations of Peruvian Diving-petrels *Pelecanoides garnotii* (see *Unexpected philopatry in an insu-*
 17 *lar seabird, the Peruvian diving-petrel*, p. 275), by running the Structure clustering algorithm
 18 (Pritchard *et al.* 2000) on filtered datasets with a minimum depth ranging from 20X to 40X, and
 19 a minimum proportion of genotyped individuals per locus ranging from 40% to 90%. While
 20 structure is perfectly recovered for balanced datasets (e.g. 30X and 60% representation), it decays
 21 with more stringent filtering (e.g. 90% representation) due to the low number of remaining loci.
 22 This particular algorithm being especially sensitive to missing data, it highlights the importance
 23 of finding a good compromise between genotype certainty and missing genotypes - the main
 24 problem posed by PCR duplication and loss of sequencing efficiency.
 25 Being PCR-free, our whole-genome resequencing libraries are immune from the duplication is-
 26 sues identified in our RAD-sequencing data (see §35 p. 93): therefore, the data processing steps

1 could be greatly simplified. For each sample, we pooled the data from both sequencing lanes (see
 2 §30 p. 86), and mapped it to the published Emperor penguin reference genome (Zhang *et al.*
 3 2011b) using Bowtie2 (Langmead & Salzberg 2012), allowing only for unique, concordant
 4 paired-end alignments. Output BAM files were sorted and filtered using Samtools 0.1.19 (Li *et*
 5 *al.* 2009) and PicardTools 1.113 (<http://picard.sourceforge.net>).

6 *§-37 Phasing of the WGS data.* A particular difficulty, in WGS data analysis, is the inference of the
 7 correct phase for the pre-established polymorphism. Two main methodological approaches have
 8 been proposed (Browning & Browning 2011): (i) an observed-linkage approach, and (ii) a com-
 9 putational approach. The observed-linkage or read-based approach (i) exploits the naturally
 10 phased base calls in the paired-end Illumina reads, in a process similar to scaffolding, but using
 11 the prior positional information provided by the read mapping step. At sufficiently high cover-
 12 age, the overlap between reads may allow the reconstruction of long stretches of phased hap-
 13 lotypes. However, the process is often broken - either necessarily whenever the distance between
 14 two successive heterozygous sites exceeds the library insert length, or stochastically through varia-
 15 tion in read overlap distribution. Thus, at medium coverage, read-based phasing tends to produce
 16 short blocks of phased haplotypes, separated by stretches of unphased sequence. The computatio-
 17 nal approach (ii), on the other hand, relies on a prior model of recombination to estimate the
 18 most likely local haplotypes given the observed linkage data. Ideally, a specific recombination
 19 model is established beforehand in the form of a genetic map, or an ascertained phased reference
 20 panel is used. In a non-model species, however, computational phasing may rely on population
 21 data: across a large number of individuals, assuming a constant recombination rate, the measure
 22 or linkage disequilibrium between closely located SNPs is considered informative of their physi-
 23 cal phasing, and used to infer the most likely underlying local haplotypes.

24 Ideally, both approaches should be combined, as is the case in the ShapeIt PIRs pipeline
 25 (O'Connell *et al.* 2014): read-based physical linkage is assessed first in order to build local hap-
 26 lotypes, and the phasing between those is further assessed computationally using population data.

1 In our case, however, the low number of sequenced individuals (3 per species) prevented the use
 2 of any computational method. Therefore, we reconstructed the local haplotypes using the graph-
 3 based algorithm implemented in HapCompass (Aguilar & Istrail 2012), and used the resulting
 4 partially-phased data whenever phase information was necessary. However, the resulting phasing
 5 proved insufficient, and results visibly lacked robustness, presenting artefact patterns similar to
 6 those encountered in unphased data: we therefore did not retain it for final analyses (see *The King*
 7 *synnome*, p. 163 for details).

8 *§-38 A point on data management.* Handling of high-throughput sequencing data presents impor-
 9 tant scale challenges, both in terms of computational power and of data storage and manage-
 10 ment. The methods used to address these challenges vary according to the specific analysis stages,
 11 although some general guidelines can be retained. When processing sequencing data, four main
 12 «generations» of data files can be identified, of increasing specificity and decreasing redundancy:
 13 (i) raw data formats, (ii) alignment formats, (iii) variant formats, and (iv) analysis input formats.

14 (i) **Raw data** is mostly stored as (compressed) fasta or fastq format - a sequencer-output format
 15 listing sequence base calls (as performed, in our case, by Illumina's CASAVA software), physical
 16 read coordinates on the flow cell, and base quality encoded as Phred33 quality scores. Redundan-
 17 cy is high, as several reads, or dozens of reads, may cover the same position in the target genome.

18 (ii) **Alignment files** are output by read mapping algorithms, usually formatted as SAM (Sequence
 19 Alignment Map), or as its binary equivalent BAM (Binary Alignment map). After alignment, the
 20 complete information from read sequence and quality scores is retained, but mapping coordi-
 21 nates are added, relative to a reference genome: at that stage, redundancy is still high, as each
 22 individual read is represented by a single entry (and not collapsed by genomic position). SAM/
 23 BAM files are typically handled using the flexible Samtools package (Li *et al.* 2009), or the Broad
 24 Institute's Picard Tools package (<http://picard.sourceforge.net>). BAM format is the preferred data
 25 storage format in this work, as it retains all the necessary information for alternative SNP calls.

1 *(iii) Variant calls* are produced through SNP and genotype calling algorithms (see §33 p. 90),
2 usually stored as a VCF (Variant Call Format, or its binary version, the BCF) file, and usually
3 handled using VCFtools (Danecek *et al.* 2011). The VCF format removes most of the redundan-
4 cy of raw and aligned formats, as each genomic position is represented by a single entry, and the
5 number of reads covering that position is summed numerically for each individual. Thus, a major
6 shift of paradigm occurs between the aligned and the called formats: the identity of the original
7 reads is lost, and information switches from a «horizontal» (sequence) format, to a «vertical»
8 (per-site) format. Usually only polymorphic sites are retained at that step, although the Broad In-
9 stitute gVCF (genomic Variant Call Format) includes also the sequence at non-polymorphic sites,
10 as «blocks» of sequence data. The main characteristic of VCF formats is that it is an «analysed»
11 format, storing an interpretation of the raw genetic data, rather than the data itself: it contains
12 information that is the intersection of a dataset, an analysis protocol, a particular variant calling
13 model, filtering decisions, etc.

14 *(iv) Input formats* are usually particular to each analysis program. They are, in their vast majori-
15 ty, a reformulation of the VCF format, with or without data loss (e.g. 0/1 coding of alleles in-
16 stead of ATCG coding, etc). They are too specific to an individual program or version to be de-
17 tailed in this work.

18 *Genetic analysis framework*

19 The detail of the particular analyses and models used in each chapter will be detailed in the corre-
20 sponding Methods sections. Here, however, we present the conceptual framework of the methods
21 used repeatedly throughout this work.

22 *§-39 Populations, random mating and neutral drift.* As we introduced earlier (see §2 p. 25), the ba-
23 sis of population genetics is the *deme*, which, whenever it has a strong geographic component, is
24 described as a *population*. A population, in the population genetic framework, is oftentimes ap-

1 proximated through the *Wright-Fisher model* (Wright 1931; Ewens 1979), as an ensemble of *ran-*
 2 *domly mating* («*panmictic*») individuals, where each individual will be replaced by one offspring
 3 at the next generation, and where generations do not overlap. Although seemingly rather unreal-
 4 istic, this model has repeatedly been proved to perform surprisingly well under most empirical
 5 situations (Hamilton 2011), at the only cost of the necessary re-definition of some concepts, such
 6 as *population size* (see §41 p. 103 for a development).

7 One essential feature of the Wright-Fisher model is the way the transmission of alleles is con-
 8 ceived. The Wright-Fisher model describes a simple Markovian process: at each generation event,
 9 allele frequencies are re-sampled from the ancestral gene pool to the offspring gene pool under a
 10 binomial law, with stochastic variation in frequencies; thus, the gene pool at time $T+1$ only de-
 11 pends on the gene pool at time T and on a stochastic sampling event, without memory from pre-
 12 vious states. The expectation for the binomial sampling is that, in the absence of selection
 13 pressure, allele frequencies remain the same across generation (the *Hardy-Weinberg propriety*, see
 14 Edwards 2008) - an expectation marginally contradicted by the stochastic nature of the sampling,
 15 which allows frequencies to vary in a random direction at each generation. The extent of this
 16 variation, or *genetic drift*, is a function of the *size* of the Wright-Fisher population: a larger popu-
 17 lation will undergo less drift than a smaller one.

18 *Drift* thus becomes the main tenet of the Wright-Fisher neutral molecular evolution model: new
 19 alleles, that occur randomly and after an *infinite sites model* (*i.e.* no two mutations ever occur at
 20 the same site, see Kimura & Crow 1964 for details), can either be lost through random sampling,
 21 or on the contrary see their frequency increase in the population; so that the probability of loos-
 22 ing an allele through drift decreases as the frequency of the allele increases.

23 As opposed to the *coalescent* framework (see §40 p. 101), the Wright-Fisher Markovian model
 24 does not integrate evolution over the history of the population: there is a succession of states, but
 25 no *origin* in the process. Thus, the discrete generation steps of the Markov chain can be approxi-
 26 mated as a continuous process, in what is known as the *diffusion approximation of genetic drift*

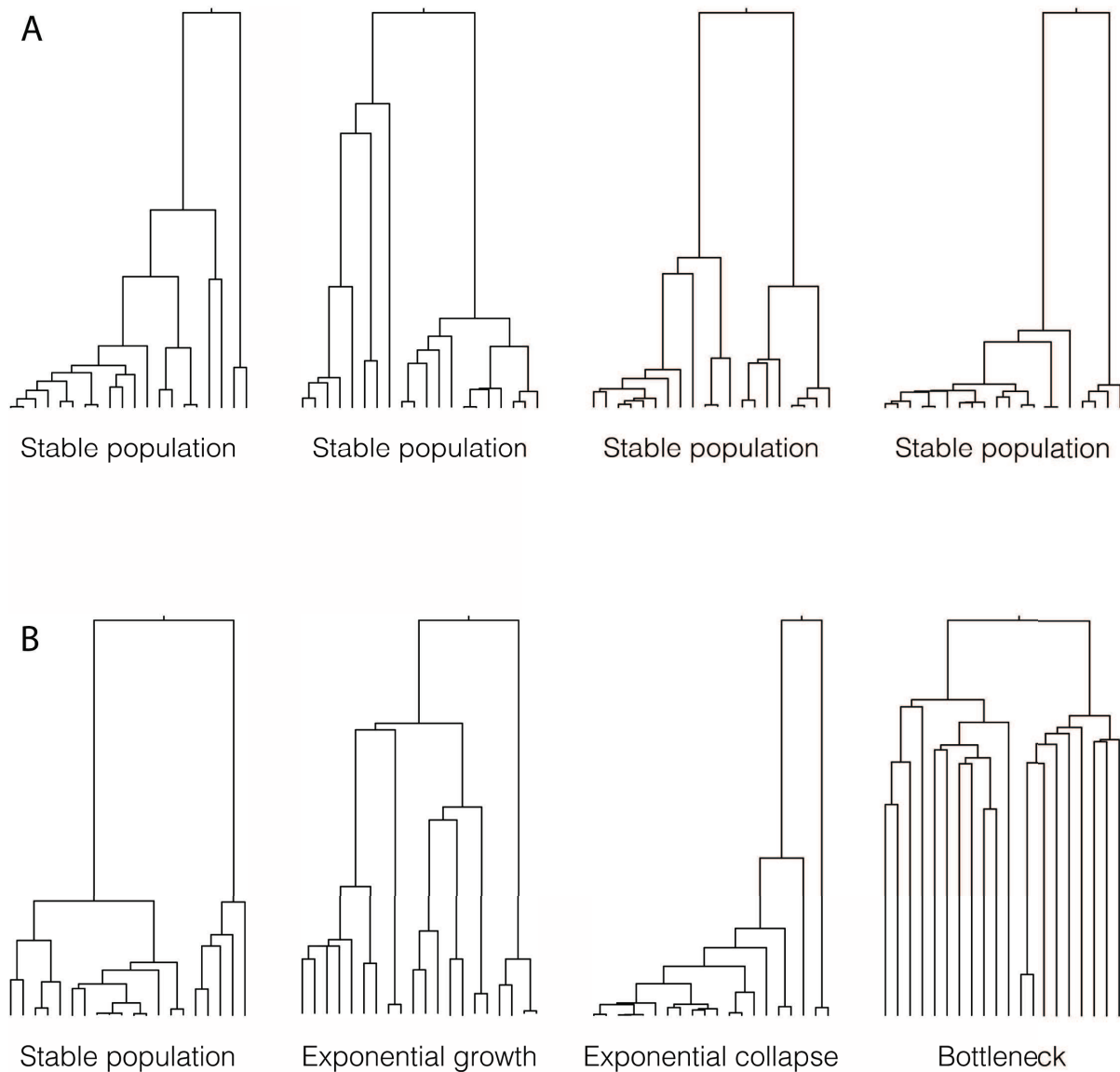
Concepts and methods - §40

1 (Kimura 1955). Under the Kimura model, alleles are idealised as *particles*, their frequencies as *co-*
2 *ordinates*, and drift as *movement*. The diffusion process results in a change in the location of alleles
3 in the frequency space, under a probability law determined by the allele's *diffusion coefficient* - it-
4 self a reflection of the population size and demographic events that may vary through time.
5 While the expectation for the movement of the allele remains zero for any diffusion coefficient
6 and any length of time (in keeping with the Hardy-Weinberg expectation), the *variance* of the fi-
7 nal position increases with increased diffusion coefficient - in other words, over a large set of loci,
8 a sharply clustered frequency distribution is expected to spread out over time. The diffusion ap-
9 proximation becomes especially useful in N-dimensional frequency spaces, *i.e.* when the frequen-
10 cy of an allele is considered in several populations simultaneously, with limited diffusion allowed
11 between populations (see §45 p. 117 for applications of this propriety).

12 Although powerful conceptual tools, the Wright-Fisher model and Kimura diffusion theory have
13 the limitation of being, by default, a site-based paradigm: alleles are represented as (usually inde-
14 pendent) entities whose relationship is merely complementarity in the frequency space. In other
15 words, the drift paradigm does not take the genealogical evolution of sequences into account -
16 thus, it can not exploit one of the richest information sources in the genome structure.

17 *§-40 The Coalescent theory.* Developed by Kingman in the 1980s, the coalescent theory is a di-
18 achronic take on the previously dominantly synchronic paradigm of population genetics, and the
19 Wright-Fisher approximations of neutral drift (Kingman 1982; Kingman 2000). As opposed to
20 essentially atemporal descriptive statistics such as fixation indices (Wright 1965) or analyses of
21 molecular variance (Excoffier *et al.* 1992b), the coalescent explicitly models the history of the ob-
22 served state of genetic diversity. When considering the tree-like history of a short DNA sequence,
23 each new mutation appears as the branching of the ancestral lineage into two new sequences: one
24 unchanged, and one new mutated sequence. If we happened to know the «true» history of that
25 sequence in a population, we would witness several such branching events, as well as several ext-
26 inction events, when a lineage becomes lost by the effects of random sampling, or natural selec-

1 tion. But considering a realistic experimental situation, we really only sample the final state of
 2 that history, *i.e.* the present state. When reconstructing that history from the data contained in its
 3 present state, each *branching* event now appears as the *merging* of two lineages into one ancestral
 4 lineage - a *coalescence* event.



5
 6 **Figure 9 | Stochasticity and demography in the coalescence process.** Gene trees simulated under
 7 different demographic models using *msHOT* (Hellenthal & Stephens 2007). (A) Constant population: all four trees
 8 are replicates of the same simulation parameters. (B) Four different demographic histories.

1 Coalescence is an essentially stochastic phenomenon (Fig. 9A p.102): interestingly, although the
 2 actual coalescence events are random, the probability of two lineages coalescing (*i.e.* having a
 3 common ancestor) at any given point of the past is a direct function of the effective population
 4 size. Given a diploid population of constant effective size N_e (thus $2.N_e$ lineages in the gene pool
 5 at each generation), and a constant evolution rate of one substitution per generation, the proba-
 6 bility of any two lineages sharing a common ancestor at the previous generation is $(1 / 2.N_e)$,
 7 and the probability of *not* sharing a common ancestor is $1 - (1 / 2.N_e)$. Expanding this to any
 8 generation t , the probability of two lineages having a common ancestor at generation t is the
 9 product of their probability of having a common ancestor given the population size at generation
 10 t , and their probability of not having already had a common ancestor at any previous generation,
 11 or $(1 - (1 / 2.N_e))^{(t-1)} \cdot (1 / 2.N_e)$. If N_e is constant and large enough, the instantaneous
 12 probability of coalescence is very small, and the cumulative probability of two lineages coalescing
 13 at or before generation t can be approximated by the exponential function, as $1 - e^{-(1 / 2.N_e) \cdot t}$.

14 Thus, the coalescence theory sets an expectation for the rate of coalescence of lineages in a ran-
 15 dom sample, which follows an exponential distribution if the population size remains constant
 16 through time. As it gives a diachronic picture of the evolution of sequences in a Wright-Fisher
 17 population, it allows us to step from the earlier synchronic paradigm of population genetics, into
 18 the reconstruction of time-dependent events such as demography and migration history.

19 *§-41 Coalescence, population size and migration.* The probability of coalescence at any given gener-
 20 ation is a function of the population size at that generation (see §40 p. 101). This has two impor-
 21 tant consequences: (*i*) the neutral, exponential coalescent provides us with a null model against
 22 which we can assess the neutrality of observed coalescent histories, and (*ii*) we can further devel-
 23 op the coalescent framework to explicitly model population size changes through time. But an
 24 important first step is to clarify what we mean by *population size* - since, as we already discussed

1 earlier (see §2 p. 25), the fundamental concept of *population* is defined very differently in the
2 ecological, and the evolutionary frameworks.

3 Revolving essentially around the notion of *individual*, the ecological paradigm naturally considers
4 the *population size* to be the *number of individuals alive at the same time* in the group under scruti-
5 ny. Here, we call *census size* this direct observation of the population size. While the census size of
6 a full species is a rather understandable concept («*how many king penguins are there at this mo-*
7 *ment on Earth?*» does have a true answer), the census size of a population is a rather more abstract
8 idea as soon as life history strategies become elaborate (see Lecomte *et al.* 2009 for an interesting
9 practical example): in these seabirds, adults will only breed up to two years out of three, and will
10 probably not come ashore at all during the third year - so that the maximum number of adults in
11 reproductive age on the colony will only represent two-thirds of the total number of breeders
12 who «belong» in that colony. Juveniles that are born on a colony regularly spend several traveling
13 years at sea, and may or may not choose to disperse after that time: should they be counted as
14 part of the colony in advance, or not until they effectively settled? What about juveniles in *r-strat-*
15 *egy* species, where the number of juvenile is very high, but mortality tremendous? Thus, although
16 the concept itself appears fairly straightforward, its application is much more subjective in wild
17 populations - all the more so if we lack precise knowledge about the recruitment and the propor-
18 tion of non-breeders that may be temporarily away.

19 The evolutionary paradigm, on the other hand, is based on the notion of *gene pool*. Thus, is only
20 counted as part of the genetic population an individual that actively contributes to the gene pool:
21 population size is there defined as the *number of successfully breeding individuals*. To distinguish it
22 from the census size, we call this measure the *effective size* of the population. Here again, this defi-
23 nition is simple in the ideal case of a Wright-Fisher population, with non-overlapping genera-
24 tions and replacement of each parent by one offspring (see §39 p. 99). But it becomes more diffi-
25 cult to estimate whenever life history diverges from this abstract state - *e.g.* when a breeder may
26 have several offsprings, or when breeding success is heterogeneous within the population. Effec-

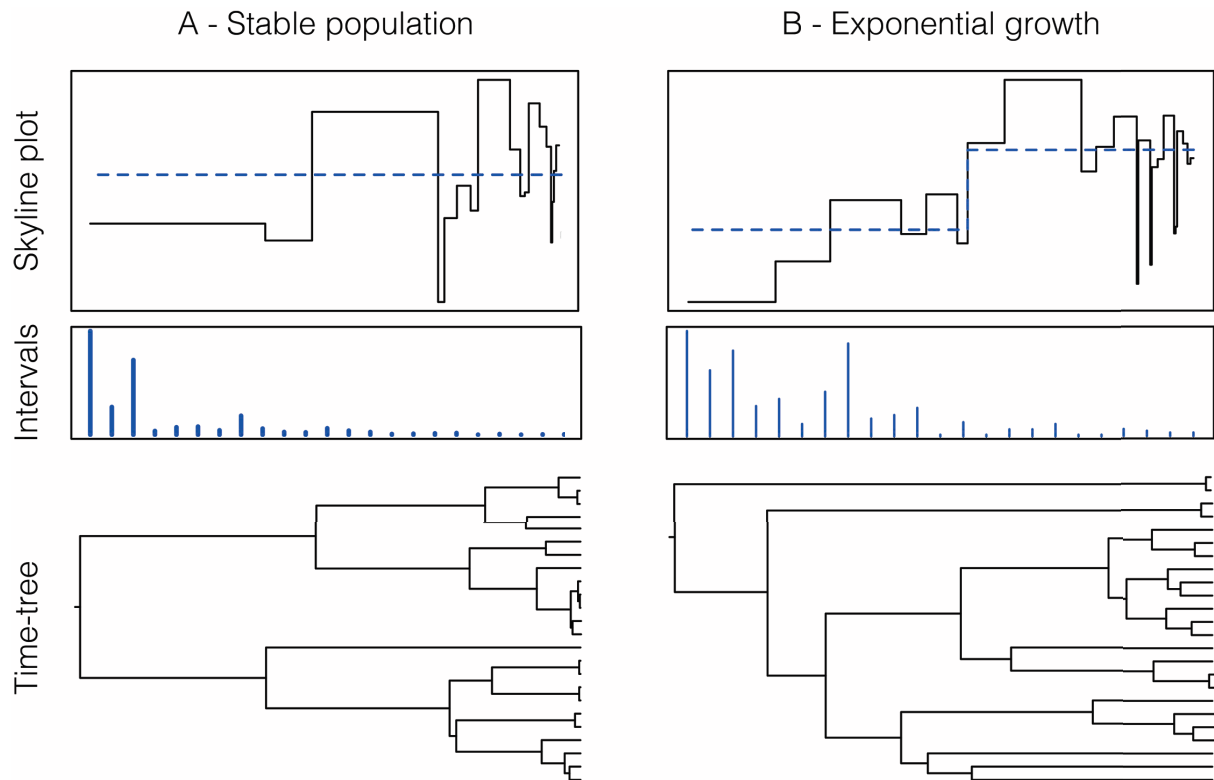
1 tive population size also diverges widely from census size (sometimes by orders of magnitude, see
 2 for example *The Emperor synnome*, p. 211) as soon as non-breeding individuals are an important
 3 part of the species, *e.g.* when sexual maturity is delayed (as in most long-lived species), or when
 4 an adult may choose to not breed if its body condition is not adequate - or when a senescent in-
 5 dividual chooses not to die (as in humans and albatrosses). As it then becomes increasingly diffi-
 6 cult to define the number of individuals contributing to the species' genetic diversity through ex-
 7 ternal evidence, the effective population size is commonly defined as a direct index defining the
 8 characteristics of the observed gene pool, in other words, as the *size of a theoretical Wright-Fisher*
 9 *population of similar genetic characteristics* (Hamilton 2011). As we mentioned earlier (§40 p.
 10 101), in a population of constant size, and with a unit substitution rate, the probability of two
 11 lineages coalescing at generation t is $(1 - (1 / 2.Ne))^{(t-1)} \cdot (1 / 2.Ne)$. Provided the sample is
 12 large enough, this probability can be approximated by the observed density of coalescence events,
 13 and allows us to estimate the effective population size. Yet since substitution rate μ is, in reality,
 14 never equal to 1, this estimated population size is scaled by the substitution rate as $\vartheta = 4.Ne.\mu$ (in
 15 a diploid organism) - with the underlying idea that a small, but slowly mutating population will
 16 undergo the same amount of genetic drift as a larger, but faster evolving one.

17 Thus, assuming that substitution rate is constant through time (which is reasonable at the in-
 18 traspecific level), the distribution of times to coalescence throughout gene trees is an estimator of
 19 the probability of coalescence throughout the history of the population - and hence of the popu-
 20 lation size through time. This is the fundamental idea of the Skyline Plot approach (Fig. 10 p.
 21 107): in its simplest form, each coalescence interval in a single gene tree is converted by maxi-
 22 mum-likelihood into a population size during a time interval. Of course, the stochasticity of coa-
 23 lesence events is directly transferred to the population size estimates (as can be seen on the first
 24 row of Fig. 10 p. 107). Several approaches have been proposed to filter out this noise - ranging
 25 from simple ideas such as the generalised skyline plot (which imposes thresholds for taking
 26 changes into account, and filters out excessively quick «jumps» in population size), to more com-

1 plex ones line the Bayesian Skyline Plot (which tries to estimate a general demography with a low
 2 number of epochs, that maximises the likelihood of the observed gene tree - with a set of assump-
 3 tions, like the autocorrelation of the population size through time, reducing the demography
 4 space to be explored). More complex and accurate approaches have been proposed to better ad-
 5 dress the problem of estimating past population size changes (see §44 p. 112 to §47 p. 122), but
 6 most of them ultimately rely on this basic tie between coalescence rate and scaled effective popu-
 7 lation size.

8 A fundamental feature of the link between coalescence and population size, as it is currently con-
 9 ceived, is the irrelevance of gene tree topology - reconstructed topology is only a means to accu-
 10 rately estimate coalescence intervals. Migration, on the other hand, mostly affects the structure of
 11 gene trees through the topology of the coalescent. In geographically complex groups of individu-
 12 als, migration (as defined in §4 p. 30) interacts with population size in shaping the gene pool: in
 13 a given deme, genetic diversity may only arise through (i) mutation and drift (both functions of
 14 the effective population size) or (ii) migration. Unlike population size, however, migration is not
 15 estimated absolutely from the genetic data, but rather in reference to the present-day distribution
 16 of the sequences. In the coalescent framework, if we consider two present-day *demes* (see §2 p. 25
 17 for a definition), *migration* can be defined as the probability (as usual looking back in time) for a
 18 given lineage to move from its current deme to the other at the previous generation - *i.e.* the
 19 probability of any «effective individual» to have one of its parents being an immigrant. Thus, the
 20 *probability of coalescence* is, within each deme, the probability of a lineage merging with another
 21 from the same deme at a given generation (a function, as we exposed, of population size) - and
 22 the *probability of migration* is the probability of a lineage stemming back to another deme at the
 23 same period. Identifying migration through the coalescent structure across several demes has
 24 some limitations: in particular, it requires that deme identity is stable throughout species'
 25 analysed history (*i.e.* within $\sim 4N_e$ generations), and that during that time gene flow is low
 26 enough, that back-and-forth lineage migration is separated by at least one coalescence event.

- 1 Technically, accurate estimation is restricted to demes *stricto sensu* (see §2 p. 25), where mutation
- 2 and drift are major contributors to genetic diversity, an migration a marginal contributor.



3 **Figure 10 | Coalescence and population size.** Reconstruction of the past population size of two simulated
 4 populations after the original Skyline Plot model (Pybus *et al.* 2000). (A) Constant size population, and (B) Expo-
 5 nential growth population. *Bottom row:* simulated gene trees, generated using *msHOT* (Hellenthal & Stephens
 6 2007), *Middle row:* length of the coalescence intervals along the tree, *Upper row:* direct demographic function (black
 7 line) and generalised skyline plot (dashed blue line).

8 §-42 *The Ancestral Recombination Graph.* The coalescence process gives an accurate description of
 9 the evolution of a non-recombining sequence of DNA, for which, under an infinite-sites model,
 10 the only possible event is a divergence due to a mutation (when looking forward in time), *i.e.* a
 11 coalescence event when looking backwards in time (as presented in details in §40 p. 101). How-

1 ever, the recombination process interferes with the pure coalescence process when we consider
 2 longer sequences, and *a fortiori* genomic-scale data. During recombination events (usually the
 3 manifestation, at the sequence level, of the crossing-over of chromosomes during meiosis), se-
 4 quences sharing different ancestries (and thus different coalescent histories) become concatenated
 5 into one new sequence. By reshaping the associations between alleles, recombination is a major
 6 driver of genome evolution (Reich *et al.* 2001; Zhang *et al.* 2002), and must be accounted for
 7 when reconstructing the evolutionary history of a set of haplotypes. In the context of sequence
 8 evolution, recombination events can be seen as the counterpart of coalescence events: looking
 9 back in time, a coalescence event is the merging of two different sequences into one ancestral se-
 10 quence, while a recombination event is the splitting of one single sequence into two distinct an-
 11 cestral sequences.

12 The *Ancestral Recombination Graph* (ARG) is the representation of the full ancestry of a set of se-
 13 quences, that includes both coalescence, and recombination events (Rasmussen *et al.* 2014). Just
 14 like a classical phylogenetic tree, it is constituted of nodes, and edges - however, unlike a in phy-
 15 logenetic tree, nodes can be of two different types: *coalescent* nodes, or *recombination* nodes.
 16 When estimated over a very short stretch of non-recombining DNA, the ARG is essentially a co-
 17 alescent tree, but its complexity increases when it is estimated over longer, recombining se-
 18 quences. The full ARG of population-scale genomic data virtually encompasses all the possible
 19 information about population size changes, migration events, selection processes, *etc.* However,
 20 the complexity of the structure makes it extremely challenging to reconstruct and analyse, and it
 21 is mostly simplified down to a lower-dimensionality representation, such as the multi-dimension-
 22 al allele frequency spectrum (see §45 p. 117).

23 The Sequentially Markovian Coalescent (SMC, Wiuf & Hein 1999 - see §47 p. 122) is a trans-
 24 versal interpretation of the ARG. As a set of orthologous¹ sequences is considered transversally,

1. *I.e.* stemming from a single common ancestral copy.

1 the coalescent process is «pure», i.e. without recombination, at each single nucleotide position.
2 As long as successive nucleotides share the same ancestral history, a single coalescent tree holds.
3 However, a recombination event will introduce a change in the tree topology, as the ancestral re-
4 lationships between sequences are shuffled - and this new topology will hold until the next re-
5 combination point. This structure can be summarised as a Markovian process: a single coalescent
6 tree is a state in the process, a recombination event is a state change, and the topology of the coa-
7 lescent tree before the state change has no bearing on the topology after the state change. One
8 particularity of the SMC model is that it allows inference of multiple tree topology, *i.e.* coalescent
9 histories, when sampling along a single set of recombining sequences.

10 Sampling along the SMC can be conducted in several ways in order to infer population history:

11 (a) A «classical» approach relies on the extreme discretisation of the genome to sample virtually
12 independent states of the ARG. Thus, when sampling short sequences separated by very large
13 chromosomal distances, we can assume that 1) no recombination occurs within the sequences
14 under consideration, and 2) recombination occurred freely between the sequences, so that their
15 coalescent histories are fully independent draws from the population's ARG. Since, in that case,
16 sampled topologies are comparatively few and far in-between, it is possible to estimate the full
17 coalescent tree for each locus, and to draw information from every node. This is the approach im-
18 plemented in several multi-locus algorithms, for example the Extended Bayesian Skyline Plot
19 (Heled & Drummond 2008, see §44 p. 112), or the Migrate-n algorithm (Beerli 2006; Beerli &
20 Palczewski 2010).

21 (b) A continuous approach, on the contrary, is based on the reconstruction of the states and state
22 changes in the hidden SMC model through transversal scanning of the set of sequences. Thus, a
23 coalescent tree is reconstructed for each non-recombining interval between two state changes, de-
24 composing the full ARG into a set of coalescent-only graphs. This is the approach implemented
25 in the «SMC» family of models (see §47 p. 122) (c) the ARG may also be sampled indirectly,

Concepts and methods - §43

1 through a set of summary statistics - for example the distribution of allele frequencies , either rep-
2 resented as a spectrum (see §45 p. 117) or as single statistics, such as Wright's F-statistics or Taji-
3 ma's D (see §43 p. 110). These different approaches will be developed below.

4 *§-43 What is a summary statistic?* The full complexity of population genomic data could ideally be
5 represented without loss in a highly dimensional structure such as the Ancestral Recombination
6 Graph (see §42 p. 107). However, such an endeavour has limitations both in means, and in aims.
7 Accurately reconstructing a population's ARG would require flawlessly ascertained genomic data -
8 which we can not yet produce - and monumental computational resources (see Lam *et al.* 2010).
9 On the other hand, the full ARG is hardly intelligible by itself (Arenas 2013), and its major char-
10 acteristics need to be extracted in order to produce meaningful information. The need to reduce
11 the ARG to a simpler object therefore stems both from the failure to gather exhaustive data and
12 computational resources, and from the necessity of obtaining more easily comparable indices.
13 Thus, the common character of most population genetic approaches - several of which have been
14 devised long before the theorisation of the ARG - is to be in some respect a reduction of the full
15 graph into a lower-dimensionality object, such as a tree, a distribution, or a single statistic.

16 In the «classical» population genetics approach (see §28 p. 84), the nature of the markers used
17 (single mtDNA sequence or low number of nuclear loci) naturally makes the ARG a very distant
18 concept. The low information content of the data only allows for the estimation of «point» indi-
19 cators, not of the shape and extent of variation in signal across the genome. Thus, low-dimen-
20 sionality statistics have initially been devised to match the information content of the data that
21 was possible to produce: these focus on describing the amount and distribution of polymorphism
22 between a set of samples. As we mentioned above (§40 p. 101 and §42 p. 107), the present-day
23 distribution of genetic variation is a result of the history of *coalescence*, *recombination*, and *migra-*
24 *tion* within the species - if the sequences under scrutiny are not evolving neutrally, *selection* must
25 be added to that list, but neutrality is normally verified, or assumed. In the Wright-Fisher «drift»

1 framework (see §39 p. 99), on the other hand, only that final state is really considered and de-
 2 scribed at all.

3 A first and essential descriptor of an instantaneous state of genetic variation is the *density of segre-*
 4 *gating sites*, often called S , that is a measure of the depth of the local coalescent tree in a perfect
 5 Wright-Fisher constant-size population - and therefore an indirect estimator of \mathfrak{D} , the mutation-
 6 scaled population size, also described as Watterson's estimator \mathfrak{D}_W (Watterson 1975). However,
 7 Watterson's \mathfrak{D} technically discretises allele frequencies as a binary character (« minor allele frequen-
 8 cy > 0 »), and thus discards the obvious informative content of the observed allele frequency dis-
 9 tribution. Therefore, a more refined estimator is *nucleotide diversity*, normally designated as π ,
 10 and defined as the *average number of pairwise differences* in a set of sequences - *i.e.* the mean value
 11 of the *mismatch distribution* (Nei & Li 1979). Shifting from the polymorphism in a full set of se-
 12 quences to the average polymorphism in pairs of sequences sampled from that set, π retains infor-
 13 mation as to the allele frequency distribution in the sample, and is also an estimator of \mathfrak{D} under a
 14 neutral model, as \mathfrak{D}_π .

15 These two estimators of \mathfrak{D} become inaccurate when the true population departs from the Wright-
 16 Fisher model: interestingly, however, the inaccuracy of these indices does not arise in the same
 17 way, and is in itself informative. \mathfrak{D}_π is sensitive to the frequency of alleles, and intermediate-fre-
 18 quency variants will weight most in the estimator: thus, histories with a high density of ancient
 19 coalescence events will lead to higher values of \mathfrak{D}_π . \mathfrak{D}_W , on the other hand, since it handles allele
 20 frequencies in a binary way (*i.e.* « polymorphic» or « monomorphic»), is not sensitive to the de-
 21 pth of pairwise coalescence of sequences, and hence is not affected by the structure of the coales-
 22 cent: it only reflects the depth of the full local coalescent. Thus, the difference of \mathfrak{D}_π and \mathfrak{D}_W is an
 23 indicator of the neutrality of the coalescence process, as proposed by Tajima 1989: if both estima-
 24 tors are similar, a neutral constant-size coalescent process cannot be rejected, but if \mathfrak{D}_π is larger
 25 than \mathfrak{D}_W , we can infer an excess of ancient coalescence events, and therefore population growth.
 26 Other indicators (Fu's F_S , Fay and Wu's H , *etc.*) have been devised with a similar idea, all reflect-

1 ing, in some way or some other, skews in the coalescent structure, and consequently directional
2 departures from the neutral population model.

3 Another central (and still very much in use) application of point-value summaries of the ARG is
4 the quantification of population structure, or, in other words, of the amount of difference be-
5 tween allele frequencies in distinct sets of samples. The most common approach is the family of
6 «*F*-statistics» or *fixation indices* (Wright 1950), initially described in the synchronic drift frame-
7 work (see §39 p. 99), but that can be re-thought in the coalescent framework. In the original for-
8 mulation, the fixation index idea relies on the *Wahlund effect* (Wahlund 1928), a corollary of the
9 Hardy-Weinberg principle that implies that observed heterozygosity is reduced by sub-population
10 structure. By comparing the observed heterozygosity in a pair of populations to the expected het-
11 erozygosity under Hardy-Weinberg equilibrium across that combined set of individuals, one can
12 obtain a measure of the Wahlund effect, and hence of the strength of sub-population structure,
13 or population separation. Fixation statistics are declined at several levels (individual *vs.* total het-
14 erozygosity, *etc.*), and more refined estimators have been proposed for practical sampling cases, to
15 increase robustness to *e.g.* asymmetric sample sizes, or missing data (see Reynolds *et al.* 1983;
16 Weir & Cockerham 1984; Reich *et al.* 2009; Willing *et al.* 2012; Bhatia *et al.* 2013, and §129 p.
17 242), and are also the basis of the Analysis of Molecular Variance approach (Excoffier *et al.*
18 1992a). In the coalescent framework, fixation statistics are a function of the N-dimensional Allele
19 Frequency Spectrum (see §45 p. 117), and therefore a translation of the relative coalescence and
20 migration probabilities (see §41 p. 103).

21 Although summary statistics, by the low dimensionality, necessarily miss a large part of the infor-
22 mative content of genomic data, they have the important propriety of being easily comparable
23 across organisms, markers, or genomic regions (*e.g.* genome-wide statistics *vs.* mitochondrial
24 DNA statistics, *etc.*), and therefore to provide a more unified (with the trade-off of being simplis-
25 tic) picture of genetic variation.

1 §-44 *The Extended Bayesian Skyline Plot approach.* The Extended Bayesian Skyline Plot (EBSP)
2 method was proposed by Heled and Drummond (Heled & Drummond 2008), as a Bayesian
3 multilocus extension of the Skyline Plot (SP). In its original form (Pybus *et al.* 2000), the Skyline
4 Plot is a maximum-likelihood non-parametric method that exploits the relationship between
5 population size and coalescence probability (see §41 p. 103) to reconstruct effective population
6 size at different time points. The density of coalescence events is therefore indicative of the popu-
7 lation size (a higher rate of coalescence generally means a smaller population). However, the clas-
8 sical Skyline Plot has severe limitations: for example the fact that the gene tree is assumed error-
9 free, and estimated independently, or the fact that, as a single-locus approach, it may be severely
10 affected by the random sampling processes - a true random coalescent may diverge vastly from
11 the neutral expectation by mere chance (Ho & Shapiro 2011). The first limitation was overcome
12 by the replacement of the maximum-likelihood framework by a Bayesian approach (as Bayesian
13 Skyline Plot or BSP, Drummond *et al.* 2005), in which gene tree and demographic history are
14 co-estimated, and averaged over a very large Markov-chain Monte-Carlo (MCMC) sampling.
15 However, the locus-by-locus stochasticity limitation of the Skyline Plot remains with the BSP.
16 Another limitation of the BSP is that it is not fully model-flexible: the number of demographic
17 groups needs to be fixed *a priori*. The EBSP, as implemented in BEAST2 (Bouckaert *et al.* 2014),
18 overcomes all these major limitations. As a natively multilocus approach, EBSP relies on the
19 Bayesian co-estimation of gene trees and demographic histories from a set of unlinked loci, thus
20 considering large pannel of stochastic representations of the underlying «true» coalescent history
21 of the population. EBSP is also a truly model-flexible approach, in the sense that the number of
22 periods in the demographic function is informed solely by the genetic data. In its current form,
23 its main limitation is that, as it relies on the full likelihood of the DNA sequences, it is computa-
24 tionally intensive, and hardly scalable to full-genome-scale sequence data.

25 In this work, we exploit the specific features of RAD sequencing to deploy EBSP demographic
26 reconstructions (see Supplementary sections of *The King synnome*, p. 163, and of *The Emperor*

1 *synnome*, p. 211), as first proposed by Trucchi *et al.* (Trucchi *et al.* 2014). RAD sequencing (see
 2 §29 p. 85) allows the production of a large number of short sequenced loci that are widely inter-
 3 persed in the genome, so that they can be assumed unlinked. Within loci, recombination rate
 4 can also be assumed null, and the random sampling process in the genome allow us to consider
 5 the loci as neutrally-evolving (a condition rarely met in traditional multilocus approaches, that
 6 often rely on gene sequencing). Thus, the three main assumption of EBSP are *a priori* verified by
 7 RAD data. However, two characteristics of RAD sequencing require some adaptation: (i) the se-
 8 quencing process makes locus- and allele-dropout a significant problem and (ii) short loci neces-
 9 sarily have less segregating sites than longer loci, so that a selection must be made in order to base
 10 the analysis on informative data.

11 Allele dropout happens either as a result of cut-site mutation, that prevents one of the alleles to
 12 be sampled during the RAD library preparation (see §29 p. 85), or of under-sequencing, that
 13 causes one allele only to be sampled during the sequencing process (see §35 p. 93). The direct
 14 consequence of allele dropout is a mis-representation of true haplotype frequencies (see §35 p.
 15 93), that is likely to bias the reconstruction of the coalescent history of the population, through
 16 an over-representation of null branch lengths in the gene tree topologies. A simple workaround is
 17 to sample only one haplotype per individual: while halving the dataset, this guarantees that the
 18 distribution of coalescence events will be faithfully reproduced. Locus dropout is a different issue,
 19 that is more easily solved: both during library preparation and during sequencing, loci may be
 20 lost in some individuals. Thus, it may be difficult to find a sufficient number of informative loci
 21 that are sequenced in the same set of individuals. However, coalescent reconstructions are not in-
 22 dividual-, but rather population-centred: as long as panmixia is verified in the population, the
 23 identity of the individuals from which haplotypes were sampled has no bearing on the sampling
 24 of the haplotype pool. In order to avoid potential skews due to, for example, the sampling of re-
 25 lated individuals (i.e. individuals with correlated haplotypes at all loci), it may actually be desir-

1 able to choose haplotypes randomly from the full (haploid) population for each locus, rather
2 than to choose individuals.

3 The information content of loci is a different problem. Since RAD loci as considered neutrally
4 evolving, we suppose that the general genome background substitution rate applies to the
5 RADome. However, since substitution is essentially a rare, stochastic process, we expect the actu-
6 al number of substitutions occurring in a given genomic interval to follow a Poisson distribution,
7 with a parameter λ equal to the genome-wide background substitution rate (see §48 p. 124). Giv-
8 en the low substitution rate in long-lived vertebrates, most short (here 95 bp) loci are expected to
9 contain zero or 1 SNP. However, the general Skyline Plot rationale implies that several coales-
10 cence events must be reconstructed in order to obtain information on the variation of coalescence
11 probability (i.e. population size) through time. Indeed, in the absence of recombination, the
12 number of coalescence events is, at most, equal to the number of polymorphic sites: a locus with
13 3 SNPs is, at best, informative as to 3 coalescent nodes (less if multiple substitution occurred on
14 the same branch). In our experience, loci with less than 4 SNPs are not informative enough to ro-
15 bustly infer population history. In the same population, the number of SNPs observed at a locus
16 is a function of the number of haplotypes sampled and of the haplotype frequency distribution:
17 generally, sampling more individuals will increase the number of polymorphic sites discovered,
18 until all extant alleles have been represented. However, loci with excessive diversity are also suspi-
19 cious: in a Poisson process, the probability drops rapidly towards the upper tail of the distribu-
20 tion - in our case, very few loci are expected to have more than 7 or 8 SNPs. Processing artefacts,
21 on the other hand, such as a low-quality read, or collapsed paralogous loci¹, can have the effect of
22 increasing the observed number of SNPs at a given locus. For that reason, we restrict the analysis
23 to loci that have from 4 to 6 SNPs amongst 50 randomly sampled haplotypes. As this selection is
24 only made amongst neutrally evolving loci, the underlying substitution and diversification

1. That is, regions of the genome that stem from a common ancestral copy through a duplication event. Each paralogous copy further evolves separately, in a generally tree-like way. If the duplication event is recent, paralogs may retain enough similarity to be merged or confused during analysis, resulting in a chimæric, non-tree-like ensemble.

1 processes are not different from those at work in the rest of the RADome. By doing so, we are
2 only filtering the stochastic output of the substitution Poisson process.

3 This locus selection also has implications for the parametrisation of the EBSP model. EBSP, like
4 any other model, may be easily subject to over-parametrisation: however, due to the huge compu-
5 tational burden of the MCMC algorithm, precisely estimating the number of necessary parame-
6 ters (through, e.g. Bayes factor comparison of path-sampled alternative models' marginal likeli-
7 hoods) is a difficult task. However, the structure of RAD data allows for a particularly efficient
8 parametrisation. In Trucchi *et al.* (Trucchi *et al.* 2014), an independent substitution model was
9 defined for each locus in the analysis: even reducing the sites model to a simple HKY+G with
10 empirical base frequencies, a substitution rate and a transition/transversion ratio (*kappa*)¹ had to
11 be defined of each locus (or 100 parameters for a 50-locus analysis), each of them informed with
12 very little data. On the other hand, if we assume the RADome to be a globally neutral ensemble,
13 we may assume that a single substitution model holds for all loci: we need only, in theory, define
14 one substitution rate and one *kappa* parameter for all 50 loci, informed this time by the full
15 dataset. Substitution rate, however, is strongly biased in our model by the selection of loci with
16 higher observed polymorphism: although the full RADome verifies the expectation of a Poisson-
17 distributed number of polymorphisms per locus with a single parameter, we use a truncated dis-
18 tribution that is not accounted for by a single substitution rate. Therefore we use a single *kappa*
19 parameter for all sites, but a different substitution rate for each class of loci (by «class», we mean
20 whether the locus includes 4, 5 or 6 polymorphic sites). This reduced parametrisation scheme has
21 two advantages: since we use 4 instead of 100 parameters, MCMC convergence time is greatly
22 reduced. And since we obtain a single consensus substitution rate for each locus class, it is possi-

1. A transversion being the replacement of a purine base (A or G) by a pyrimidine base (C or T), and a transition the replacement of a purine by the other purine, or of a pyrimidine by the other pyrimidine. Thus, there are 8 possible transversions, and only four possible transitions. The ratio of the transition to transversion rates, or κ (*kappa*) is characteristic of a mutation model.

1 ble to link the observed RAD substitution rates, and the whole-genome background substitution
 2 rate (see §48 p. 124).

3 We use this much simplified parametrisation in *The King synnome*, p. 163, and *The Emperor syn-*
 4 *nome*, p. 211, and find that the slow MCMC convergence issues encountered by Trucchi *et al.* are
 5 mostly solved, and that repeatability is high between different random sets of loci. However, large
 6 improvements could be achieved by designing a RAD-specific parametrisation for the EBSP, that
 7 would take into account the observed distribution of polymorphisms in the whole RADome, and
 8 use its λ parameter as an underlying rate for the loci included in the analysis.

9 *§-45 The Allele Frequency Spectrum approach.* The Allele Frequency Spectrum (AFS) is a rich sum-
 10 mary of the Ancestral Recombination Graph. This data structure gathers the counts (or propor-
 11 tions) of sites in the genome for which the derived allele has a frequency f in the population un-
 12 der consideration. While the «true» AFS for a population could only be compiled through
 13 exhaustive sampling, the AFS can also be estimated from a random subset of individuals sampled
 14 from a panmictic population. Thus, the maximum resolution of the reconstructed spectrum will
 15 be $1/(2S+1)$ (with S being the number of sampled diploid individuals). Since usually the sample
 16 size is much smaller than the effective population size, only alleles that are already present in a
 17 significant fraction of the population (*i.e.* $f \gtrsim S/Ne$) are effectively represented in the spectrum.
 18 Under its most simple expression, the AFS can be reconstructed directly from allele counts in the
 19 sample, with the observed frequency taken as an estimator of the true frequency in the popula-
 20 tion. However, the difficulty of ascertaining genotypes reliably at a large number of loci, in a large
 21 number of individuals, makes such a direct estimation rather error-prone. Maximum-likelihood
 22 estimates of the AFS, that take into account the possibility of genotype errors, and the impact of
 23 missing data, are used whenever possible, and have been shown to provide a much more robust
 24 estimate of the «true» population AFS (Nielsen *et al.* 2012). The ML estimate of the AFS is not a
 25 direct transformation of the observed sites counts, and should better be interpreted as the distrib-

1 uation of probabilities, for a given allele, to be found with a frequency f in the population - here,
 2 we used this latter method as implemented in ANGSD (Korneliussen *et al.* 2014).

3 Another difficulty of AFS estimation lies in the correct identification of an ancestral and derived
 4 states at each position. Accurately estimating the full AFS for a population requires that we can
 5 determine whether a given site should be considered as having a frequency $f = 1/2N_e$, or rather f
 6 $= 1 - (1/2N_e)$. In order to do this, it is necessary to reconstruct the state at each position in the
 7 genome at the parent node for the taxon, thus requiring sequence and polymorphism data for a
 8 sister taxon, and for an outgroup. If this is not possible, the AFS may be *folded*: instead of consid-
 9 ering the frequency of the *derived* allele, one considers the frequency of the *minor* allele (the allele
 10 with the lesser frequency - which raises additional problems when considering multiple popula-
 11 tions). Although this approach removes the errors due to ancestral state mis-identification, it
 12 greatly increases the number of possible scenarios leading to the same observed AFS, thus reduc-
 13 ing the robustness of inferences.

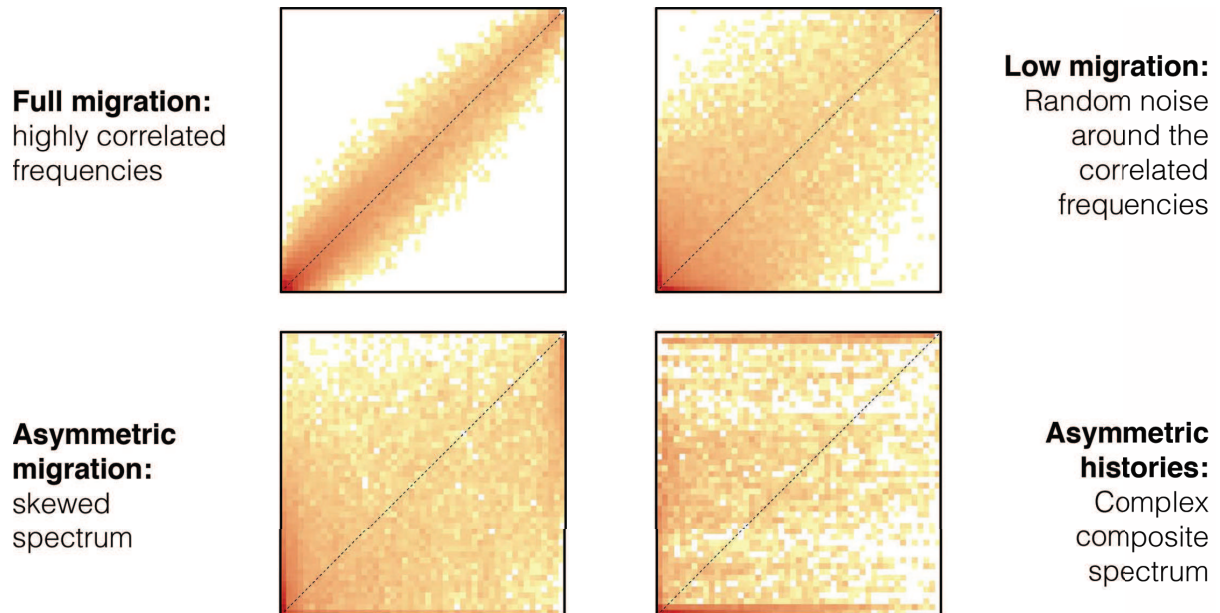
14 The AFS is highly informative as to demography and gene flow. When starting from a single an-
 15 cestral sequence, each new mutation appears as single copy in the whole population, *i.e.* with a
 16 frequency $f = 1/(2N_e)$. If the fitness value of that new mutation is neutral, its frequency is expect-
 17 ed to vary only due to the stochastic process associated with binomial sampling at each genera-
 18 tion, with variations in population size being a major forcing on allele frequency drift. For exam-
 19 ple, a population bottleneck will have the immediate effect of removing rare variants from the
 20 population, and biasing the AFS towards intermediate allele frequencies. Thus, the AFS for a giv-
 21 en population reflects the successive demographic events in the population. But more informative
 22 is the N-dimensional AFS, that summarises the distribution of alleles frequencies amongst N
 23 populations, with $P(i,j,k)$ being the probability, for a given allele, to have the frequency i in pop-
 24 ulation I, the frequency j in population J, and the frequency k in population K. If populations I,
 25 J and K are three random subsets of one single panmictic population, we expect the frequencies i ,
 26 j , and k to be highly correlated for all sites, since the only source of variation is the stochastic er-

1 ror due to the sampling process. Isolation between populations, on the other hand, will manifest
 2 itself as a decay in the correlation between allele frequencies between these populations. The N-
 3 dimensional AFS (or « multi-AFS ») is therefore a result both of the populations' demographic
 4 history, and migration history (see Fig. 11 p. 120).

5 Whereas the ARG explicitly represents population history by integrating time as a variable in its
 6 topology, the AFS is a « *temporally flat* » object, that only contains the outcome of demographic
 7 and migratory events. Inferring population history from the AFS therefore requires a more elabo-
 8 rate procedure than would be needed using an explicit coalescent history. The first approach that
 9 has been proposed to exploit the full N-dimensional AFS is that of Gutenkunst *et al.* 2009. This
 10 approach, implemented in $\delta a \delta i$, uses a diffusion approximation to represent the AFS: the discrete
 11 structure of real populations (discrete number of individuals - hence discrete frequency classes -
 12 and discrete generations) is approximated in continuous time, with continuous frequency transi-
 13 tions, as proposed by Kimura 1955 (see above §39 p. 99), so that demographic parameter values
 14 become parameters in a diffusion equation, of which the observed AFS is the numerical solution.
 15 In order to find the most likely parameter values, $\delta a \delta i$ uses a composite-likelihood approach (*i.e.*
 16 a method where the true log-likelihood function is approximated by a simpler function, see Lind-
 17 say 1988), and the uncertainty of the results is assessed through parametric or non-parametric
 18 bootstrapping (*i.e.* either through resampling of the AFS itself, or by simulating new datasets
 19 based on the estimated parameters, and estimating new parameters based on these simulations -
 20 both approaches described by Gutenkunst *et al.* 2009).

21 Recently, a different approach has been suggested to infer population history from the observed
 22 N-dimensional AFS, based on coalescent simulations. This approach, proposed by Excoffier *et al.*
 23 2013, overcomes the previous limitations of the simulation-based methods by implementing a
 24 continuous-time coalescent simulation (as opposed to the « natural » generation-by-generation co-
 25 alescent, too slow to be used for demographic inference). Thus, the composite likelihood of the
 26 model is calculated based on a large number of simulated coalescent histories and resulting AFS -

1 typically 100,000 simulations per step in an Empirical Bayes approach, in the *fastsimcoal2* imple-
 2 mentation (Excoffier & Foll 2011; Excoffier *et al.* 2013). Thus, as opposed to the $\delta a \delta i$ approach,
 3 the *fastsimcoal2* approach explicitly models the coalescence process that underlies the observed
 4 AFS.



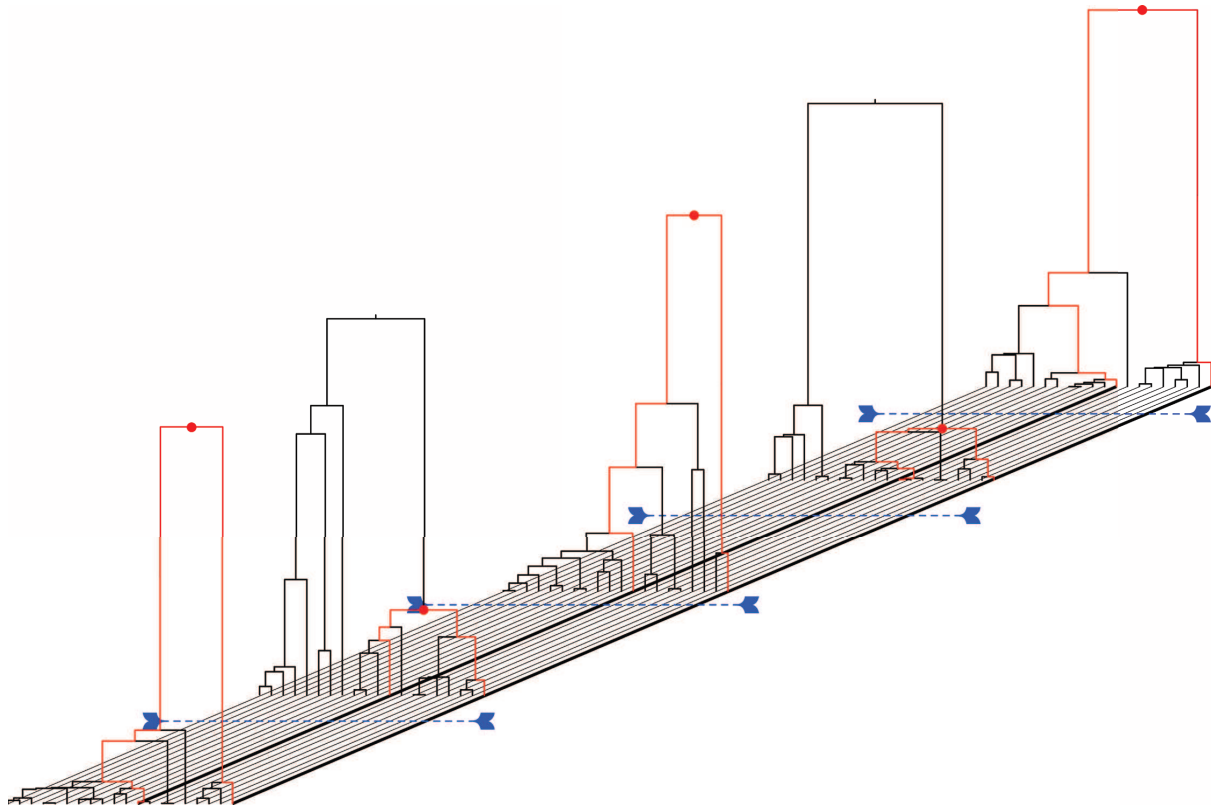
5
 6 **Figure 11 | Influence of demographic processes on the joint AFS.** Four different pairwise situations,
 7 simulated using *fastsimcoal2*. Under **full migration**, allele frequencies are highly correlated between the two popula-
 8 tions. With **low migration**, random dispersion increases around the full correlation line. With **asymmetric migra-**
 9 **tion**, dispersion becomes skewed, and the correlation departs from the diagonal. Our last example is an **asymmetric**
 10 **history**, where the first population(vertical marginal spectrum) underwent a strong bottleneck biasing it towards in-
 11 termediate allele frequencies, while the second population remained stable.

12 One limitation of both methods is that they are strongly model-constrained: the parameter space
 13 has to be carefully defined from prior knowledge. Since possible population histories can be ex-
 14 tremely diverse, and since the AFS does not represent the temporal component of the ARG, it is
 15 necessary to define a demographic function, with a fixed set of free parameters that will be opti-
 16 mised by maximum composite likelihood - these parameters including the timing of major
 17 events, the population size and growth rate before and after those events, and the migration rates

1 between populations. Thus, the general structure of the population history is set beforehand, and
 2 only the value of the parameters may be optimised over. Although it is possible to compare differ-
 3 ent historical models (normally through AIC model choice), it may happen that only models that
 4 are fully incompatible with the true underlying history have been tested: thus, strong external ev-
 5 idence is required to formulate relevant historical hypotheses before testing them.

6 *§-46 The Stairway plot approach.* The Stairway plot (STY) model has been recently proposed by
 7 Liu & Fu 2015 as a model-flexible alternative to the general AFS-based methods (see §45 p.
 8 117). By considering only a 1-dimensional AFS, marginalised for a single population, the STY
 9 excludes all information relative to historical migration rates, and focuses on population size
 10 change events, which allows for a considerable simplification of the parameter space. Thus, in-
 11 stead of requiring the specification of a number of demographic events, epochs, or growth rates,
 12 it is able to use a piecewise-constant population size prior, with a flexible number of population
 13 size changes informed by the data (through a likelihood ratio test, as models with decreasing
 14 numbers of demographic epochs are fully nested within each other) - a characteristic shared with
 15 the EBSP approach (see §44 p. 112, although EBSP includes the number of population size
 16 changes as a parameter to be co-estimated together with substitution models, gene trees and pop-
 17 ulation sizes during the MCMC sampling, rather than a likelihood ratio test).

18 Instead of explicitly modelling recombination events (as the PSMC approach does, see §47 p.
 19 122), the STY assumes each single variable site to be independent, and to stem for a separate coa-
 20 lescent process (a simplification defining the composite likelihood function). The probability of
 21 a site having a derived allele frequency f is a function of population size at each point of time in
 22 the past: thus, we can search for population size values (mutation-scaled as θ) that maximise the
 23 composite likelihood of the observed spectrum. Although the STY approach does not explicitly
 24 model the full coalescent process as EBSP does, it is able to use information from a much larger
 25 number of loci, and thus may better capture the information contained in the full dataset.



1

2 **Figure 12 | The Pairwise Sequentially Markovian Coalescent model.** We represent the independent
 3 local trees along a set of recombining sequences. Each blue arrow figures a recombination event, or state change in
 4 the hidden Markov model. The bolder lines are the two sampled chromosomes, and their pairwise coalescence is out-
 5 lines in red in the local trees. The result of the pairwise sampling is a distribution of TMRCA's (see §47 p. 122 for
 6 details).

7 *§-47 Sequentially Markovian Coalescent approaches.* Another, independent take on the problem of
 8 estimating past population size changes has been proposed by Li & Durbin 2011, and recently
 9 extended by Schiffels & Durbin 2014. In its original form, the Sequentially Markovian Coales-
 10 cent (SMC) model is a transversal interpretation of the classical coalescent, with recombination,
 11 introduced by Wiuf & Hein 1999, and can be interpreted as a reduction of the general Ancestral
 12 Recombination Graph (see §42 p. 107), in which recombination density is retained, but recom-
 13 bination topology is discarded. Considering that the distribution of times to the most recent
 14 common ancestor (TMRCA) along two chromosomal sequences follows a Markov model, where

1 changes of states represent recombination events (McVean & Cardin 2005; Li & Durbin 2011),
 2 the pairwise sequentially Markovian coalescent approach treats the «true» coalescence and re-
 3 combination history of a pair of chromosomes as a hidden Markov model, where the hidden
 4 states are the TMRCA, the state transitions the recombination events, and the observed states are
 5 the densities of SNPs along the sequence (a higher density implying a more remote MRCA - see
 6 Fig. 12 p. 122). Past population size is assumed piecewise-constant, but can change state at time
 7 intervals: these intervals need to be specified beforehand, as opposed to the approach implement-
 8 ed in the STY (see §46 p. 121). Considering the number of recombination events in a typical
 9 population, any pair of chromosome contains thousands of independent coalescent samples.
 10 However, recent history is difficult to retrieve, since fewer recombination events can be detected,
 11 which is the main shortcoming of the method. To overcome this problem, the multiple sequen-
 12 tially Markovian coalescent approach (MSMC, Schiffels & Durbin 2014), considers more than
 13 two haplotypes at a time. Thus, instead of considering the TMRCA between a pair of chromo-
 14 somes as an indicator of the ARG, it takes the shallowest TMRCA amongst any pair of chromo-
 15 somes in the sample as its hidden Markov state. The structure of the coalescent (see §40 p. 101)
 16 implies that this TMRCA tends to become shallower as the sample size becomes larger. Thus,
 17 switching from PSMC to MSMC increases considerably the resolution over the most recent peri-
 18 ods. A particularity of the MSMC algorithm, consequently, is the link between sample size and
 19 reconstruction depth. Since MSMC focuses on the first coalescence event between any two se-
 20 quences in the sample, a larger sample size will increase the amount of information about recent
 21 coalescent structure, at the cost of older events, while reducing sample size will increase accuracy
 22 for deep-time events (Schiffels & Durbin 2014). However, this requires precise phasing of each
 23 haplotype, which is a considerable challenge in non-model species (see §37 p. 97): indeed, recon-
 24 structing the local mismatch distribution is impossible without an accurate representation of the
 25 local haplotypes. If phasing is not accurate, the precise local distribution of pairwise differences
 26 will be averaged to its mean value, and the estimated local MRCA will consistently be closer to

1 the population's average MRCA, thus losing nearly all the informative content of the distribu-
 2 tion of TMRCAs along the genome. Thus, the MSMC seems more suited for cases where the
 3 phasing can be established with high confidence - which is nearly never the case in non-model
 4 species (see §37 p. 97).

5 *§-48 Rate of evolution.* Accurate reconstruction of past and present population size changes re-
 6 quires an accurate estimate of the substitution rate - a notoriously difficult parameter to assess
 7 (Warnock *et al.* 2012; Warnock *et al.* 2015). A robust estimate of the substitution rate of the mi-
 8 tochondrial hyper-variable control region (HVR) or the Adélie penguin, based on mitochondrial
 9 heteroplasmy transmission through lineages, has been proposed by Millar *et al.* (Millar *et al.*
 10 2008). In order to infer the genome-wide evolution rate, we performed a joint analysis of mito-
 11 chondrial HVR and RAD data in a multilocus Extended Bayesian Skyline Plot framework, using
 12 the HVR rate as a calibration, by setting a normally-distributed prior reflecting the published
 13 point estimate and 95%CI interval, with a high sampling rate along the MCMC (10x the sum of
 14 sampling rates for all other substitution rates) in order to ensure free sampling of the prior.

15 In a coalescent framework, we need to assess the frequency of coalescence events along time:
 16 thus, calibration must address the rate of transmitted mutations only, i.e. the species' substitution
 17 rate, as opposed to its intrinsic mutation rate (a purely physiological parameter of little relevance
 18 here). However, the germ line cell divisions are, within a taxon, independent from the generation
 19 time - thus, what we may assume constant is the number of substitutions per generation, rather
 20 than per units of calendar time. Since the generation time differs widely between the Adélie pen-
 21 guin (6.46 years, see Millar *et al.* 2008) and the King penguin (10.48 years, see §95 p. 186), we
 22 must thus convert the established rate (in substitutions per site per Myr: median = 0.55, 95% CI
 23 = 0.29–0.88, see Millar *et al.* 2008) in order to reflect the generation time of the King penguin,
 24 to 0.34 substitutions.site⁻¹.Myr⁻¹ (95% CI = 0.18–0.54).

1 The EBSP analysis was setup as exposed in *The Emperor synnome*, p. 211, and is briefly reminded
 2 here. We downsampled the data to haploid individuals, and using independent 50 loci with 50
 3 haplotypes each, with 3 to 6 polymorphic sites, in addition to 50 randomly selected HVR hap-
 4 lotypes. We specified one independent site model for each locus class (3, 4, 5 or 6 SNPs, and
 5 HVR). For each class, specified a HVR model, allowing for invariant sites for the HVR, but not
 6 for the short nuclear loci, and for gamma-distributed rate heterogeneity discretised in 4 classes.
 7 Transition-transversion ratio *kappa* was linked across nuclear models. All chains were run in du-
 8 plicate to check for convergence and for a sufficient length to gather $ESS^1 > 200$ for all parame-
 9 ters, which necessitated 500,000,000 to 1,000,000,000 steps.

10 Since we parametrised each locus class separately, we expect our model to fit a class-specific sub-
 11 stitution rate as a function of the observed number of segregating sites, rather than a common
 12 substitution rate. However, as we focus on neutrally evolving regions of the genome, we expect
 13 the number of segregating sites to follow a Poisson distribution, of parameter λ equal to the mean
 14 number of segregating sites per RAD locus. On a large number of sequences, the expected value
 15 $E(\lambda)$ converges towards the “true” underlying constant mutation rate, multiplied by the total tree
 16 length for each locus. Thus, if we fix the tree length, λ becomes an estimator of the substitution
 17 rate μ . However, under the EBSP model, the observed number of segregating sites is taken as an
 18 estimator of λ , and consequently of the substitution rate μ . Therefore we expect the inferred value
 19 of μ for each locus class to be a posterior probability of the “true” substitution rate, conditional
 20 on the mean number of segregating sites observed for that class (Trucchi *et al.* 2014). In order to
 21 retrieve the underlying common substitution rate μ , we first fitted a log-linear model to the in-
 22 ferred substitution rates ($\mu_3= 0.0159$, $\mu_4= 0.0218$, $\mu_5= 0.0275$, $\mu_6= 0.0389$. Fitted model: inter-
 23 cept $i = -5.02$, slope $s = 0.292$, $R^2=0.997$). A Poisson model of parameter λ equal to the mean ob-
 24 served number of segregating sites was a good fit for the empirical distribution of number of

1. ESS or *Effective Sample Size* is the number of uncorrelated states sampled along the Markov chain. Correlation between successive states indicates a trend in the state of the model, and hence incomplete mixing. Thus, the number of steps needed for the correlation between states to decay, or *lag vector*, is indicative of the efficiency of the mixing for the MCMC.

1 segregating sites per locus ($\lambda=1.47$, chi-squared test of goodness-of-fit p-value= 0.232). Thus, we
2 extracted μ as $e^{(s\lambda+i)} \sim 1.02e-2$ substitutions per site per Myr, or $1.08e-7$ substitutions per site per
3 generation. This rate is ca. twice slower than the one reported by Trucchi *et al.* ($2.6e-7$ sub-
4 st.site⁻¹.generation⁻¹, see Trucchi *et al.* 2014), but much faster than the one reported by Li *et al.*
5 ($8.11e-9$ subst.site⁻¹.generation⁻¹, see Li *et al.* 2014). While the former was not used in Trucchi *et*
6 *al.*'s analysis, but rather derived from it, Li *et al.*'s result, on the other hand, relies on two exterior
7 and uncertain assumptions: 1) the divergence time between the Emperor and the Adélie penguin
8 is set to ~ 23 Myr, which may be a large overestimate (based on a state-of-the-art total evidence
9 bayesian analysis, Gavryushkina *et al.* 2015 proposes ~ 9 Myr instead), and 2) the generation
10 time is taken to be 5 years in both species; however it has been shown to be 16 years in the Em-
11 peror penguin (Jenouvrier *et al.* 2014), and 6.46 years in the Adélie penguin (Millar *et al.* 2008) -
12 thus 11 years would be a closer (although inaccurate because assuming a single rate) estimate of a
13 common generation time. Applying these corrected estimates to Li *et al.*'s findings would give a
14 rate of $\sim 4.55e-8$, which is more than five times faster than proposed, and ca. half our estimate.
15 And still, this calculation does not take into account the possible rate heterogeneity between lin-
16 eages, and most importantly the changes in generation time between the *Aptenodytes/Pygoscelis*
17 common ancestor and the extant species, which may explain the remaining difference.

18 Generally, the rate of evolution of penguins has been a rather challenging subject, with a wide
19 discrepancy between the paleontological and molecular evidence (see §1 p. 65). While fossil data
20 has been recognised to support a very recent radiation of penguins (about 10 Myr BP, see Slack *et*
21 *al.* 2006), molecular data has been interpreted as implying a much more ancient origin (~ 45 Myr
22 for Baker *et al.* 2006). This molecular-derived radiation has successively been brought to a closer
23 agreement with the fossil evidence by Subramanian *et al.* (~ 20 Myr, Subramanian *et al.* 2013)
24 and Gavryushkina *et al.* (~ 12.5 Myr, Gavryushkina *et al.* 2015). The rate that we propose here is
25 in accordance both with the hypothesis of a very fast diversification of the spheniscids, and with
26 the findings of Trucchi *et al.*

1 *Habitat and climate analysis.*

2 §-49 *Habitat and niche.* The set of external conditions required by a species to thrive define its
3 *habitat*. Generally, these conditions include both *abiotic*, and *biotic* factors - *i.e.* factors that are
4 only dependant on (a) physical and chemical characteristics, or (b) interactions with other living
5 organisms. Abiotic habitat determinants usually include temperature and precipitation mean and
6 variability, a range of elevations, sometimes wind or currents regime, but also the length and vari-
7 ability of the photoperiod or the intensity of insolation. These parameters (although some, such
8 as temperature and precipitation, are themselves in a certain measure influenced by biotic com-
9 munities), may be accurately estimated, and predicted on medium time scales, and are the focus
10 of general circulation models (see §50 p. 130). Their impact on biological system is multiple.
11 They primarily interact with species' physiology: each organism has an optimum set of abiotic
12 conditions in which its metabolic processes are least costly, and in which it thrives. But they also
13 impact species indirectly, through direct effects on other organisms which themselves interact
14 with the focal organism - in other words, through the mediation of *biotic interactions*. For exam-
15 ple, prey or food resource availability, predation pressure, vegetal cover for shelter, parasites and
16 pathogens, *etc.* play a central role in defining the suitable habitat for a species. The respective
17 parts of abiotic and biotic factors in the definition of a habitat are debated: their contribution is
18 certainly species-dependent, but may also vary across the different margins of a species distribu-
19 tion - for example, it has been proposed that the polar boundaries of habitats are more often de-
20 fined by abiotic limits (*e.g.* physiologic effects of the coldest winter temperatures), while the equa-
21 torial boundaries are rather defined by biotic constraints (such as a n i ncreasing d ensity of
22 competitors and parasites equatorwards - themselves prevented from a polewards spread by their
23 own abiotic limits). Biotic and abiotic factors may also interact across phases (*e.g.* overwintering
24 and summer grounds, foraging and breeding areas, *etc.*), or evolve through time - a habitat main-

1 ly determined by biotic factors may come under increasingly strong abiotic constraints as a result
2 of climate change.

3 The geographical extent of a species' habitat defines its *potential range* - that is, the range it would
4 occupy if biological systems were fully mechanistic. Several factors may prevent a species from oc-
5 cupying its whole potential range: for example, suitable habitat may occur outside of the species'
6 dispersal range, or the species may have gone locally extinct through hunting pressure, *etc.* The
7 actual range occupied by the species, which is necessarily fully nested within its potential range, is
8 its *realised range* (Hutchinson 1957) - the one provided in atlases and distribution maps (*e.g.* Fig.
9 7). A mismatch between the *potential* and the *realised* range is common at the global scale, espe-
10 cially in continental taxa: while similar ecosystem are normally found zonally around the world,
11 dispersal across ocean basins is a rare event for larger organisms. Thus, the a bundant endemic
12 fauna of Australia and New Zealand owes more to the lack of connectivity between similar habi-
13 tats than to restricted potential ranges. Indeed, the ease and rapidity of biological invasions shows
14 that once the (often punctual) barrier-to-dispersal problem is solved (*e.g.* once the ocean has been
15 crossed in some way), whole swathes of new habitat may be colonised rapidly (Thuiller *et al.*
16 2005; Trucchi *et al.* 2015), bringing the realised range closer to the potential range. The dispersal
17 problem is less drastic in marine systems (see §6 p. 33), and is most important meridionally -
18 while connectivity is rather large between ocean basins at the same latitudes (especially at higher
19 latitudes were connectivity is enhanced by inter-ocean circulation, see §7 p. 35), the equatorial
20 region may constitute an important barrier between the Arctic and the Antarctic, resulting in ful-
21 ly independent organisms occupying similar ecological functions and displaying convergent traits
22 (for example the auklets and the diving-petrels, see §20 p. 71).

23 This convergence, in similar habitats, of species fulfilling similar ecosystem functions, allows us to
24 define the concept of *niche* (Hutchinson 1957). Whereas a *habitat* is a species-focused, and large-
25 ly geographical concept («is that place, given its biotic and abiotic constraints, suitable for that
26 species ?»), a *niche* is a functional, ecosystem-focused concept («is there already a species here

1 that can eat anchovy in the coastal waters while nesting on these scree slopes ?»). Although both
2 concepts are often confused, they are extremely different. Modelling a niche would rely on explic-
3 it modelling of each relevant species and their relationships in the ecosystem (Wiens *et al.* 2009;
4 Chevin *et al.* 2014) and is very seldom done in climate change biology. Modelling a habitat is a
5 slightly easier task, as the problem may be restricted with a reasonable level of approximation to
6 non-functional aspects (*e.g.* static distribution of abiotic parameters, and a few key biotic parame-
7 ters such as prey abundance). Yet even within the approximation of habitat modelling, a funda-
8 mental distinction arises between fully *correlative*, and what we could call *causal* models. The
9 overwhelming majority of studies of species' response to climate change use the «climate enve-
10 lope» or «climate space» approach (Box 1981; Berry *et al.* 2002): through a multivariate correla-
11 tive approach, the observed distribution of a species is correlated to a set of (usually abiotic only)
12 environmental parameters, and this set is used to project the species' future distribution (*e.g.*
13 Virkkala *et al.* 2008; Durner *et al.* 2009) or demography (*e.g.* Jenouvrier *et al.* 2012; Jenouvrier *et*
14 *al.* 2014) based on numerical climate models predictions. One obvious limitation of this ap-
15 proach is that it takes the species *realised range* as an estimator of its *potential range* - in some cas-
16 es, this can be just an approximation, but in some other it can be a true error (Pearson & Daw-
17 son 2003). This becomes even worse when only abiotic habitat characteristics are included in the
18 model (as it is the case in the 2015 Audubon society report on future North American bird dis-
19 tributions, for example - see Langham *et al.* 2015), since biotic constraints are expected to reduce
20 the set of abiotically exploitable areas. Thus, the abiotic habitat estimated from the realised range
21 is very likely to underestimate the range of abiotic conditions that a species can tolerate, as it con-
22 fuses the abiotic limitations of the biotic interactors with the abiotic limitations of the focal
23 species itself. For example, a species may be limited to a very little part of its otherwise potential
24 habitat by the presence of a strong competitor: the future distribution of that species will then
25 mostly depend on the future evolution of the competitor, and not of the local climate. If, by
26 chance, the competitor is less tolerant to changes in abiotic conditions, then climate change may

1 manifest itself as a range extension of our example species. If we had in the first place modelled
2 that species habitat by hypothesising that it can only tolerate the precise abiotic conditions in
3 which it is found, it is very likely that this overly precise set of conditions will be disrupted by
4 any change of climate, and that we will forecast that species' extinction. The same goes to demo-
5 graphic, rather than distributional, modelling: there again, observed correlations are assumed, but
6 not shown, to imply a causal relationship between environmental parameters and the species'
7 growth rate. However this line of thought is usually flawed, as has been convincingly exemplified
8 by Trivelpiece *et al.* 2011 using the Chinstrap penguin *Pygoscelis antarctica* as a model: the pen-
9 guin's breeding success is explained primarily by krill abundance, and the krill abundance is part-
10 ly explained by sea ice extent. A climate-envelope model would therefore find a rather robust cor-
11 relation between penguin breeding success and sea ice extent. However a disruption of the
12 correlation between sea ice and krill can be introduced by cetacean demography (cetaceans are
13 important competitors, and their demography is also controlled by lower-latitude breeding-
14 ground conditions) or by human exploitation (a notoriously difficult parameter to forecast in our
15 days): thus, the observed correlation between sea ice and penguin demography is in fact falla-
16 cious, and may change abruptly whenever cetacean demography or human exploitation of krill
17 will change regime.

18 A *causal* approach to modelling, on the other hand, will endeavour to identify parameters that
19 have a know, direct influence on the species' demography. In the case of the Chinstrap penguin,
20 this would mean including krill abundance, and not sea ice extent, as an explanatory variable.
21 This is the approach we use in *The King synnyme*, p. 163, as we are convinced it allows for much
22 more robust forecasts of population trajectories.

23 *§-50 Climate projections in the Coupled Model Intercomparison Project.* Our current understanding
24 of the Earth's climate system is summarised in Earth System models, that mainly implement
25 General Circulation Models (GCM) - i.e. numerical models which reproduce energy transfer be-
26 tween atmospheric, oceanic, land and ice masses. These models usually involve a discretisation in

1 cells of varying dimensions, and the state of each cell is computed for each time step as a function
2 of the other cells in the system. The Coupled Model Intercomparison Project (CMIP) proposes
3 an international framework to assess and understand the differences between the representation
4 of climate in different models, by defining standard experiments that can be compared between
5 systems. The current framework, the CMIP phase 5 (CMIP5, detailed in Taylor *et al.* 2012),
6 builds upon the previous CMIP3 ensemble to achieve more consistent and accurate model pre-
7 dictions. In particular, CMIP5 models have a generally higher resolution, and a more accurate
8 representation of bathymetry, than the previous generation of models, which leads to a much im-
9 proved simulation of the Southern Ocean (Meijers 2014). While most CMIP5 models perform
10 generally well, sea ice representation in the Antarctic Ocean remains a weakness, as most models
11 predict a slight decrease in sea ice extent in East Antarctica in the early 21st century, while the
12 observed trend is a slight increase (Zhang 2007; Turner *et al.* 2013; Xu *et al.* 2013; Wang *et al.*
13 2014) - a discrepancy that may be linked to different factors. First, most current GCMs treat
14 ocean, atmosphere and sea ice as three separate models, with boundary interfaces at the surface of
15 the ocean, and the upper and lower surfaces of the sea ice layer. Relative to its own internal dy-
16 namics, boundary conditions are thus comparatively very important for sea ice, and the current
17 approach may lead to incorrect representation. Secondly, the spatial resolution of most CMIP5
18 models (usually a $1^\circ \times 1^\circ$ grid) may be too coarse to explicitly model latent-heat polynya, that act
19 as sea-ice formation zones (see §118 p. 225 and Kushara *et al.* 2015). But lastly and most im-
20 portantly, the input of freshwater (with a higher freezing temperature) from the melting conti-
21 nental ice sheet has an important impact on the sea ice formation: however, the Antarctic ice
22 sheet itself is not modelled in the CMIP5, but merely treated as a specified boundary condition,
23 which is likely to lead to an incorrect representation of sea ice (see Bintanja *et al.* 2015 and §118
24 p. 225).

25 As CMIP5 models include different representations of climatological phenomena, their outputs
26 may vary, both for the reconstruction of past climate, or for the prediction of future changes. A

1 common way to handle these divergences is the *ensemble* approach, in which several different
 2 models are used jointly (Ainley *et al.* 2010a; Doak & Morris 2010; Péron *et al.* 2012; Raybaud *et*
 3 *al.* 2013; Jenouvrier *et al.* 2014; Raybaud *et al.* 2015). Thus, it may be expected that the most ro-
 4 bust features, represented similarly in all models, will be preserved, while the model-specific and
 5 possibly artifactual divergent traits will be evened out (Cabré *et al.* 2015). Of course, this «demo-
 6 cratic» approach to model selection is not failsafe: first, because despite the controls included in
 7 the CMIP5 validation process, not all model are of similar «quality» at a given location - the de-
 8 gree of precision and realism varies between models, and so does the size of the spatial grid over
 9 which phenomena are represented. Another source of imprecision is the similarity between differ-
 10 ent models, that may in some cases amount to pseudo-replication when several models of related
 11 architecture give similar outputs because they represent climate in the same way (Cabré *et al.*
 12 2015). In order to assess the robustness of the predicted trends to the choices in model para-
 13 metrisation, an ensemble approach may be used within the framework of a single model, by us-
 14 ing multiple equally possible parameterisations in a *perturbed physics ensemble* approach (Ackerley
 15 *et al.* 2009). But a more common approach is to use altogether different models when the focus is
 16 the prediction itself, and not its robustness.

17 Although ensembles of AO-GCMs usually perform better than single models, their aim is to rep-
 18 resent the earth system as a whole: thus, their still may not accurately capture local aspects of cer-
 19 tain parameters. In the case of less reliably represented areas, it is common to include only mod-
 20 els that already accurately describe the state of climate during the historical period on a few
 21 accurately observed parameters (Ainley *et al.* 2010a; Raybaud *et al.* 2013), *e.g.* satellite-derived
 22 measures of surface temperature or sea-ice cover, which are directly comparable to model outputs.

23 *§-51 The Representative Concentration Pathways.* While the uncertainty arising from climate repre-
 24 sentation within models can be partly addressed through the ensemble approach, another impor-
 25 tant source of uncertainty is the nature of future forcings. Atmospheric greenhouse gas (GHG)
 26 concentration, in particular, is a major driver of effective radiative forcing (Pachauri *et al.* 2014;

1 Stocker *et al.* 2014), but their future concentration is largely determined by political and socio-
 2 economical trajectories. Thus, four different forcing scenarios or *Representative Common Pathways*
 3 (RCP) have been established, each corresponding to a different atmospheric concentration of
 4 GHG in 2100 (Meinshausen *et al.* 2011). Under the RCP-2.6 scenario, efficient political action
 5 leads to a peak in radiative forcing at $\sim 3 \text{ W/m}^2$ before 2100, followed by a decline. Under the
 6 RCP-4.5 and RCP-6 scenarios, radiative forcing stabilises before 2100, either at 4.5 W/m^2 or at 6
 7 W/m^2 . Under the RCP-8.5, no political action is taken and radiative forcing follows the current
 8 trend, reaching 8.5 W/m^2 by 2100. These four scenarios lead to widely different predictions as to
 9 global temperature increase (Meinshausen *et al.* 2011) - yet the direction taken by political deci-
 10 sions is difficult to forecast. Thus a common practise when assessing species response to climate
 11 change is to repeat predictions under several RCPs, in order to assess the range of possible re-
 12 sponses (see e.g. Péron *et al.* 2012; Raybaud *et al.* 2013; Jenouvrier *et al.* 2014; Jenouvrier *et al.*
 13 2014; Raybaud *et al.* 2015). However, variability between models under a single forcing scenario
 14 remains high, so that ensemble predictions may overlap between different RCPs (Goberville *et al.*
 15 2015). Thus, CMIP5 RCP experiments should not be understood as a positive take on future cli-
 16 mate: rather, «*the goal of working with scenarios is not to predict the future, but to better understand*
 17 *uncertainties in order to reach decisions that are robust under a wide range of possible futures*» (Moss
 18 *et al.* 2010).

19 *§-52 Palaeoclimate experiments.* In order to assess the current GCMs' ability to represent different
 20 climatic conditions, it has been proposed that comparatively well-known ancient climate periods
 21 could be used as a benchmark. Thus, the Palaeoclimate Modelling Intercomparison Project
 22 (PMIP, currently in phase 3, see Braconnot *et al.* 2012) has been developed as a counterpart to
 23 the CMIP experiments. PMIP3 experiments focus on past climatic periods with abundant avail-
 24 able data, in particular the LGM («21 kyr experiment»), the Early Holocene («8.2 kyr experi-
 25 ment»), the mid-Holocene period («6 kyr experiment»), and the last millennium («past1000 ex-
 26 periment»). Further back, Interglacial and mid-Pliocene experiments are also included. Of these,

1 three have been included in the CMIP5 panel: the 21 kyr, 6 kyr, and past1000 (Braconnot *et al.*
2 2011). While model parametrisation follows the same rules as piControl, historical and rcp runs,
3 initial conditions, specification of external forcings, and boundary conditions are different, and
4 are determined as part of the experiments definition. These rely either on direct computation (*e.g.*
5 the state of solar radiative forcing at each point of time conditional on Milankovitch cycles, etc.),
6 or on evidence from palæontology and sedimentology (see §9 p. 42 for more details). In the
7 Southern Ocean, one particularly important parameter needs to be addressed, in the configura-
8 tion of the Antarctic Ice Sheet. The continental ice sheets themselves (either in Antarctica or in
9 Greenland) are currently not explicitly modelled in the PMIP3/CMIP5 panels (Kusahara *et al.*
10 2015), and need to be specified as *boundary conditions*, *i.e.* fixed states at the limit of the dynamic
11 cells of the models. Reconstruction of the Antarctic Ice Sheet in the PMIP3 experiments relies on
12 external evidence, in particular the CLIMAP multiproxy reconstructions (CLIMAP 1981). Over-
13 all, PMIP3 experiments support the current generation of AO-GCMs, as reconstructed palæocli-
14 mate reconstructions closely match the available direct evidence (Braconnot *et al.* 2011; Bracon-
15 not *et al.* 2012; Harrison *et al.* 2015).

16 In this work, we use PMIP3/CMIP5 palæoclimate experiments as an explanatory framework for
17 our coalescent-based palæodemographic reconstructions for the King and the Emperor penguin
18 (*The King synnome*, p. 163 and *The Emperor synnome*, p. 211).

1 Chapter 3: Fine-scale colony structure

2 This chapter has been published as: [Cristofari R, Trucchi E, Whittington JD, Vigetta S, Gachot-](#)
3 [Neveu H, Stenseth NC, Le Maho Y, Le Bohec C \(2015\) **Spatial heterogeneity as a genetic mix-**](#)
4 [ing mechanism in highly philopatric colonial seabirds. PloS one 10: e0117981.](#)

5 *Context*

6 *§-53 Local structure and random sampling in population genetics.* Population genetic studies rely on
7 the molecular analysis of individuals sampled from one or several demes or populations (see §2 p.
8 25). With very few exceptions, these individuals represent a small fraction of the focal popula-
9 tions: therefore, when comparing different ensembles through reduced subsets of individuals, an
10 underlying assumption is that the population-level sampling process is unbiased, and that funda-
11 mental parameters (such as allele frequencies, see §45 p. 117) observed in the sampled individu-
12 als are reliable estimators of the same parameters at the population scale. While this hypothesis
13 is normally verified when true random sampling is possible, it is often only assumed in less easily
14 accessible wild species.

15 Thus, an important first step before engaging in large-scale population genetic studies is to estab-
16 lish whether variation in local sampling schemes is likely to influence results on a broader geo-
17 graphic scale. Previous studies of colonial seabirds have suggested that natal philopatry could ex-

Fine-scale colony structure - §54

1 tend not only to the colony, but also to the sub-area within the colony where the individual was
2 born (e.g. in Black guillemots *Cepphus grylle*, see Frederiksen & Petersen 1999). In turn, strong
3 fine scale philopatry may lead to very heterogeneous genetic landscapes, and in particular to
4 strong spatial autocorrelation of genotypes at short distances (e.g. as has been observed in Superb
5 Fairy-wrens *Malurus cyaneus*, see Double *et al.* 2005). A direct consequence of this is that unin-
6 formed local sampling would be likely to oversample autocorrelated genetic groups (*i.e.* loosely
7 related individuals), and therefore to overestimate the amount of genetic drift between demes,
8 while also violating the genetic independence assumption of most models. Indeed, very local pat-
9 terns of genetic variation do not occur on the same level as between-deme genetic structure.
10 Broad-scale genetic structure between demes normally involves long-term reduction of gene flow
11 and random drift in allele frequencies: population-scale genetic sampling, rather than direct kin-
12 ship, plays the main part in the process. Thus, although strong local genetic heterogeneity (e.g.
13 kinship structure) at the local level is likely to promote larger-scale genetic structure (*e.g.* through
14 reduction of dispersal between demes), this is not necessarily the case, as local kinship structures
15 may be regularly dissolved into the broader deme gene pool.

16 *§-54 Impact of local habitat on populations.* While anthropogenic climate change is a global phe-
17 nomenon (see §11 p. 50), its realisation is heterogeneous, and its variability happens at all scales -
18 from continental to local ones (§15 p. 57). Understanding the relationship of species with their
19 environment, whether in a context of stability or of change, therefore requires the integration of
20 very broad-scale patterns (that will be examined later, in *The King synnyme*, p. 163), but also of
21 fine-scale, extremely local parameters. Heterogeneity between seabird colonies and breeding loca-
22 tions is a rather straightforward phenomenon that can be intuitively hypothesised, and that im-
23 plies contrasts between groups of individuals (see *e.g.* Kauffman *et al.* 2003) - but heterogeneity
24 between singular breeding places and individual fitness and breeding success within a colonial
25 framework is a rather new concept in the study of coloniality (see Cornet 2014). Here, we con-

1 sider the link between local habitat heterogeneity, and the shape of genetic variation at the indi-
2 vidual level.

3 *§-55 Aims of this chapter.* In the King penguin, fine scale local philopatry has been anecdotally
4 documented (Nesterova *et al.* 2010), yet the importance of this behaviour in shaping the species
5 genetic landscape is unknown. In this study, we use a dense local sampling of king penguins
6 chicks to assess the importance of very-fine-scale spatial patterns in the structure of the species'
7 colonial units. Our aim is twofold: first, assessing the strength of local genetic heterogeneity at
8 the colony level will allow us to set an upper limit on the strength of philopatric behaviour in the
9 King penguin from a dense local sample, and thus to set a prior expectation for the study of
10 world-wide gene flow. Second, it will allow informed sampling for larger-scale genetic studies in
11 the following chapters of this work (see *The King synnome*, p. 163 or *The Emperor synnome*, p.
12 211).

13 *Abstract*

14 How genetic diversity is maintained in philopatric colonial systems remains unclear, and under-
15 standing the dynamic balance of philopatry and dispersal at all spatial scales is essential to the
16 study of the evolution of coloniality. In the King penguin, *Aptenodytes patagonicus*, return rates
17 of post-fledging chicks to their natal sub-colony are remarkably high. Empirical studies have
18 shown that adults return year after year to their previous breeding territories within a radius of a
19 few meters. Yet little reliable data are available on intra- and inter-colonial dispersal in this
20 species. Here, we present the first fine-scale study of the genetic structure in a king penguin
21 colony in the Crozet Archipelago. Samples were collected from individual chicks and analysed at
22 8 microsatellite loci. Precise geolocation data of hatching sites and selective pressures associated
23 with habitat features were recorded for all sampling locations. We found that despite strong natal
24 and breeding site fidelity, king penguins retain a high degree of panmixia and genetic diversity.

1 Yet, genetic structure appears markedly heterogeneous across the colony, with higher-than-ex-
2 pected inbreeding levels, and local inbreeding and relatedness hotspots that overlap predicted
3 higher-quality nesting locations. This points towards heterogeneous population structure at the
4 sub-colony level, in which fine-scale environmental features drive local philopatric behaviour,
5 while lower-quality patches may act as genetic mixing mechanisms at the colony level. These
6 findings show how a lack of global genetic structuring can emerge from small-scale heterogeneity
7 in ecological parameters, as opposed to the classical model of homogeneous dispersal. Our results
8 also emphasise the importance of sampling design for estimation of population parameters in
9 colonial seabirds, as at high spatial resolution, basic genetic features are shown to be location-de-
10 pendent. Finally, this study stresses the importance of understanding intra-colonial dispersal and
11 genetic mixing mechanisms in order to better estimate species-wide gene flows and population
12 dynamics.

13 *Introduction*

14 *§-56 Coloniality and population structure.* How colonial systems are maintained in non-coopera-
15 tive species remains an important question in evolutionary biology (Danchin & Wagner 1997;
16 Bowler & Benton 2005). Philopatric behaviour is usually considered to be the basis of coloniality
17 (Bowler & Benton 2005). It is thought to offer several selective advantages, such as a good
18 knowledge of the higher quality breeding spots and of the pool of breeding partners linked to
19 these spots (Wheelwright & Mauck 1998; Heg *et al.* 2011; Arnaud *et al.* 2012), and it favours
20 the selective value of proximal defensive behaviour (Dunford 1977) or allofeeding (Lecomte *et al.*
21 2006) through kinship selection. However, the drawback of such behaviour is the fragmentation
22 of genetic diversity at the colony level (Banks & Peakall 2012), which leads to an increase in in-
23 breeding within clusters of closely-related individuals, in turn potentially leading to local inbreed-
24 ing depression (Frankham 1995; Frankham 2005). Dispersal is therefore an essential balancing
25 force in the conservation of colonial systems (Serrano *et al.* 2004; Bowler & Benton 2005; Bick-

1 nell *et al.* 2012; Fernández-Chacón *et al.* 2013). It is known to promote gene flow (Winters &
 2 Waser 2003) and genetic adaptability (Bowler & Benton 2005) as well as plasticity in response to
 3 habitat changes (Travis & Dytham 1999; Serrano *et al.* 2004).

4 Both philopatry and dispersal occur at all spatial scales in many colonial species (Murrell *et al.*
 5 2002; Barlow *et al.* 2013), yet it has often been assumed that strong philopatric behaviour is a
 6 common trait of most pelagic seabirds (Frederiksen & Petersen 1999; Bried *et al.* 2007; Dugger
 7 *et al.* 2010). Indeed, the benefits of philopatry and coloniality (Coulson 2002), the barriers to
 8 dispersal (Friesen *et al.* 2007) and the dynamics of colony formation and extinction
 9 (Matthiopoulos *et al.* 2005) have traditionally been explored under that hypothesis. Thus several
 10 studies have documented a lack of genetic structure amongst distant colonies, which, in this con-
 11 text, seemed incompatible with assumed philopatric behaviour (the so-called "seabird paradox"
 12 of Milot *et al.* 2008).

13 It is often considered that the relatively weak evidence for structure, together with the high level
 14 of genetic diversity observed in many seabird species (e.g. Akst *et al.* 2002; Bouzat *et al.* 2009;
 15 Gómez-Díaz *et al.* 2009), are mostly the consequence of the random dispersal of a few juveniles.
 16 On the other hand, the importance of intra-colonial dispersal in fitness enhancement has been
 17 studied through phenological methods (Steiner 2005), and dispersal distance, even at very fine
 18 scale, has been shown to depend on species-specific ecological dynamics (Murrell *et al.* 2002;
 19 Banks & Peakall 2012). However, strict correlation between dispersal, site fidelity, and popula-
 20 tion structure has been questioned (Pearce *et al.* 2008), and the importance of intra-colonial
 21 movements on genetic mixing has received comparatively little attention.

22 *§-57 The King penguin as a model.* The King penguin, *Aptenodytes patagonicus*, which breeds in
 23 large colonies all around the sub-Antarctic area, is an ideal model to investigate these aspects of
 24 coloniality. While the King penguin's sister-species, the Emperor penguin, *Aptenodytes forsteri*,
 25 does not display territorial behaviour (Williams 1995), the King penguin is known to choose and
 26 defend a breeding territory within the colony, even though it does not build a nest (Bried & Jou-

1 ventin 2001). Breeders of this species have been described as having a marked philopatric behav-
 2 iour across reproductive seasons and have been observed to use the same breeding site, within a
 3 radius of a few meters, year after year within their colony (Barrat 1976; Bried & Jouventin
 4 2001). Return rates of juveniles to their natal colony are also typically high (Le Bohec *et al.* 2008;
 5 Saraux *et al.* 2011b). However, it is not yet clear whether natal philopatric behaviour also extends
 6 to the specific birth site within the colony, and whether this behaviour may result in local struc-
 7 turing of populations. In order to test this, a genetic study was conducted in the "Baie du Marin"
 8 colony on Possession Island, Crozet Archipelago. If philopatric behaviour indeed extends to
 9 breeding territory selection, it is expected that the colony will have a non-random genetic struc-
 10 ture, with closely related individuals distributed in spatial clusters, as has already been observed in
 11 several non-avian philopatric species (e.g. Rivers *et al.* 2005; Kanno *et al.* 2011). However, even
 12 in known philopatric species, detection of genetic structure can be hindered by a variety of be-
 13 haviours, such as juvenile and adult dispersion or extra-pair paternity (Double *et al.* 2005); there-
 14 fore, the methods used to study it must be sensitive enough to detect even a weak signal (Jones &
 15 Wang 2012).

16 *§-58 Expectations and hypotheses.* When considering the genetic structure of a species, expected
 17 patterns may be analysed according to (i) global homogeneous processes or (ii) local heteroge-
 18 neous processes whose effect varies across the area (Osborne *et al.* 2007). (i) The predictions asso-
 19 ciated with strong and consistent philopatric behaviour belong to the former class. Without
 20 taking spatial distribution of individuals into account, a skew in global population inbreeding de-
 21 scriptors may be expected (Wright 1965), with an excess of homozygosity reflecting higher-than-
 22 random relatedness between paired individuals. However, a prior to F-statistics is that popula-
 23 tions are in a state of equilibrium, which may be in contradiction with fine spatial and temporal
 24 scale analysis, such in our study (Whitlock & McCauley 1999; Paetkau *et al.* 2003). Alternatively
 25 spatial autocorrelation should be detected at the colony level, reflecting a gradual isolation-by-
 26 distance trend, though this should only hold true if the process leading to spatial structure is ho-

1 mogenous at the individual level (Osborne *et al.* 2007). (ii) If, on the other hand, philopatric be-
2 haviour is a more variable trait dependent on local ecological conditions, the processes involved
3 in colony structure are expected to belong to the second class. Local heterogeneous signatures
4 should therefore be found, such as divergent population parameters for different areas of the
5 colony (Jones & Wang 2012). Although inbreeding is not in itself an indicator of spatial struc-
6 ture, higher-than-expected individual inbreeding levels may reveal a non-random mating system,
7 favouring relatives as potential partners. In such cases, a spatially-correlated distribution of in-
8 breeding suggests a space-driven mating system (Ruzzante *et al.* 2002; Vogl *et al.* 2002). Pairwise
9 relatedness, calculated across spatially-defined windows, is another insight into the same phe-
10 nomenon (Ruzzante *et al.* 2002; Stevens *et al.* 2011).

11 In this study, we use a combined approach including genetic, spatial and local ecological data to
12 investigate the mechanisms of genetic mixing inside a seabird colony. We focus in particular on
13 how fine-scale habitat heterogeneity patterns may influence local dispersal behaviour, and how
14 lack of genetic structuring can emerge from small-scale heterogeneity in ecological parameters.

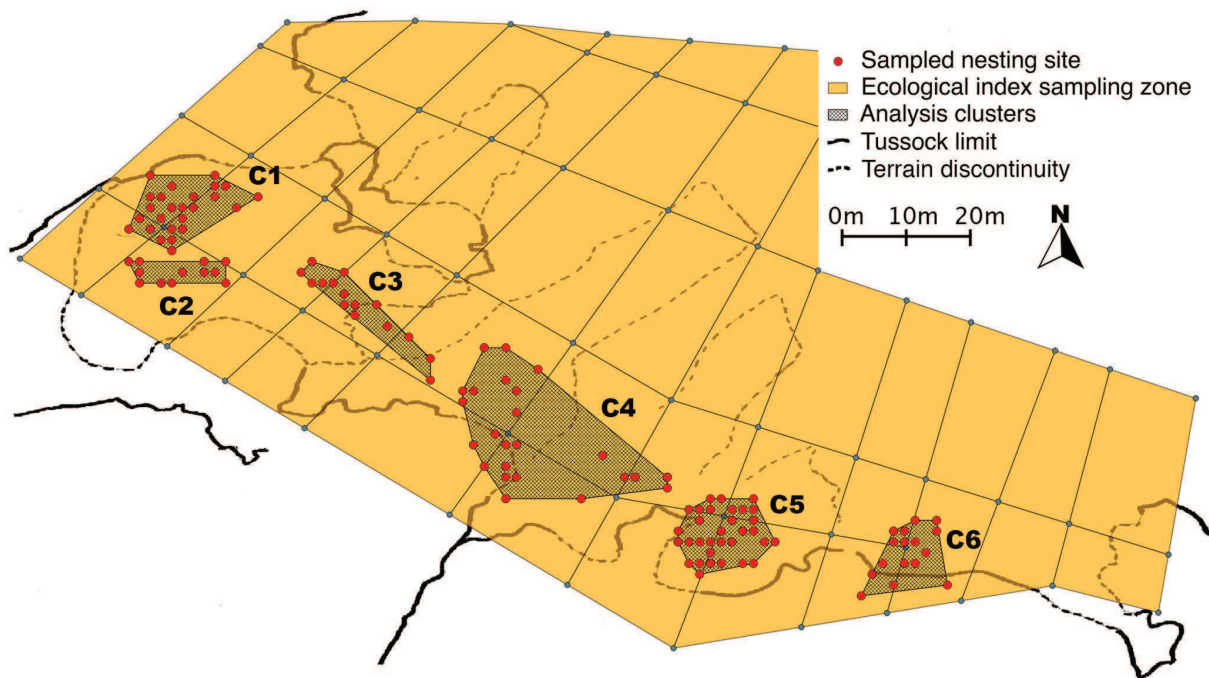
15 *Materials and Methods*

16 *§-59 Permits and ethics statement.* All animals in this study were handled only once in order to
17 mark them with a plastic tag and to conduct morphological measurements. All procedures em-
18 ployed during this field work were approved by the Ethical Committee of the French Polar Insti-
19 tute (Institut Paul Emile Victor – IPEV) and conducted in accordance with its guidelines, also
20 complying with French laws including those relating to conservation and welfare. Authorizations
21 to enter the breeding site (permits n°2009-57 issued on august 26th 2009) and handle birds (per-
22 mits n°2009-59 issued on august 29th 2009) were delivered first by the French “Ministère de
23 l’Aménagement du Territoire et de l’Environnement” and then by the Terres Australes et Antarc-
24 tiques Françaises (TAAF).

Fine-scale colony structure - §60

1 Handled animals were removed from the colony in order to minimize the disturbance to neigh-
2 bouring birds and taken to a shelter a few meters away for manipulation. They were hooded to
3 reduce their stress and manipulations lasted between 5 and 10 minutes. All blood-sampling and
4 tagging material was sterilized (either sealed, or through chemical sterilization). Moreover, Vété-
5 dine soap and alcoholic antiseptic solutions were used to disinfect the skin before bleeding and
6 tagging. Flesh wounds did not seem infected thereafter (personal observations on a subset of re-
7 captured birds).

8 *§-60 Study site and sampling.* Our study was conducted during the summer season 2010/2011 in
9 the king penguin colony of "La Baie du Marin", on Possession Island, Crozet Archipelago
10 (46°25'S, 51°45'E), which comprises around 16,000 breeding pairs (Delord *et al.* 2004). Chicks
11 born in a sub-section of the colony comprising approximately 10,000 breeding pairs (Antavia,
12 Gendner *et al.* 2005) were randomly selected and captured along a peripheral 120m-transect, in
13 order to maximise separation distance. To minimize disturbance of breeding penguins, sampling
14 was conducted no deeper than 4-5 nests towards the center the colony. Sampling was conducted
15 at a high spatial resolution (~2 m between successive samples, i.e. every second nest at the time of
16 sampling) in order to assess fine-scale genetic structure of the colony (sampling map is given in
17 Fig. 13). The sampling area was partitioned according to natural terrain discontinuities (such as
18 tussocks, rocks, or flooded patches), in a total of 6 clusters encompassing all sampled nests (Fig.
19 13). 175 early-hatched chicks were thus sampled during early brooding (~ 2 week old chicks) and
20 temporarily tagged with a small external plastic pin (Floytag®). Each chick's hatching site was ge-
21 olocated with a Global Positioning System device (Garmin eTrex®), and further confirmed using
22 visual grid markings in the field (*ca.* 10x10m cells). King penguins are monogamous during a re-
23 productive season and lay a single egg each year (Stonehouse 1960; Barrat 1976), thus we exclude
24 the possibility of full siblings in the study population. Blood (~ 100 µL) samples were collected
25 from the brachial veins using heparin-coated tubes, transferred to a Whatman filter paper, dried,
26 and later frozen at -20°C for the genetic analysis.



1

2 **Figure 13 | Sampling distribution.** Sampling was restricted to the periphery of the colony. Orange zone
 3 boundaries are marked on the ground for remote parameter assessment. Shaded clusters run from C1 (north-west) to
 4 C6 (south-east). Orange zone: extent of the colony at the peak of the breeding season.

5 *§-61 DNA extraction, PCR and genotyping.* DNA was extracted using a DNeasy® Blood and Tis-
 6 sue Kit (Qiagen) with a modified protocol. A prolonged initial incubation of the impregnated
 7 Whatman paper in PBS buffer (24 h) was used to dissolve dried blood cells, and a final elution in
 8 50 µL instead of 100 µL was made to obtain higher DNA concentration. The final DNA concen-
 9 tration was assessed by spectrometry and adjusted to 25 ng·µL⁻¹.

10 Genotyping was performed with 10 microsatellite loci (Akst *et al.* 2002; Billing *et al.* 2007;
 11 Ahmed *et al.* 2009; Hart *et al.* 2011). PCR was conducted using MasterMix® (Qiagen) premixed
 12 TAQ-polymerase, dNTPs and MgCl₂, in a total reaction volume of 12.5 µL (6.75 µL Master-
 13 Mix, 2 µL of each primer pair, 0.5 µL DNA, H₂O up to a final volume of 12.5 µL). Amplifica-
 14 tion followed a touchdown protocol, with a 5-minute denaturation at 95°C, 12 amplification cy-

Fine-scale colony structure - §61

cles of a 30 second denaturation at 95°C, 30 seconds of annealing beginning at 63°C in the first cycle and decreasing one degree at each cycle until 52°C, and 30 seconds of elongation at 72°C, followed by 23 cycles of 30 seconds at 95°C, 30 seconds at 52°C, and 30 seconds at 72°C. Amplification was finished with 10 minutes at 72°C. The same protocol was used for all primers. FAM and HEX fluorochrome-labelled primers were used for sequencing: several microsatellite loci could be multiplexed in the same reaction (Table 1). When multiplex amplification failed, samples were amplified at each locus separately, and pooled together for electrophoresis. Sequencing was performed on an ABI3730 capillary sequencer. 1 µL of PCR product was diluted in 9 µL H₂O, and 1 µL of this solution was added to 10 µL formamide and 0.2 µL GeneScan 500-ROX size standard (Applied Biosystems). Output files were analysed on GeneMapper (Applied Biosystems). A random 10% of the samples were amplified and genotyped a second time to check the repeatability and accuracy of the readings.

Table 1 | Microsatellite analysis: multiplexing parameters and summary statistics.

	Ech030		Ech036		Ech051		Ech071		AM13		B3-2		Emm4		RM3	
PCR	Multi_A	Multi_A	Multi_A	Multi_A	Multi_B	Multi_B	Multi_B	Multi_B	Single	Single	Multi_C	Multi_C	Single	Single	Multi_C	Multi_C
Genotyping	Seq_A	Seq_A	Seq_A	Seq_A	Seq_B	Seq_B	Seq_B	Seq_B	Seq_C	Seq_C	Seq_C	Seq_C	Seq_A	Seq_A	Seq_C	Seq_C
Label	HEX	HEX	HEX	HEX	FAM	FAM	FAM	FAM	HEX	HEX	HEX	HEX	FAM	FAM	FAM	FAM
Pattern	(CTAT) _n	(GT) _n	(GT) _n	(GT) _n	(GT) _n	(CTCAT) _n	(CTCAT) _n	(CTCAT) _n	(GT) _n	(GT) _n	(GT) _n	(GT) _n	(CT) _n	(CT) _n	(CA) _n	(CA) _n
Allele count	9	8	8	8	8	13	13	13	6	9	9	9	9	9	4	4
Gene diversity	0.773	0.750	0.750	0.750	0.544	0.866	0.866	0.866	0.702	0.642	0.642	0.642	0.658	0.658	0.541	0.541
Allelic richness	8.988	7.994	7.994	7.994	7.988	12.994	12.994	12.994	6.000	8.994	8.994	8.994	9.000	9.000	4.000	4.000
	Allele	Freq.	Allele	Freq.	Allele	Freq.	Allele	Freq.	Allele	Freq.	Allele	Freq.	Allele	Freq.	Allele	Freq.
	267	0.024	190	0.024	188	0.012	215	0.009	109	0.006	285	0.009	233	0.003	202	0.012
	271	0.045	194	0.179	192	0.006	220	0.027	111	0.066	288	0.033	235	0.012	207	0.518
	275	0.232	196	0.003	194	0.580	225	0.051	115	0.326	290	0.009	237	0.407	208	0.438
	279	0.330	198	0.015	196	0.345	230	0.057	117	0.371	292	0.238	239	0.410	210	0.033
	283	0.223	200	0.244	198	0.039	235	0.152	119	0.228	294	0.536	241	0.087	-	-
	287	0.113	202	0.378	200	0.012	240	0.185	121	0.003	296	0.113	243	0.033	-	-
	291	0.027	204	0.125	204	0.003	245	0.176	-	-	298	0.039	245	0.024	-	-
	295	0.003	206	0.033	206	0.003	250	0.176	-	-	300	0.021	247	0.015	-	-
	299	0.003	-	-	-	-	255	0.092	-	-	302	0.003	249	0.009	-	-
	-	-	-	-	-	-	260	0.048	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	265	0.015	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	270	0.012	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	280	0.003	-	-	-	-	-	-	-	-

Fine-scale colony structure - §63

1 §-62 *Analysis of microsatellite data.* Hardy-Weinberg equilibrium was tested in *adegenet* (Jombart
2 2008) with 100,000 permutations, and in Fstat (Goudet 1995). F-statistics were assessed in Fstat
3 together with microsatellite summary statistics (allele range, count and frequencies, and linkage
4 disequilibrium). Global intrinsic differentiation in the sample, regardless of any spatial grouping,
5 was tested in Structure (Pritchard *et al.* 2000), with 200,000 burn-in and 1,000,000 steps, under
6 Admixture model, with K values ranging from 1 to 6 (allele frequencies correlated). Log-proba-
7 bility of the data was calculated in Structure Harvester (Earl 2012).

8 §-63 *Spatial autocorrelation analyses.* In a population with limited, yet homogeneous, internal dis-
9 persal and gene flow, genetic drift is expected to generate a pattern of isolation by distance. Relat-
10 ed genotypes will tend to become clustered in spatial patches, and correlation between genotypes
11 will decay with distance (Epperson 2005). In this case, correlation between genetic and spatial
12 distance between samples should be observed. In order to test that prediction, spatial autocorrela-
13 tion analysis was performed across the whole sampling area using Genalex (Peakall & Smouse
14 2006), through multivariate analysis of pairwise euclidean and square genetic distances, regardless
15 of any spatial grouping. Significance of a autocorrelation was assessed through random permuta-
16 tion of samples across the whole area, and subsequent bootstrapping of the samples contained
17 within pre-defined distance-classes around each individual: permutation tests are used to assess
18 the randomness of the distribution, while bootstrapping allows for the evaluation of the weight
19 of potential individual outliers. We used distance classes starting at 2 meters (the average mini-
20 mum distance between two sampled nests), and no longer than 8 meters (the average radius of
21 our clusters). Longer distance classes were excluded as they would give too much weight to cross-
22 cluster comparisons. 100,000 random permutations were performed.

23 However, heterogeneous spatial processes may not always be detected by global spatial autocorre-
24 lation analyses (Miller 2012). Heterogeneous fine-scale genetic structure was therefore assessed in
25 Genalex using the 2D-Local Spatial Analysis algorithm (2D-LSA, Double *et al.* 2005): each indi-
26 vidual was tested for genetic relatedness to its n nearest neighbours in order to identify fine-scale

Fine-scale colony structure - §64

1 patches of lower genetic diversity. This method differs from standard autocorrelation tests in that
2 it completely removes further distance classes from the analysis, and gives the same weight to all
3 individuals considered as neighbours. Thus, it takes into account the possibility of locally restrict-
4 ed areas, within which all individuals are equivalently related and therefore do not exhibit spatial
5 autocorrelation. Once again, significance levels were estimated through random permutation of
6 the samples. Analyses were performed with a number of neighbours ranging from 8 to 10 (the
7 approximate closest 5% of the colony), and 100,000 random permutations.

8 *§-64 Inbreeding and relatedness.* Individual inbreeding coefficients and pairwise relatedness were
9 also taken as an index of local genetic structure: if potential mates are more related to each other
10 in some areas, then offspring sampled in that area are expected to be more inbred, and potentially
11 related to each other. Inbreeding was studied both at population and individual levels. At popula-
12 tion level, Ballou's maximum likelihood method was preferred (Ballou 1997), since it has been
13 shown to better estimate population parameters provided sample is large enough (Weir 1990).
14 However, MLE methods are known to be sensible to small sample size, and they are therefore less
15 suitable for estimating individual parameters. Individual inbreeding and relatedness were thus
16 calculated using Ritland's method-of-moments estimators (Ritland 1996). Ballou's MLE index
17 was calculated in *adegenet* (Jombart 2008). Individual means were extracted from a subset of
18 2,000 values sampled from 20,000 estimates, and were then averaged across the whole sample.
19 Ritland's MME estimators were computed in *Coancestry* (Wang 2011). As expectations for indi-
20 vidual inbreeding levels vary, and cannot be reliably imported from another study organism, ex-
21 pected inbreeding distribution was estimated through simulation. A virtual population of entirely
22 non-related individuals (as is the expectation for a large, ideal and randomly sampled population)
23 was generated with the same allele frequencies and per-locus homozygosity levels as the sampled
24 colony, and empirical data was compared to the simulated individual inbreeding distribution.
25 Difference in average individual inbreeding between the observed and the simulated distribution
26 was assessed through a bootstrapping method as implemented in *Coancestry*: 10 independent

1 simulated datasets were generated. For each of them, both observed and simulated individuals
 2 were pooled, and 100,000 randomly redistributed partitions were generated. Group-averaged in-
 3 breeding difference was computed for each of these partitions, and significance of the observed
 4 result was checked against this random distribution. Observed individual inbreeding was then
 5 mapped onto the colony to identify potential high-inbreeding clusters using *Quantum GIS*
 6 (Quantum GIS Development Team, 2013).

7 *§-65 Pairwise cluster analysis.* If our second hypothesis (heterogeneous structuring processes)
 8 holds true, we expect these processes to apply differently in specific areas of the colony depending
 9 on local characteristics. We therefore performed the following analyses on the discrete natural
 10 clusters defined in the sampling area (Fig. 13), as opposed to the continuous models applied
 11 above. Genetic differentiation between clusters was tested in *Genepop* (exact G-test, default
 12 Markov chain parameter values) and *Fstat* using chi-square assessment of genotypic frequencies.
 13 Pairwise divergence in inbreeding and relatedness distribution between clusters was also assessed:
 14 pairs of individuals belonging to one cluster were compared to pairs of individuals belonging to
 15 different clusters, and significance of distribution divergence was tested in *Coancestry* (10,000
 16 bootstrapping samples).

17 *§-66 Site-quality descriptors analysis.* To measure breeding site quality, several ecological variables
 18 were used as proxies. (i) For 3 consecutive breeding seasons (2010-2012), chick survival rate be-
 19 tween hatching (January/February t) and fledging (November/December t) was measured as the
 20 proportion of marked chicks still alive at fledging (Weimerskirch *et al.* 1992). (ii) Parasite infesta-
 21 tion has been shown to influence breeding success and adult survival rate in king penguins (Man-
 22 gin *et al.* 2003). Thus, for 8 consecutive years (2005-2012), tick infestation was measured as the
 23 proportion of infested adults in a randomly-selected sample (N=50) within each colony grid
 24 zone. (iii) Finally, as early breeders lay eggs in November/December, and late breeders after Janu-
 25 ary (Brodin *et al.* 1998; Descamps *et al.* 2002), an estimate of the ratio of incubating and brood-
 26 ing birds in early March was used, as preferred nesting sites are expected to be occupied first by

Fine-scale colony structure - §66

1 early breeders, while late breeders will occupy remaining, lower-quality areas. For 8 consecutive
2 years (2006-2013), site occupancy timing was thus indexed as the proportion of brooding birds
3 among 50 randomly-chosen breeders (i.e. incubating or brooding birds, leaving non-breeders
4 aside). This results in a score of 1 for preferred early-breeding sites, and a score of 0 for locations
5 only occupied by late breeders.

6 Correlation between successive year per unit of space was assessed for all three variables (Pearson's
7 r test performed between successive years with all grid zones), and spatial autocorrelation of the
8 data was assessed in *spdep* (Bivand 2013, Moran's I, two-sided test, using the closest 5%, or 8
9 nearest neighbours). Our underlying hypothesis being that the breeding-site location choice of an
10 individual breeder depends on its long-term experience and knowledge of breeding-site quality,
11 ecological data were then concatenated for all available sampling years on each space unit. Adult
12 tick infestation on each grid-location was averaged across all 8 sampling years. Chick survival rate
13 was spatially averaged through local spatial analysis: each sampling location was weighted with
14 the mean survival of the closest 5% of the colony (8 nearest sampling sites) within each breeding
15 year, and data for all years were simultaneously mapped onto the colony. Areas with generally
16 high score therefore have a consistently high local survival rate across several breeding seasons.
17 Spatial autocorrelation within each dataset was assessed through Moran's I test implemented in
18 *spdep* R package (Bivand 2013), with the closest 5% of the colony defining the spatial weights.
19 For representation purposes, each nesting location was attributed its local tick load, chick survival
20 score and occupancy score, and values were interpolated across the whole colony in Quantum
21 GIS, using squared inverse distance algorithm.

22 In order to assess the relation and overlap of the three included ecological variables, we per-
23 formed a principal component analysis (as implemented in R package FactoMineR, Lê *et al.*
24 2008). Since we focus on three variables that seem to determine habitat quality on the periphery
25 of the colony, but omit some factors that may interfere in other zones than our study area (e.g.
26 vulnerability to flooding, or density pressure), we restricted the PCA to the peripheral area where

1 our genetic sampling is restricted. Each variable was averaged by grid zone, across all available
 2 years. The first principal component was plotted on the colony using Quantum GIS.

3 *Results*

4 *§-67 Microsatellite data.* Of the 10 microsatellite loci tested for genotyping, only 8 were success-
 5 fully amplified, and were used in the subsequent analyses. Ech011 amplified irregularly in king
 6 penguins under the tested PCR conditions, and Ech081 was repeatedly scored as more than 2 al-
 7 leles per genotype. Only Ech030, Ech036, Ech051, Ech071, AM13, B3-2, Emm4 and RM3
 8 were therefore retained for genotyping. For these, scoring repeatability was of 100% across our
 9 random replicates.

10 The total population was found to be in Hardy-Weinberg equilibrium. After 100,000 randomisa-
 11 tions, no locus showed significant departure from Hardy-Weinberg equilibrium (*P-values*:
 12 Ech030 = 1; Ech036 = 0.9997; Ech051 = 0.3880; Ech071 = 1; AM13 = 1; B3-2 = 0.4948;
 13 Emm4 = 1; RM3 = 0.0697). F_{IT} was not found to be significantly different from zero for the
 14 whole sample ($F_{IT} = 0.0259$), although it was significantly higher for one locus (Ech071: $F_{IT} =$
 15 0.1275). Considering this might be a sign of weak linkage to a region under selective pressure,
 16 subsequent analyses were therefore tested both including and excluding this locus. No locus was
 17 found to be under linkage disequilibrium. Allele frequencies, counts, and genetic diversity are
 18 given in Table 1.

19 Population inbreeding indicator mean value across our sample was $F = 0.2321 \pm 0.0958$ (Ballou's
 20 MLE estimator, standard error: 0.0074). At the individual level, MME inbreeding coefficients
 21 were found to deviate from a random distribution. In 80% of our simulations, the empirical data
 22 fell outside the 95% interval of the random distribution, and in all cases the difference was clear-
 23 ly positive (empirical mean inbreeding clearly higher than simulated). In particular, the lower
 24 (outbred) range was strongly under-represented as compared to a random, non-related sample as

Fine-scale colony structure - §67

1 simulated in *Coancestry* (Fig. 14b-c). Ritland's inbreeding coefficient was then mapped on the
2 colony. Two more inbred regions appeared, at the south-east and in the centre of the sampling
3 area (Fig. 15a), corresponding to clusters C3 and C5-C6.

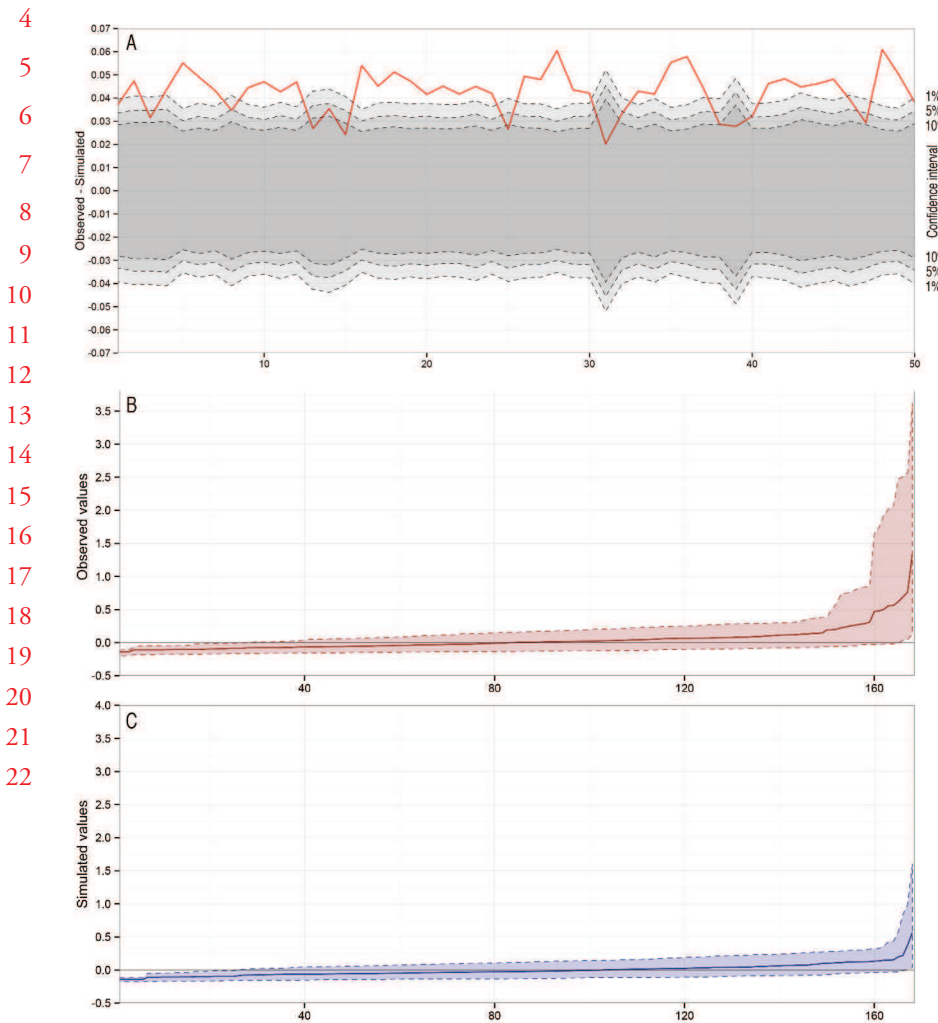
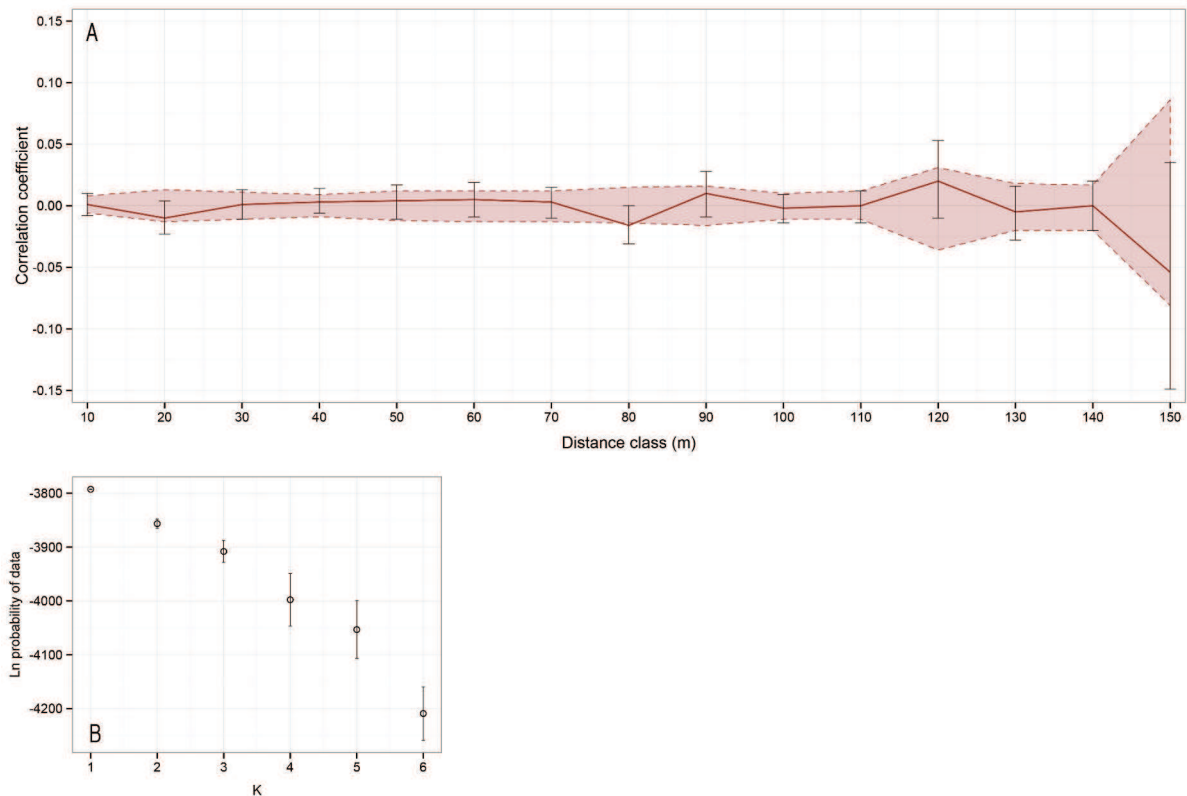


Figure 14 | Observed individual inbreeding distribution deviates from expected distribution in a non-related sample. (A) Difference between simulated and observed mean population inbreeding level (Ritland's coefficient) for 50 simulated datasets. Solid red line: observed difference. Gray intervals: 1%, 5% and 10% confidence intervals. (B) Observed distribution in the sample. (C) Expected distribution in a non-related population, given the same parameters as the observed population.

Fine-scale colony structure - §68



1 **Figure 15 | Individual inbreeding and nearest-neighbours-relatedness tend to have a clustered**
2 **distribution throughout the colony.** (A) Distribution of Ritland's individual inbreeding coefficient along the
3 sampling area. Shaded zones represent analysis clusters C1 to C6. Gray circles signal point displacement for represen-
4 tation purposes (B) 2D-LSA scores (9 neighbours analysis). Red triangles represent individuals that are significantly
5 more related to their 9 nearest neighbours than to random individuals.

6 *§-68 Spatial analysis.* Spatial autocorrelation tests performed on the whole population in *Genalex*
7 did not yield any significant structure across the sampling area (Fig. 16a), suggesting no visible
8 decay in correlation in the first distance classes, as would be expected if genetic diversity varied
9 along a gradient across the colony. Intrinsic clustering of the dataset, assessed in Structure, did
10 not yield any significant results either (Fig. 16b). In our 2D-LSA analyses however, a maximum
11 of 10 individuals across our sample were found to have significant genetic correlation with their 9
12 nearest neighbours ($P < 0.05$). It is noteworthy that these individuals were clustered in two
13 groups that strongly overlapped with the higher-inbreeding regions identified previously (clusters

1 C3 and C5, see Fig. 15b). Genetic correlation decayed quickly when we increased the number of
 2 neighbours used in the analysis above the nearest 5% of the colony.

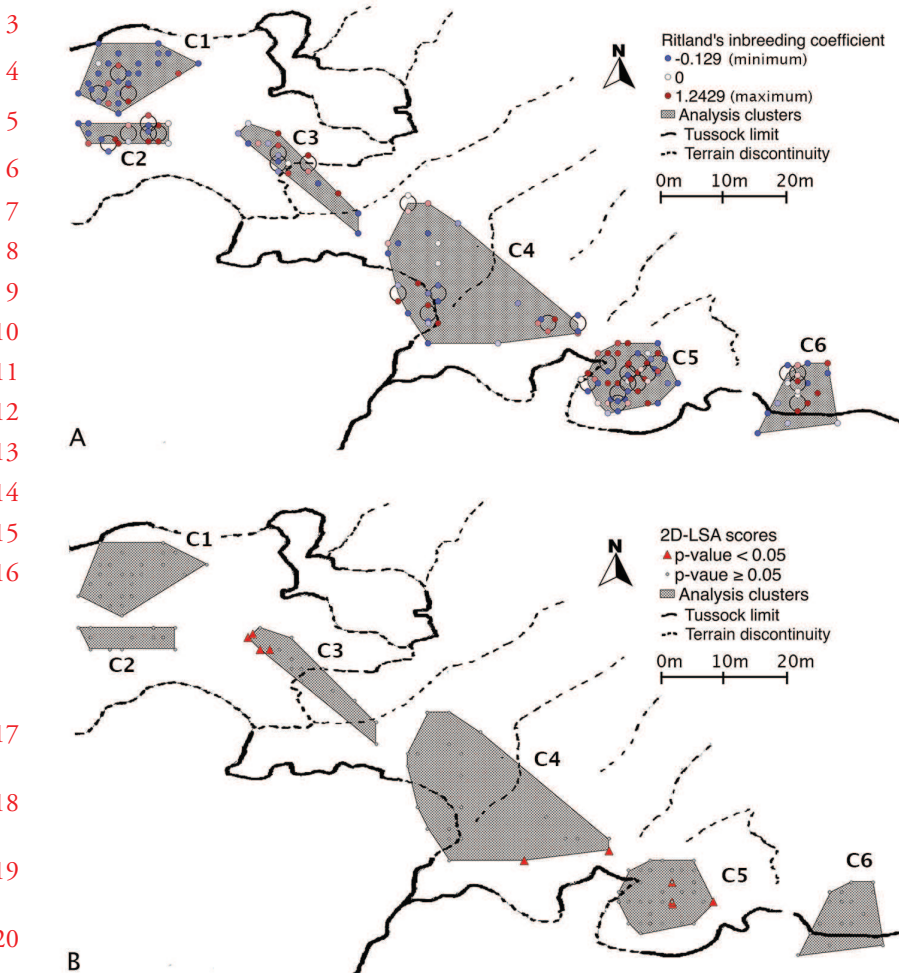


Figure 16 | Spatial correlation and Intrinsic clustering tests do not support strong structure. (A) Spatial correlogram. Solid red line: observed correlation, by distance classes. Dashed red lines: random expectation confidence interval, determined by bootstrapping. (B) Ln probability of the genetic data under six different clustering situations (from 1 to 6 populations).

17 Cluster differentiation analyses performed in
 18 *Genepop* showed that cluster C3 was significantly
 19 different from three other
 20 clusters (genotypic differentiation, exact G -test: $P_{C1} = 0.0277$, $P_{C4} = 0.0326$, and $P_{C5} = 0.0268$).
 21 However, pairwise F_{ST} comparisons did not yield any significant result.

24 Heterogeneity in inbreeding distribution across different clusters was tested pairwise in *Coancestry*
 25 (bootstrapping: 10,000): cluster C1 was found to have significantly ($P_{C1} < 0.05$) less inbreeding
 26 than clusters C2, C3, C5 and C6 (but not C4). Most significant was the comparison between
 27 C1 and C5 (Table 2). In the same way, pairwise individual relatedness was found to be signifi-

Fine-scale colony structure - §68

1 cantly higher in C6 than in C1 and C4 (other pairwise comparisons were non-significant, see
2 Table 2).

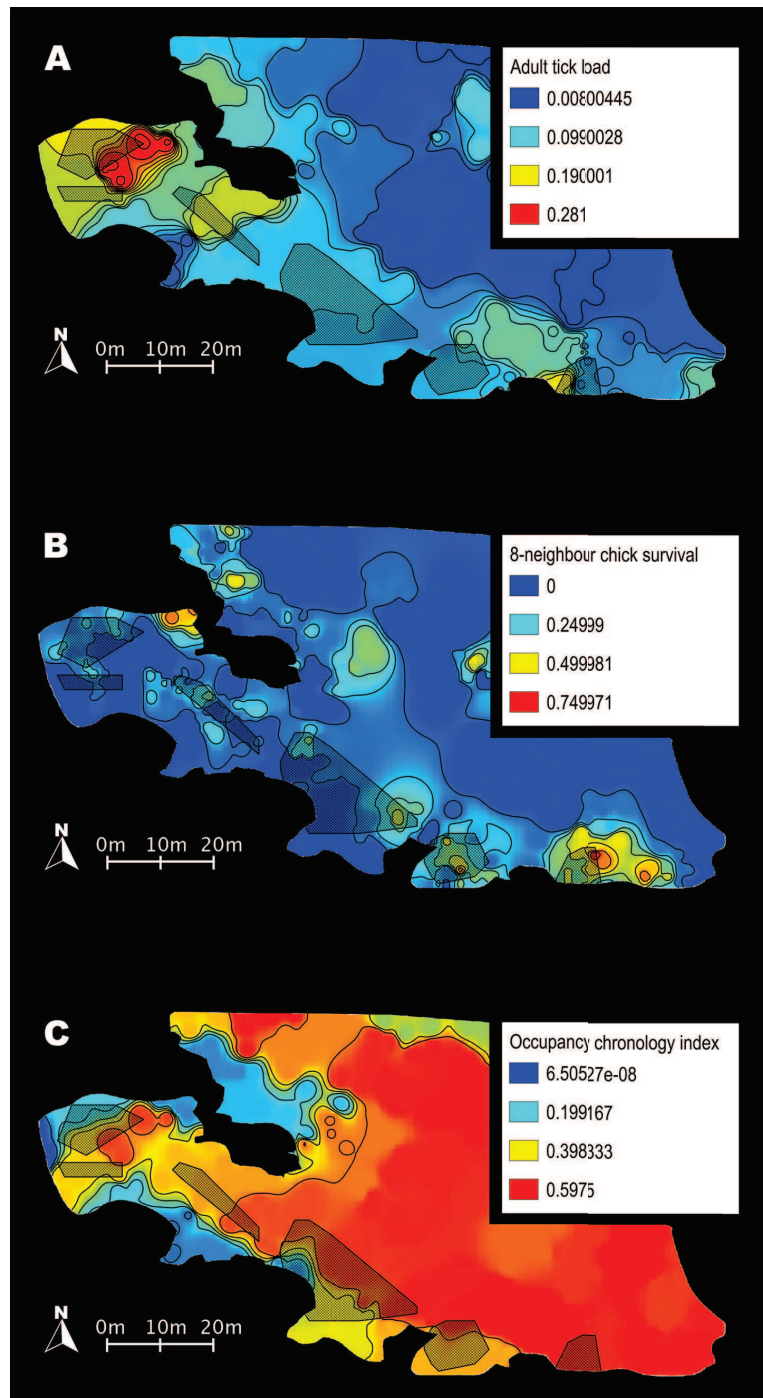
3 **Table 2 | Pairwise cluster comparison of inbreeding and relatedness distributions reveal genetic**
4 **heterogeneity in the sampling area.** *Upper triangle:* Ritland's individual inbreeding coefficient. *Lower triangle:*
5 Ritland's pairwise relatedness coefficient. The empirical coefficient is shown in bold, followed by the 95% confidence
6 interval. Significantly divergent comparisons are shown in bold type with asterisks.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1		-1.1002e-1	-1.5355e-1	-3.9188e-2	-7.0628e-2	-6.3993e-2
		8.4695e-2 C1<C2*	1.0299e-1 C1<C3*	4.0492e-2 C1=C4	5.2465e-2 C1<C5*	5.8778e-2 C1<C6*
		-8.5559e-2	-1.2204e-1	-4.3607e-2	-5.1926e-2	-5.8408e-2
Cluster 2	1.0118e-2		-4.3531e-2	7.0834e-2	3.9394e-2	4.6029e-2
	2.0834e-2 C1=C2		1.5727e-1 C2=C3	7.9085e-2 C2=C4	8.1326e-2 C2=C5	1.0041e-1 C2=C6
	-2.3307e-2		-1.6080e-1	-7.1760e-2	-7.3237e-2	-9.9416e-2
Cluster 3	-4.0898e-3	-1.4208e-2		1.1436e-1	8.2925e-2	8.9560e-2
	2.1038e-2 C1=C3	3.2650e-2 C2=C3		1.1821e-1 C3=C4	1.1181e-1 C3=C5	1.4201e-1 C3=C6
	-2.3405e-2	-3.4437e-2		-8.6573e-2	-8.5461e-2	-1.2877e-1
Cluster 4	4.5949e-3	-5.5231e-3	8.6848e-3		-3.1439e-2	-2.4804e-2
	1.1684e-2 C1=C4	1.9016e-2 C2=C4	-1.7182e-2 C3=C4		4.7974e-2 C4=C5	4.7573e-2 C4=C6
	-1.1589e-2	-1.6720e-2	1.7959e-2		-4.4700e-2	-5.3248e-2
Cluster 5	8.1903e-3	-1.9277e-3	1.2280e-2	3.5953e-3		6.6350e-3
	1.4813e-2 C1=C5	2.4833e-2 C2=C5	2.7328e-2 C3=C5	1.2583e-2 C4=C5		5.8991e-2 C5=C6
	-1.3828e-2	-2.0848e-2	-2.2622e-2	-1.1928e-2		-6.3224e-2
Cluster 6	1.8131e-2	8.0130e-3	2.2221e-2	1.3536e-2	9.9407e-3	
	1.6586e-2 C1<C6*	2.6127e-2 C2=C6	2.3997e-2 C3=C6	1.3248e-2 C4<C6*	1.8263e-2 C5=C6	
	-1.7630e-2	-2.4755e-2	-2.3723e-2	-1.3465e-2	-2.1122e-2	

14 Our analyses revealed strong overlap between this clustered pattern and spatial distribution of
15 ecological site-quality indicators. Tick infestation and site occupancy timing showed strong year-
16 to-year correlation across all 8 sampled years (slope c_t , intercept i_t and significance level of the
17 correlation p_t for tick load from one year to the previous one, starting in 2004: $c_{2004}=0.400$
18 $i_{2004}=0.006$ $p_{2004}\approx 0$; $c_{2005}=0.465$ $i_{2005}=0.034$ $p_{2005}\approx 0$; $c_{2006}=0.146$ $i_{2006}=0.058$ $p_{2006}<0.05$; $c_{2007}=0.406$
19 $i_{2007}=0.021$ $p_{2007}\approx 0$; $c_{2008}=0.577$ $i_{2008}=0.068$ $p_{2008}<0.001$; $c_{2009}=0.424$ $i_{2009}=0.062$ $p_{2009}<0.01$;
20 $c_{2010}=0.192$ $i_{2010}=0.037$ $p_{2010}<0.01$; $c_{2011}=0.743$ $i_{2011}=0.063$ $p_{2011}\approx 0$. For site occupancy timing from
21 each year to the previous one, $c_{2007}=0.526$ $i_{2007}=0.279$ $p_{2007}\approx 0$; $c_{2008}=0.712$ $i_{2008}=0.112$ $p_{2008}\approx 0$;
22 $c_{2009}=0.462$ $i_{2009}=0.331$ $p_{2009}\approx 0$; $c_{2010}=0.664$ $i_{2010}=0.226$ $p_{2010}\approx 0$; $c_{2012}=0.159$ $i_{2012}=0.062$ $p_{2012}<0.01$;
23 $c_{2013}=0.010$ $i_{2013}=0.250$ NS. Too much missing data in 2011 led us to discard this year, and to

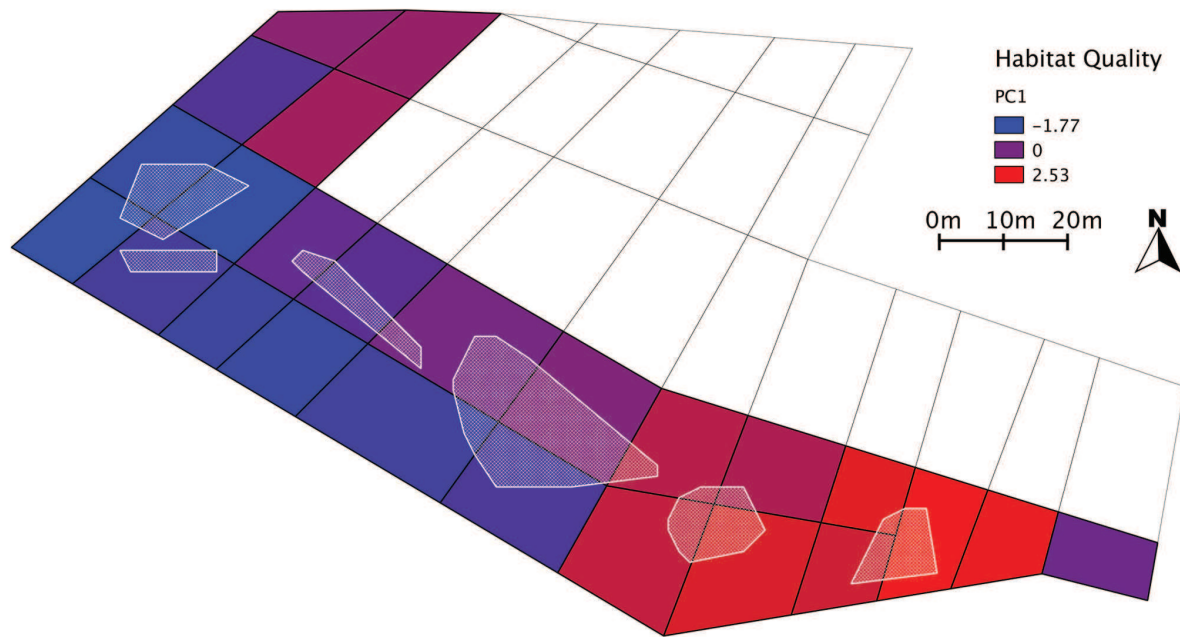
Fine-scale colony structure - §68

1 correlate 2012 values with 2010). Chick survival was found to be correlated between 2010 and
2 2011 (slope 1.548, intercept 0.168, p-value 0.0427), sampling did not overlap sufficiently be-
3 tween years 2011 and 2012 for comparison. All three variables showed a markedly structured,
4 gradient-driven distribution across our sampling range (Bivand 2013) and were found to be
5 strongly autocorrelated (Moran's I, two-sided test, using the closest 5%, or 8 nearest neighbours,
6 all p -values < 0.01). Occupancy timing and chick survival distributions appeared to largely over-
7 lap, and to be partly complementary to tick infestation distribution. Comparison between C1
8 and C5 was especially significant for all variables: tick infestation varied from 0.188 ± 0.095 in
9 C1 to 0.083 ± 0.019 in C5, while averaged chick survival ranged from 0.208 ± 0.058 in C1 to
10 0.398 ± 0.130 in C5, and occupancy timing index from 0.303 ± 0.206 in C1 to 0.488 ± 0.056
11 in C5 (Fig. 17). Principal component analysis results support this view: the first axis, which ac-
12 counted for 40.5% of the total variation (correlation coefficients, contributions: chick survival
13 0.82, 55.6%, occupancy timing 0.62, 31.6%, tick load -0.39, 12.8%), separates between the
14 south-eastern end of the colony, of higher habitat quality, and the western end, of lower habitat
15 quality (Fig. 18). However, neither of the individual habitat quality descriptors correlates directly
16 with individual-level or cluster-level genetic descriptors.



1 **Figure 17 | Ecological descriptors of breeding-site quality exhibit a strongly heterogeneous dis-**
 2 **tribution across the colony.** (A) Adult tick load, averaged for years 2005-2012. Contour lines: 0.02 steps. (B)
 3 8-neighbour chick survival, averaged for years 2010-2012. Contour lines: 0.1 steps. (C) Site occupancy chronology,
 4 averaged for years 2006-2013. Ratio of brooding birds amongst 50 randomly selected breeders. Contour lines: 0.1
 5 steps.

Fine-scale colony structure - §69



1 **Figure 18 | First axis of principal component analysis as a summary habitat-quality index.** White
2 polygons: clusters C1 to C6.

3 *Discussion*

4 *§-69 A near-panmictic population.* Our data on colony-wide spatial structure appear to contradict
5 the hypothesis of a strong, homogeneous process shaping the genetic structure of our studied
6 king penguin sub-population. Hardy-Weinberg equilibrium held across all loci, population-wide
7 F_{IT} was not significantly skewed (Wright 1965; Holsinger & Weir 2009), and spatial autocorrela-
8 tion tests performed across the whole sampling area were non-significant (Epperson 2005; Kanno
9 *et al.* 2011). Analyses of intrinsic partitioning indeed systematically pointed to a non-structured
10 population (Falush *et al.* 2003). Although tests of genotypic differentiation did yield significant
11 results for cluster C3, this finding is subject to caution, as it is not confirmed by pairwise F_{ST}
12 comparisons. This discrepancy may be due to the non-homogeneous inbreeding probability, as in
13 inbred individuals, genotype distribution departs from the neutral expectations (and thus be bio-
14 logically meaningful); however, G-test has been shown to often be over-sensitive (Balloux &

1 Lugon-Moulin 2002), and to yield numerous false positives (Ryman *et al.* 2006), this result
 2 should therefore be taken with caution, as it is not supported by other tests. These findings allow
 3 us to reject strong underlying structure resulting from constant, homogeneous processes such as
 4 multi-generation isolation of sub-colony units. Yet, in wild populations, a lack of global visible
 5 genetic structure does not necessarily imply the absence of structuring processes (Osborne *et al.*
 6 2007). Indeed, spatial structure may evolve as a secondary dependency to heterogeneous habitat
 7 features (Majumdar *et al.* 2011; Miller 2012).

8 *§-70 Heterogeneity in inbreeding probability.* Our investigations of inbreeding probability high-
 9 light the fact that analyses taking into account possible heterogeneity in structuring processes do
 10 point towards a non-panmictic population. Observed population inbreeding descriptor $F =$
 11 0.231 is remarkably high - it implies a homozygosity level close what is expected in a half-sibling
 12 relatedness context, for example (Wright 1978; Ballou 1997). Although homozygosity-based in-
 13 breeding calculations do not necessarily imply true (i.e. pedigree-deduced, Keller & Waller 2002)
 14 inbreeding (Balloux *et al.* 2004), our result shows a strong bias in the mating system, resulting in
 15 non-random gene mixing. This interpretation is supported by the individual inbreeding distribu-
 16 tion, which significantly deviates from an equivalent random distribution, with outbred classes
 17 being strongly under-represented as compared to a non-related group of individuals (Keller &
 18 Waller 2002). Moreover, more inbred individuals tend to be concentrated at the south-eastern
 19 end of the sampling range of our sub-colony (clusters C5 and C6), while more outbred individu-
 20 als are more frequent at the north-west end (cluster C1). Pairwise relatedness within spatially de-
 21 limited areas has also been shown to reflect weak genetic structure in wild populations (Jones &
 22 Wang 2012). Not surprisingly, we found that individuals that are significantly related to their di-
 23 rect neighbours tend to cluster in areas that largely overlap more inbred areas.

24 *§-71 Pairwise cluster differentiation.* Our pairwise cluster analysis supports such a structure. Geno-
 25 typic differentiation points towards structural divergence between the centre and the extremities
 26 of our sampling range. Yet, the genotypic structure detected is globally weak. This is an expected

Fine-scale colony structure - §72

1 result, as Waples and Gaggiotti (Waples & Gaggiotti 2006) showed that microsatellite-based
2 genotypic cluster divergence faded out when movement among clusters exceeded 10%, which is
3 likely the case in seabird colonies (e.g. Steiner 2005). More importantly, the north-west cluster
4 (C1) appeared significantly less inbred than any other cluster. Similarly, C1 and C4 both showed
5 low internal pairwise relatedness levels. This trend was not marked enough to be significant
6 throughout all pairwise comparisons (comparison between C1 and C5, for instance, was near-
7 significant).

8 These different elements outline an underlying structure in which the north-west end of the
9 colony (cluster C1 in particular) appears more outbred, with lower pairwise-relatedness levels,
10 and the south-east end (especially cluster C5) is more inbred, with higher relatedness levels. Such
11 a structure implies that the probability for a breeder to mate with a related individual is not con-
12 stant across the colony: potential mates are more likely to share identical alleles (either by descent
13 or by state) at the south-east end of the colony than at the north-west end. This result is in accor-
14 dance with the findings of Steiner and Gaston (Steiner 2005), who showed, on phenological
15 grounds, that, in colonial seabirds, even a very short-distance intra-colony dispersal may signifi-
16 cantly impact breeding success and inbreeding avoidance.

17 *§-72 Heterogeneity in breeding site quality.* Interestingly, the observed heterogeneity in genetic
18 structure we found within this king penguin colony appeared to be related to the quality of the
19 breeding sites. Tick infestation has a local maximum around C1, i.e. at the north-west end of the
20 range, and lower values elsewhere. In contrast, chick survival has local maxima around C5 and
21 C6, and very low values in C1. The chronology of settlement-site occupancy also appeared to be
22 strongly associated with the quality of the breeding sites. Indeed preferred areas at the south-east
23 end of the colony were occupied first, and less-favoured spots at the north-east end were occupied
24 by late breeders. The fact that no significant correlation could be found between habitat descrip-
25 tors and genetic parameters may be a consequence of very limited sample size, as our sampling
26 area is not large enough to allow for replicate situations. It may also be explained by the limited

Fine-scale colony structure - §73

1 number of ecological parameters monitored, as several other constraints (such as flooding, fine-
2 scale variations in predation pressure, or heterogeneous inter-annual variability) may also interact
3 to define high- and low-quality habitat.

4 However, overall, principal component analysis reveals a gradient from higher habitat quality at
5 the south-eastern end of the colony, around clusters C5 and C6, to lower quality at the western
6 end (clusters C1 and C2), and this pattern globally mirrors genetic heterogeneity structure. This
7 fact may be explained as being a consequence of differences in fine-scale philopatric behaviour:
8 while king penguins have been shown to select their breeding location according to site-specific
9 characteristics at the breeding-season scale (Bried & Jouventin 2001), our results suggest that
10 breeding experience may drive settlement site choice across multiple generations, and that this
11 process may be widespread enough to impact genetic patterns in the colony.

12 *§-73 Effects of demography and life-history.* Part of the relative weakness in the genetic signal may
13 be explained by the limited number of markers used; however microsatellite loci have a high mu-
14 tation rate in spheniciforms (Shepherd & Lambert 2005; Hart *et al.* 2011), and generally enough
15 polymorphism to allow for fine-scale genetic comparisons within a population (Gorospe & Karl
16 2013).

17 A possible bias due to the exclusively peripheral sampling design (required in order to minimize
18 colony disturbance) may also be considered: indeed, the peripheral areas have generally higher
19 tick load and later settling dates. However, central areas do suffer from specific pressure, such as
20 higher density and associated stress levels (Vibblanc *et al.* 2014). Predation pressure, on the other
21 hand, is not expected to differentiate peripheral and central areas. Although it has been shown
22 that immediately peripheral birds spend twice as much time interacting with predators than cen-
23 tral birds (Coté 2000), most of the chick mortality has been shown to happen during winter, es-
24 pecially through predation by the giant petrel (*Macronectes giganteus*, Hunter 1991), at a time
25 when chicks have gathered in creches. Moreover, our field observations tend to show that sum-
26 mer predation on eggs and small chicks by the subantarctic skua (*Stercorarius antarcticus*) is not

1 more frequent at the periphery than at the center of the colony. Finally, the overall breeding suc-
2 cess has been shown to be independent of central or peripheral location (Descamps *et al.* 2009),
3 which supports the idea that the purely geometric location in relation to the colony (such as cen-
4 ter-to-periphery gradients) is not the main determinant of habitat quality. We can therefore as-
5 sume that spatial processes driving breeding site choice identified on the edges of a colony should
6 also apply at its centre.

7 Another possible limit to our study may arise from our choice of sampling pre-dispersal chicks, as
8 opposed to adults, and to analyze nuclear autosomal markers (as opposed to sex-specific ones).
9 Indeed, a constant bias such as sex-specific dispersal may lead to more genetic structure in one sex
10 than the other, a phenomenon which may be blurred by our sampling of pre-dispersal individuals
11 of undetermined sex. Yet, although sex-specific dispersal is a common phenomenon in birds
12 (Clarke *et al.* 1997), on a multi-generation scale it is only expected to influence the structure of
13 sex-specific markers, and not of autosomal loci, which are subject to admixture (Prugnolle & De
14 Meeûs 2002). Thus, our sampling design should only reduce our ability to detect instantaneous
15 sex-specific phenomena, and not population-wide, sex-independent structures.

16 However, a more likely explanation of the weakness of the detected signal may lie in inter-annual
17 habitat variability: if local philopatric maxima are conditioned by the past breeding experience of
18 individuals, they may be subject to medium time-scale variation, that is, higher-quality breeding
19 sites may change every few generations, depending on flooding variations, ground erosion, or
20 parasite pressures. These very fine-scale shifts may be sufficient to weaken genetic structure at the
21 colony level on longer time scales.

22 Moreover, at the individual scale, it may be difficult for a breeder to retain its breeding site con-
23 sistent across all breeding seasons. Indeed, the complex breeding cycle of the King penguin ex-
24 tends for more than a year, implying that birds are alternately early or late breeder (Stonehouse
25 1960; Stonehouse 1960; Barrat 1976), regardless of individual quality (Le Bohec *et al.* 2007).
26 Therefore, even experienced breeders might sometimes be forced to settle in lower-quality spots

1 due to site-occupancy contingency at the season-level, and the output of their previous breeding
2 season. This process could maintain a sufficient rate of genetic mixing in the whole colony de-
3 spite a general trend towards local structure.

4 Finally, king penguin population genetic structure may still bear the imprint of recent demo-
5 graphic history. Most of the sub-Antarctic king penguin colonies suffered a drastic reduction at
6 the turn of the nineteenth century, as sealers slaughtered them to near-extinction on most of the
7 archipelagos for oil production. They recovered, but recently (Rounsevell & Copson 1982; Gales
8 & Pemberton 1988). Similar colonies on Macquarie island were reportedly reduced to less than
9 1,600 breeding pairs in 1937 (Falla 1937). Rapid recovery followed the cessation of slaughtering
10 (Rounsevell & Copson 1982; Delord *et al.* 2004), and population growth rate is suggested to
11 have neared intrinsic growth rate for three decades (Delord *et al.* 2004). This complex demo-
12 graphic history is also expected to have had an impact on colonial genetic patterns. Even under
13 the assumption of strong, consistent processes driving structure across several generations, recent
14 rapid growth may weaken a long-term signal in present-day colonies (Waples & Gaggiotti 2006),
15 though in that case, the use of more markers or more individuals may increase the strength of the
16 signal (Banks & Peakall 2012).

17 *§-74 Heterogeneity as a mixing mechanism.* Despite these different factors, the convergence of ge-
18 netic and ecological indicators outlines the structural heterogeneity inherent to penguin colonies,
19 and probably to other colonial seabirds (such as guillemots or eiders, Frederiksen & Petersen
20 1999; Pearce *et al.* 2005). However, our data point towards generally transient processes, which
21 are not homogenous and stable enough in time to impact colony structure on a large spatial and
22 temporal frame. Local heterogeneity appears to be dependent on medium-scale factors, such as
23 individual experience and breeding history, but also present-day site quality, and should therefore
24 be considered as a fundamental feature of coloniality in seabirds: local philopatric hotspots are
25 counterbalanced by active outbreeding regions, thus preserving global genetic diversity and mix-
26 ing at the colony level. Instead of being an exception in an otherwise strongly philopatric system,

1 we therefore believe that these lower-quality, strongly outbred areas are playing an active role in
2 mitigating the potentially drastic effects of strong natal philopatry on local genetic drift and loss
3 of diversity. Our findings illustrate how individual life-history decisions, such as site-fidelity or
4 dispersion, are related at the colony level with local environment features. These constitute true
5 colony processes that actively enhance higher-scale population functions such as genetic mixing
6 and inbreeding avoidance, thereby allowing the persistence over time of philopatric colonial
7 systems.

8 *Appendices*

9 We are very grateful to Magali Beaughey for her help in the lab. We also thank M. Le Vaillant,
10 O. Prudhomme and M. Ripoché for the sample collection. This study was supported by the
11 Institut Polaire Français Paul-Emile Victor (Programme 137), the Centre National de la
12 Recherche Scientifique (Programme Zone Atelier de Recherches sur l'Environnement
13 Antarctique et Subantarctique), the Fondation de France / Fondation Ars Cuttoli (to JDW), and
14 Marie Curie Intra European Fellowships (FP7-PEOPLE-IEF-2008, European Commission;
15 project No 235962 to CLB, and FP7-PEOPLE-IEF-2010, European Commission; project No
16 252252 to ET).

1 Chapter 4: The King synnome

2 This chapter is based on Cristofari R, Liu X, Bonadonna F, Cherel Y, Le Maho Y, Pistorius P,
3 Raybaud V, Stenseth NC, Le Bohec C & Trucchi E, **Stepping-stone range shifts in response to**
4 **climate change in the Southern Ocean.** (*in prep.*)

5 *Context*

6 *§-76 From local to global gene flow.* In the previous chapter, we investigated the patterns of very
7 fine-scale local dispersal within a colony of king penguins. The absence of intra-colonial genetic
8 structure suggests that philopatry does not apply to particular areas within the colony beyond a
9 handful of generations. However, at the colony level, philopatry is a strong and fairly well doc-
10 umented behaviour (see *e.g.* Barrat 1976; Bried & Jouventin 2001; Saraux *et al.* 2011b). Juvenile
11 return rates to the natal colony may reach up to 87% in that species (Saraux *et al.* 2011a). Yet,
12 while return rate may be measured accurately, assigning the respective roles of mortality and dis-
13 persal in the non-returning fraction of the population is a challenging task. At the single-genera-
14 tion scale, this may be a methodological dead-end: however, a lower bound may be placed on av-
15 erage dispersal rates at the coalescent scale using gene flow as an indicator of the migration
16 patterns of individuals. In the present chapter, we assess the extent of genetic structure between
17 colonies of King penguins on a global scale. Based on direct observation of philopatric behaviour

The King synnome - §77

1 (Le Bohec *et al.* 2008; Saraux *et al.* 2011b) and on the parallel with other seabird species (Friesen
2 *et al.* 2007), a certain degree of genetic isolation would be expected in this species. However, the
3 high dispersal potential of the King penguin, that is known to travel considerable distances out-
4 side of its breeding season (thousands of kilometres, see Bost *et al.* 2013), may promote long-dis-
5 tance gene flow and counteract the drift consequences of local philopatry: so that it is uncertain
6 which force, of philopatry and dispersal, has the upper hand in shaping the genetic patterns in
7 the species.

8 §-77 *Understanding the dynamics of contemporary structures.* As sentinel species of the Southern
9 Ocean (see §18 p. 65), King penguins are the object of a long-term monitoring program on Pos-
10 session Island (see §26 p. 81). However, extrapolating the findings of local surveys to the whole
11 species requires a set of assumptions that may not be easily verified. The degree of demographic
12 independence of the colonies, in particular, raises central concerns. Direct comparison of the de-
13 mographic trend of the five King penguin colonies on Possession Island showed that colonies fol-
14 lowed different demographic trajectories (Delord *et al.* 2004). However, true demographic inde-
15 pendence requires not only divergent, but fully uncorrelated trends between demes (Waples &
16 Gaggiotti 2006). In other words, the transfer of individuals from one colony to another results in
17 divergent demographic patterns, but also demographic dependency. Delineating the one from
18 the other is essential if we are to draw global conclusions from single-colony surveys: (i) true de-
19 mographic independence would imply that the main driver of colony trajectory is the local envi-
20 ronmental conditions, while (ii) complex demographic correlation between colonies may blur the
21 direct relationship between environmental conditions and local demographic patterns. In the first
22 case, climate-envelope-type model may be applied successfully, while more complex modelling
23 procedures may be necessary in the second case. Moreover, our understanding of the demograph-
24 ic trajectory of the King penguin is further complicated by its recent history of near-extinction, as
25 steep population growth in the recent past is mostly an artefact of population recovery (see §22

The King synnyme - §78

1 p. 75 for details): thus, our knowledge both of population structure, and of demographic trends,
2 should be rooted in a longer time frame in order to detach itself from these recent events.

3 *§-78 Integrating genetics with climate models.* Evolutionary and ecological proxies of species de-
4 mography are grounded in very different conceptual frameworks (see §2 p. 25 for an overview).
5 While the ecological paradigm relies on the short-term, local observation of the species in order
6 to extract larger order parameters, evolutionary studies use the information contained in the
7 species' coalescent structure to infer long-term demographic events. Relating these demographic
8 histories with environmental variables (typically climate) thus also relies on very different princi-
9 ples. A commonly used approach for ecological-scale surveys is the «climate-envelope» approach
10 (*e.g.* Pearson & Dawson 2003; Hunter *et al.* 2010; Péron *et al.* 2012; Jenouvrier *et al.* 2014), in
11 which precise, fully quantitative demographic data is the response variable explained by a wide
12 set of loosely correlated environmental parameters: the resulting correlative models can be ex-
13 tended into future climate projections, with a rather large set of of assumptions (*e.g.* fully realised
14 potential niche, or «relevant» - as opposed to coincidental - observed correlations, see Pearson &
15 Dawson 2003). Genetic data, however, will not usually lend itself to such an approach: demo-
16 graphic reconstructions are more qualitative than quantitative (in particular because both the size
17 of population and the extent of changes are scales by an unobserved long-term substitution rate,
18 and because of the loose relationship between effective and true population size, see §41 p. 103),
19 and few environmental variables are available on a comparable time scale (see §9 p. 42) - so that
20 the lack of correlation between reconstructed demography and climate (*e.g.* Morin *et al.* 2015)
21 may be either explained by the low sensibility of the species, or of the methods. The preferred ap-
22 proach must be robust to these shortcomings: for example, a hypothesis-testing framework may
23 be used instead of an agnostic exploratory one. Prior conceptions about the determinants of a
24 species' range or demographics may be examined throughout the history of the species, and their
25 long-term validity can be assessed, without necessarily establishing a direct, quantitative relation-
26 ship between them. This is the approach that we develop in this chapter. Relying on the general

1 characteristics of the King penguin's breeding and foraging biology, we extract traits that appear
2 to be proximal determinants of its survival and breeding success, and broad, easily observable in-
3 dices related to these traits, in order to assess their (mostly qualitative) realisation at different pe-
4 riods of time.

5 *Abstract*

6 Range shift is the primary short-term response of species to rapid climate change, but may be
7 challenged by natural or anthropogenic ecosystem fragmentation. Using the global King penguin
8 (*Aptenodytes patagonicus*) population as a model for top-predator climate-induced range shift in a
9 fragmented landscape, and integrating a large set of ecological and genomic data with general cir-
10 culation models, we show that the panmictic King penguin population responded strongly to cli-
11 mate change throughout the late Quaternary, and predict that up to 70% of the present-day pop-
12 ulation will be threatened before the end of the century if no immediate step is taken. By
13 showing how habitat fragmentation can exacerbate the effect of global warming towards biodiver-
14 sity collapse, we stress the urgency of tangible actions to limit global temperature increase.

15 *Results*

16 §-79 *Expected responses to climate change.* While the impact of anthropogenic global changes on
17 biological communities is beyond question (Thomas *et al.* 2004; Pacifici *et al.* 2015), the nature
18 and extent of species responses are generally still imperfectly understood (Pereira *et al.* 2010; He
19 & Hubbell 2011; Pacifici *et al.* 2015). At the species level, three fundamental mechanisms inter-
20 act in determining the response to climate change: (i) *phenotypic plasticity* (Charmantier *et al.*
21 2008; Hoffmann & Sgrò 2011), (ii) *adaptive microevolution* (Bradshaw & Holzapfel 2006; Hoff-
22 mann & Sgrò 2011), and (iii) *range shift* (Chen *et al.* 2011; Beaugrand *et al.* 2015). Although
23 important in the short-term, plasticity has a high fitness cost (Van Buskirk & Steiner 2009) and
24 may rapidly reach its limits (Auld *et al.* 2009). On the other hand, the fast rate of current climate

The King synnome - §80

1 change (Mahlstein *et al.* 2013) may outpace the potential rate of microevolutionary responses, es-
2 pecially for species that have long generation times (Hoffmann & Sgrò 2011) and low genetic di-
3 versity (Hoffmann & Sgrò 2011; Norberg *et al.* 2012). Thus, range shift is emerging as a major
4 feature of species response to fast environmental change (Wiens *et al.* 2010). However, the frag-
5 mentation of ecosystems, mostly due to a growing anthropogenic pressure, implies that newly
6 available habitats may be beyond reach for most of the species. Moreover, the uncoupling of cli-
7 mate change between geographical areas (Burrows *et al.* 2014) imposes additional constraints on
8 species that rely on different territories (*e.g.* migratory species, or long-distance central-place for-
9 agers). This is especially true in the Alpine and Polar Regions (Williams *et al.* 2007), which may
10 act as “climate sink areas” where local suitable ecological conditions disappear instead of being
11 displaced (Burrows *et al.* 2014). In addition, habitat fragmentation may exacerbate the effect of
12 climate change for higher trophic level taxa: in particular, the complex response of top predators,
13 which integrate changes occurring at all levels of the trophic network (Zarnetske *et al.* 2012) and
14 largely determine ecosystem resilience (Heithaus *et al.* 2008), may be the key to understanding
15 the depth of current changes, and forecasting ecosystem collapses (Scheffer *et al.* 2001; Barnosky
16 *et al.* 2012).

17 *§-80 Habitat constraints for the King penguin.* The highly constrained niche of the long-lived sub-
18 antarctic King penguin (*Aptenodytes patagonicus*) makes this species a useful model of top-preda-
19 tor response to environmental change (Le Bohec *et al.* 2008). While a poleward range shift is the
20 predicted response to climate warming for cold-adapted species, the highly fragmented nature of
21 the King penguin’s habitat precludes continuous population displacement. Because the species
22 breeds exclusively on year-round ice-free areas on islands scattered throughout the Southern
23 Ocean (Bost *et al.* 2013), it may only disperse in a stepping-stone manner amongst the few avail-
24 able islands. The species’ foraging grounds, on the other hand, move together with the myc-
25 tophid fish stock that strives around the Antarctic Polar Front (APF) (Cherel *et al.* 2002; Bost *et*
26 *al.* 2015). Since the King penguin’s breeding system, that spans over 12 months, makes it espe-

The King synnome - §81

1 cially sensitive to the duration of its central-place foraging trips during the early chick rearing pe-
2 riod (Barrat 1976; Bost *et al.* 2015), foraging distance at the peak of summer is a critical driver of
3 breeding success (Le Bohec *et al.* 2008; Péron *et al.* 2012; Bost *et al.* 2015). The most extensively
4 studied colony, belonging to the most important breeding area for the species (the Crozet archi-
5 pelago, Bost *et al.* 2013), appears to have benefited from Holocene warming (Trucchi *et al.*
6 2014). However, recent tracking studies have revealed a southward extension in their foraging
7 ranges as a result of climate change (Péron *et al.* 2012; Bost *et al.* 2015). As a result of the associ-
8 ated increase in energy expenditure, the Crozet population is expected to decline within the com-
9 ing decades (Le Bohec *et al.* 2008; Bost *et al.* 2015). The continuous poleward displacement of
10 the species' foraging grounds, combined with the discrete distribution of its breeding locations,
11 implies that King penguin populations must undergo abrupt location shifts from one island to
12 another to follow their habitat, thereby increasing the risk of mismatched strategies.

13 §-81 *A Synnome species.* In order to predict the limits and opportunities for this species to track
14 its fragmented habitat, we adopted here a cross-disciplinary approach, integrating information
15 from ecology, behaviour and genomics, together with multi-proxy paleoclimate reconstructions
16 and numerical climate models. The most striking feature of the present-day King penguin popu-
17 lation is its worldwide panmixia, that we explain by a high inferred migration rate among
18 colonies. Our genome-wide data, including ~35,000 independent polymorphic DNA loci gener-
19 ated using a RAD-sequencing approach (Baird *et al.* 2008), and gathering 163 individuals from
20 13 different locations covering most of the King penguin's contemporary range (see S-0 p. 177,
21 and Fig. 22), strongly contradict the alleged separation between the South Atlantic *patagonicus*
22 and the South Indian and Pacific *halli* subspecies (Mathews 1911; Barrat 1976), suggesting that
23 the traits used as a basis for subspecies delineation are better explained by phenotypic plasticity
24 than by reproductive isolation. Both classical descriptors of genetic variation and structure analy-
25 sis unambiguously support a fully-panmictic worldwide population (see S-1 p. 181 for details).
26 Full admixture among colonies is also clear when repeating these analyses at the island level (see

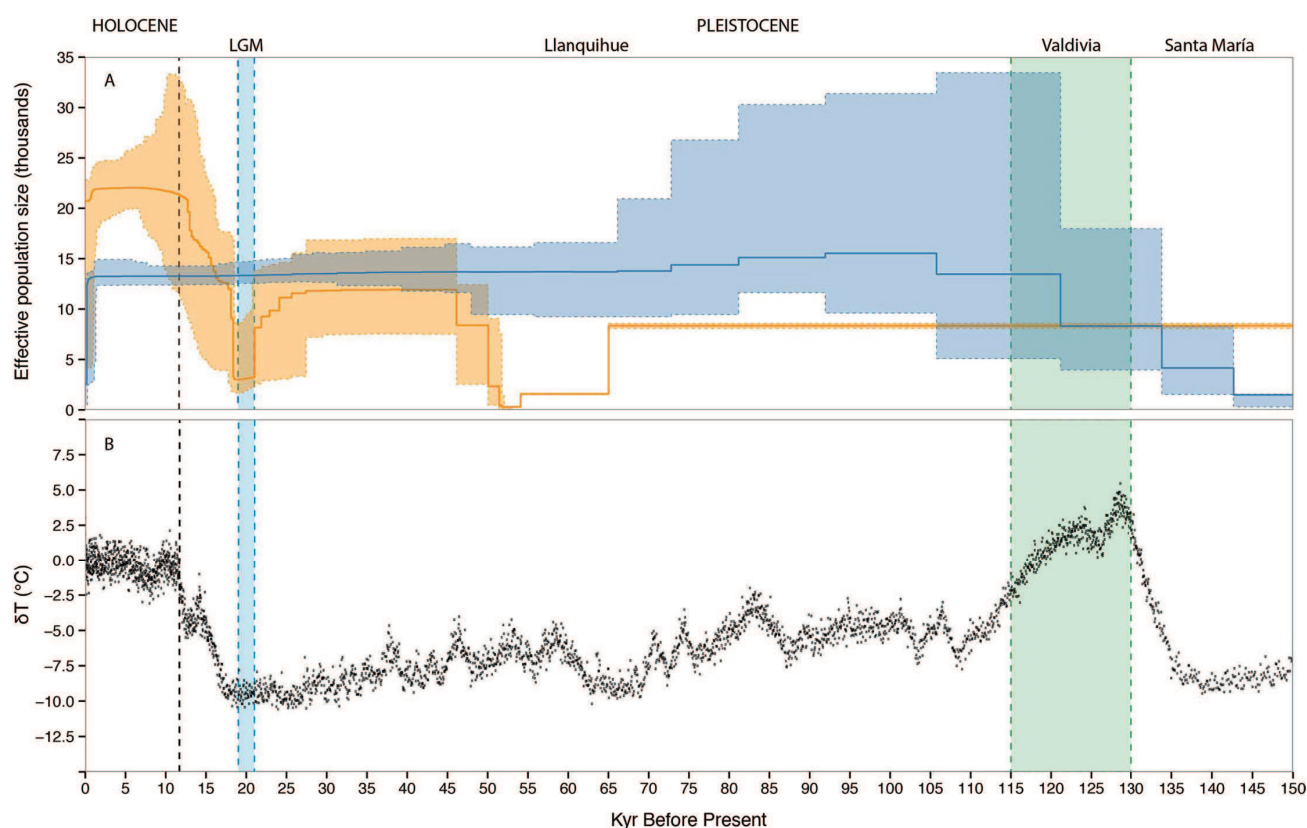
The King synnome - §82

1 S-1 p. 181). Furthermore, bio-logging experiments and empirical observations suggest that short
2 and long distance movements are significant contributors to the ongoing genetic mixing. Recent-
3 ly reported new colonies (Convey *et al.* 1999; Budd 2000; Pistorius *et al.* 2012) indicate that im-
4 migration-fuelled colony establishment can occur at a decadal scale. Contrary to previous hy-
5 potheses, recapture of tagged individuals (see S-2 p. 196 for details) shows that dispersal is also
6 strong at the generation-scale: thus, dispersal ability is not a limiting factor in the King penguin's
7 response to environmental change, and the species can be assumed to realise its fundamental
8 habitat efficiently, and to follow a single, unified demographic trajectory.

9 *§-82 Palaeodemography of the *Aptenodytes penguins*.* The value of the King penguin as a climatic
10 bio-indicator is confirmed by its strong demographic response to Quaternary climate change,
11 that we reconstructed using RAD-sequencing and full-genome data. We relied on a novel model-
12 flexible approach (the Stairway plot; §95 p. 186), based on the composite likelihood of the de-
13 rived-allele frequency spectrum (Liu & Fu 2015), as well as on extended Bayesian skyline plot
14 (§96 p. 187) and pairwise sequentially Markovian coalescent (§97 p. 190) approaches. The King
15 penguin population experienced two bottlenecks: (a) a recent one during the Last Glacial Maxi-
16 mum (LGM), and (b) a more ancient one overlapping with the previous Pleistocene glacial
17 episode (Fig. 19). During the late Pleistocene and early Holocene, a period of steep population
18 growth is followed by a long plateau. In order to assess the importance of general marine produc-
19 tivity variation in this response, we repeated the analysis for the high-Antarctic Emperor penguin
20 (*A. forsteri*), the only other extant large penguin species, using 110 individuals and an identical
21 sequencing approach. In contrast with the King penguin, the Emperor penguin shows no de-
22 tectable demographic response to the recent Pleistocene and Holocene climatic events (Fig. 19
23 and S-3 p. 197), but rather a slow increase during the Pleistocene, with a possible population
24 maximum after the Valdivian interglacial (Fig. 19). The large King penguin population fluctua-
25 tions are not mirrored in the Emperor penguin, excluding interspecific food competition. Yet, all
26 reconstruction methods support the replacement of the cold-adapted Emperor penguin by the

The King synnome - §82

- 1 warmer-adapted King penguin as the major *Aptenodytes* species already during the early
- 2 Holocene.



- 3 **Figure 19 | Penguin paleodemography in response to Quaternary climate change. A. Stairway plot**
- 4 **reconstruction of population size changes** for the King penguin (orange) and the Emperor penguin (blue). Solid
- 5 **line: median population size; shaded area: 95% confidence interval. B. Temperature anomaly in the late Qua-**
- 6 **ternary, as inferred from the EPICA Dome C ice core (Augustin *et al.* 2004). Highlighted areas: Last Glacial Maxi-**
- 7 **mum (LGM, ~21-19 Kyr BP) and Valdivian interglacial period (~130-115 Kyr BP). Dashed line: Pleistocene-**
- 8 **Holocene transition (~11.7 Kyr BP).**

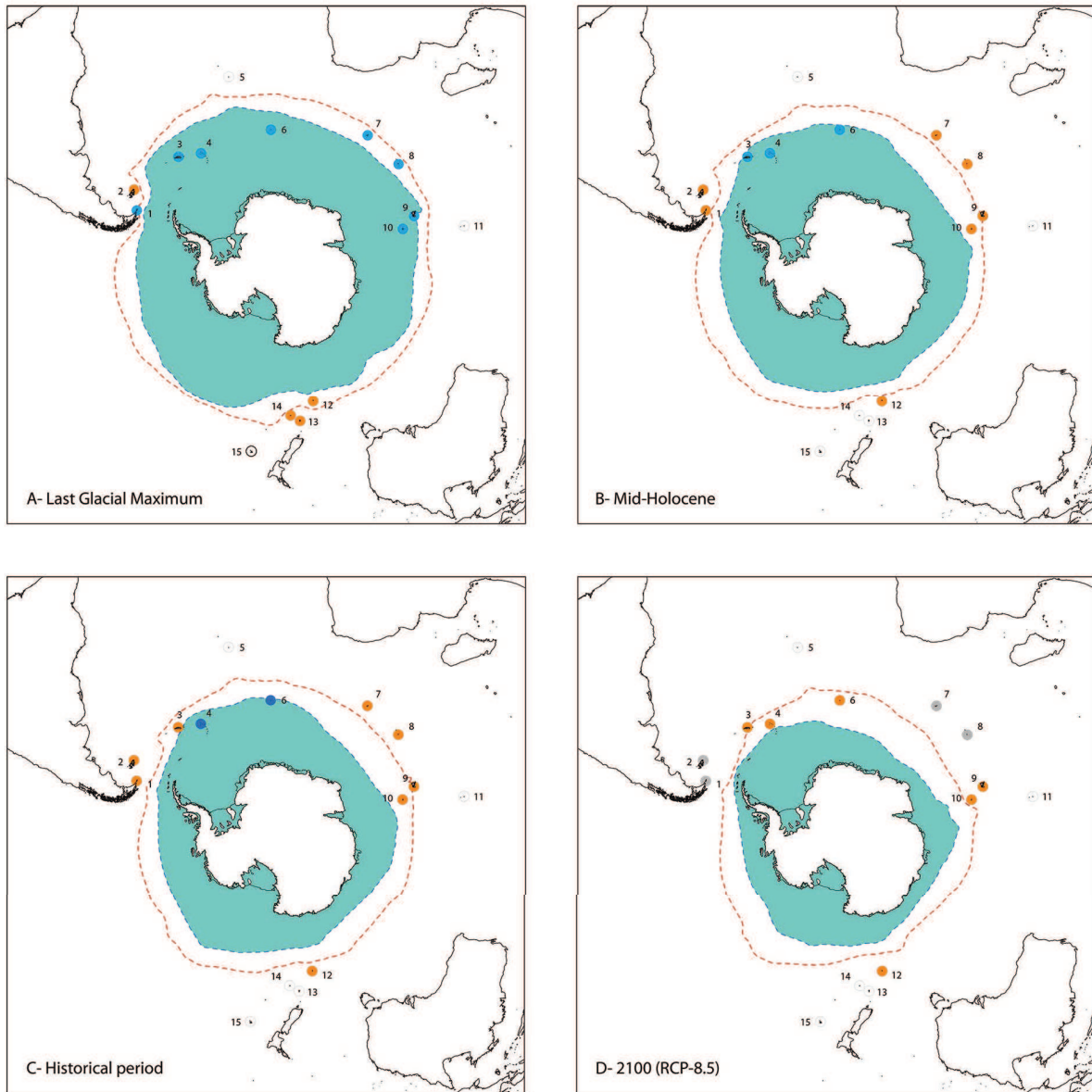
- 9 The King penguin's response to past climate change is best explained by variations in the extent
- 10 of suitable habitat (see S-4 p. 200 for details). We relied on both observed and modelled paleocli-
- 11 matic data to identify the extent of the King penguin's past fundamental niche, which we defined
- 12 based on three major traits that directly determine habitat suitability: (a) *within foraging distance*
- 13 *of the prey stock at the APF* (Bost *et al.* 2009), (b) *reduced sea ice extent to allow for overwinter*
- 14 *chick-rearing* (Barrat 1976), and (c) *insular and ice-free land* (Barrat 1976). As confirmed by the

The King synnome - §82

1 absence of LGM impact on the Emperor penguins (Fig. 19), the overall productivity of the
2 Southern Ocean did not change significantly during the Pleistocene and Holocene periods (Wolff
3 *et al.* 2006; Kohfeld *et al.* 2013). On the contrary, the location of the APF zone and the extent of
4 land ice and winter sea ice cover exhibited important latitudinal variation over the period (Ger-
5 sonde *et al.* 2005; Kohfeld *et al.* 2013; Hodgson *et al.* 2014). Thus, the location of optimal King
6 penguin breeding areas changed vastly between warm and cold conditions. APF and foraging
7 range predictions, based on historical period (1981-2005) experiments from an ensemble of
8 IPCC CMIP5 models (see S-4 p. 200 for details), closely matched both observed APF and em-
9 pirical foraging distances derived from bio-logging experiments (§103 p. 202). Our model is able
10 to capture the full present-day range of the King penguin, and our paleohabitat reconstructions
11 are also in close agreement with the species' reconstructed demography (Fig. 19 and Fig. 20). Un-
12 der LGM conditions, the equatorward displacement of the APF and increased land and sea ice
13 cover (Kohfeld *et al.* 2013; Hodgson *et al.* 2014) reduced the King penguin's range to a fraction
14 of its current extent (Fig. 20A), as suggested by the inferred population bottleneck (Fig. 19A).
15 Assuming a 700-km February foraging distance as the upper limit for successful breeding (Péron
16 *et al.* 2012), the only two possible refugial areas were found in the Falklands, and in the Camp-
17 bell plateau region, a much reduced refugial range compared to the eight pre-industrial breeding
18 areas (Bost *et al.* 2013). By mid-Holocene, on the other hand, the King penguin already occu-
19 pied most of its pre-industrial range (Fig. 20B-C). The APF occupied a position close to its
20 present-day state at most locations, while all present-day breeding archipelagos (except for South
21 Georgia) were free from sea ice, and land ice receded early on Kerguelen and South Georgia - al-
22 though it persisted until early Holocene on Crozet and Prince Edward archipelagos (Hodgson *et*
23 *al.* 2014). The King penguin rapidly exploited these newly available locations, as evidenced by
24 the steep growth and following plateau in our demographic reconstructions. Thus, the King pen-
25 guin's response to past climate change strongly supports the idea that modifications in the posi-

The King synnome - §82

- 1 tion of the APF and in the distribution of land and sea ice, by modifying the extent of available
- 2 habitat, have a major impact on the species' demographic trajectory.



- 3 **Figure 20 | Past and future breeding range of the King penguin.** Reconstructed position of the Antarctic
- 4 Antarctic Polar Front in February (SST = 5°C, dashed red line), and extent of sea ice in September (SIC > 15%, light blue
- 5 blue area) at four contrasting time periods: **A. Last Glacial Maximum** (21-19 Kyr BP), **B. Mid-Holocene** (6 Kyr BP), **C.**
- 6 **Historical period** (1981-2005), **D. Projection for 2100** (RCP-8.5 forcing scenario). Occupation status of the is-
- 7 lands: **orange:** presence of King penguin breeding colonies, **blue:** sea and/or land ice preventing colony foundation,
- 8 **grey:** too far from the Antarctic Polar Front for foraging, **white:** never occupied by King penguins. **Islands:** 1: Tierra
- 9 del Fuego, 2: Falklands, 3: South Georgia, 4: South Sandwich, 5: Gough, 6: Bouvet, 7: Marion and Prince Edward,
- 10 8: Crozet, 9: Kerguelen, 10: Heard and McDonald, 11: Amsterdam, 12: Macquarie, 13: Auckland, 14: Campbell,
- 11 15: Chatham.

The King synnome - §83

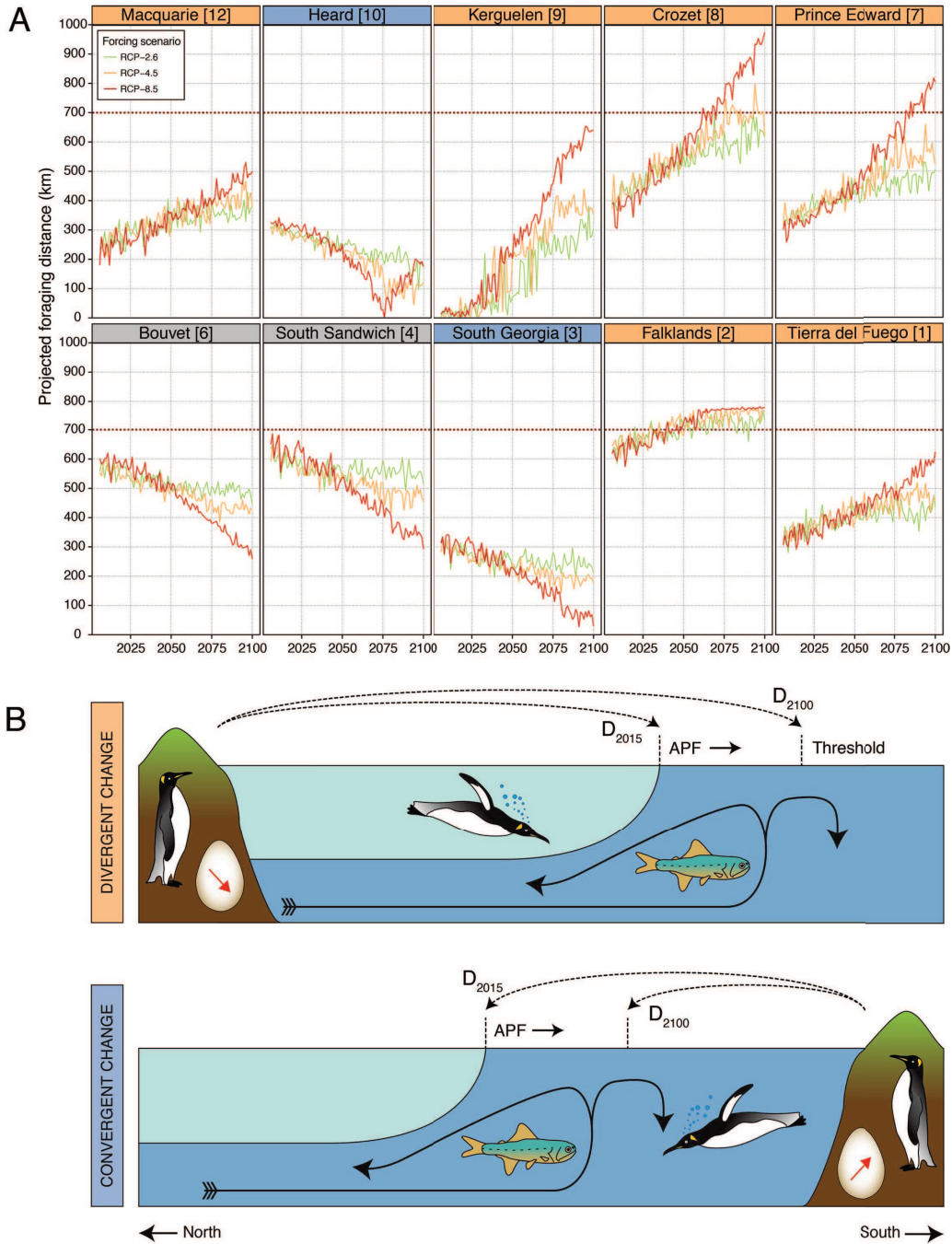
1 *§-83 21st century range shifts.* Projected changes for the 21st century are expected to have a deep
2 impact on the King penguin's range and population size (see S-4 p. 200 for details). The uncou-
3 pled trends in (i) the mobile food resources of the APF and (ii) the static breeding locations may
4 have opposite effects depending on the initial state (Fig. 21): foraging distance increases steadily
5 until the end of the century for the world's largest colonies, located north of the APF; but the
6 conditions become more favorable on the colder archipelagos south of the APF, with shorter for-
7 aging distances and decreased sea ice - a trend that is consistent across individual models (Fig. 21,
8 see §104 p. 204), and supported by three different Representative Concentration Pathways (rcp)
9 forcing scenarios (Meinshausen *et al.* 2011). With its low genetic diversity and long generation
10 time, the species is not expected to undergo rapid adaptive evolution to adapt to the new condi-
11 tions at the northern end of its range (Hoffmann & Sgrò 2011; Norberg *et al.* 2012): local ext-
12 inction or dispersal, rather than adaptation, is therefore the forecasted outcome.

13 Colony loss is likely to bring about a decrease in population size, although high dispersal ability
14 also implies that newly available locations may be colonised rapidly. Under the “*business-as-usual*”
15 rcp-8.5 scenario, 70% of the present-day 1.6 million King penguin breeding pairs (Bost *et al.*
16 2013) are expected to abruptly relocate or disappear before the end of the century. Forty-nine
17 percent of the world population are projected to lose their habitat completely by 2100 (on Crozet
18 and Prince Edward), and ~21% may also see their habitat degrade strongly due to regularly near-
19 limit foraging distances (on Kerguelen, Falklands and Tierra del Fuego). These losses may be
20 partly compensated by the predicted colonisation of Bouvet, and by a possible additional growth
21 on Heard and South Georgia due to improved foraging conditions. These last two locations, to-
22 gether with Macquarie Island, are likely to become the major refugia for the King penguin in the
23 coming decades. Under the low-emission rcp-2.6 scenario, however, only Crozet and Falkland
24 populations come under direct threat, while other colonies may retain good foraging conditions

The King synnome - §84

1 (Fig. 21), and undergo minimal demographic impact: thus, our results stress the importance of
2 immediate action to limit radiative forcing, as efficient attenuation strategies may still have a pos-
3 itive outcome for the Southern Ocean biodiversity. We also insist on the importance of taking
4 preemptive conservation measures in areas of the Polar Regions, such as Bouvet, that may act as
5 cold biodiversity refugia for the coming warm-earth conditions if we fail to limit radiative forcing
6 in time.

7 *§-84 Uncertainties and perspectives.* Thus, although the impact of these changes on the global
8 King penguin population size will largely depend upon the relative roles of dispersal, breeding
9 failure, and mortality in colony extinctions, the forecasted re-shaping of the species' distribution
10 is considerable. Nevertheless, our projection is likely to be an underestimate, as we only take into
11 account the maximum foraging distance after which no successful breeding may take place.
12 However, increasing foraging distances even below the 700 km-limit have been shown to impact
13 breeding success strongly, and may trigger a colony decrease well before the extinction threshold
14 is reached (Le Bohec *et al.* 2008; Bost *et al.* 2015). In addition, our model does not take into ac-
15 count aggravating effects of climate change, such as sea level rise (Rahmstorf 2007) or decrease in
16 ocean productivity due to ocean acidification (Jackson 2008) and reduction of the global thermo-
17 haline circulation (Kuhlbrodt *et al.* 2009). The abrupt nature of the predicted King penguin
18 range shift may additionally accelerate the restructuring and concentration of biotic interactions
19 (*e.g.* range overlap and competition with other penguin species): this in turn can generate com-
20 plex feedback effects (Davis *et al.* 1998; Norberg *et al.* 2012), thereby decreasing the predictabili-
21 ty of Southern Ocean biodiversity, and prompting us to adopt an explicitly precautionary ap-
22 proach to marine resource management.



1 **Figure 21 | Projected foraging distance under three RCP scenarios.** (A) *Mean projected summer foraging*
 2 *distance* for an ensemble of 15 CMIP5 general circulation models, over the 21st century, for the 8 currently occu-
 3 *pied archipelagos, and the two possible future breeding archipelagos (Bouvet and South Sandwich), under three*
 4 *different forcing scenarios (for inter-model variability, see S5). Horizontal red line: 700 km limit, beyond which no*
 5 *successful breeding is expected. Header numbers refer to Fig. 21. Header colour according to present-day status and*
 6 *projected trend at the location; orange: north of the APF, increasing foraging distance, blue: south of the APF, de-*
 7 *creasing foraging distance, gray: currently unoccupied islands south of the APF. RCP-8.5: colonies are predicted to*
 8 *(i) disappear from Crozet and Prince Edward, (ii) undergo significant population decline or disappear in Kerguelen*
 9 *and newly-colonised Tierra del Fuego, (iii) remain unchanged in Macquarie Island, (iv) grow on South Georgia and*
 10 *Heard, and (v) settle on Bouvet, and possibly the South Sandwich, as the winter sea ice disappears. RCP-2.6 and*

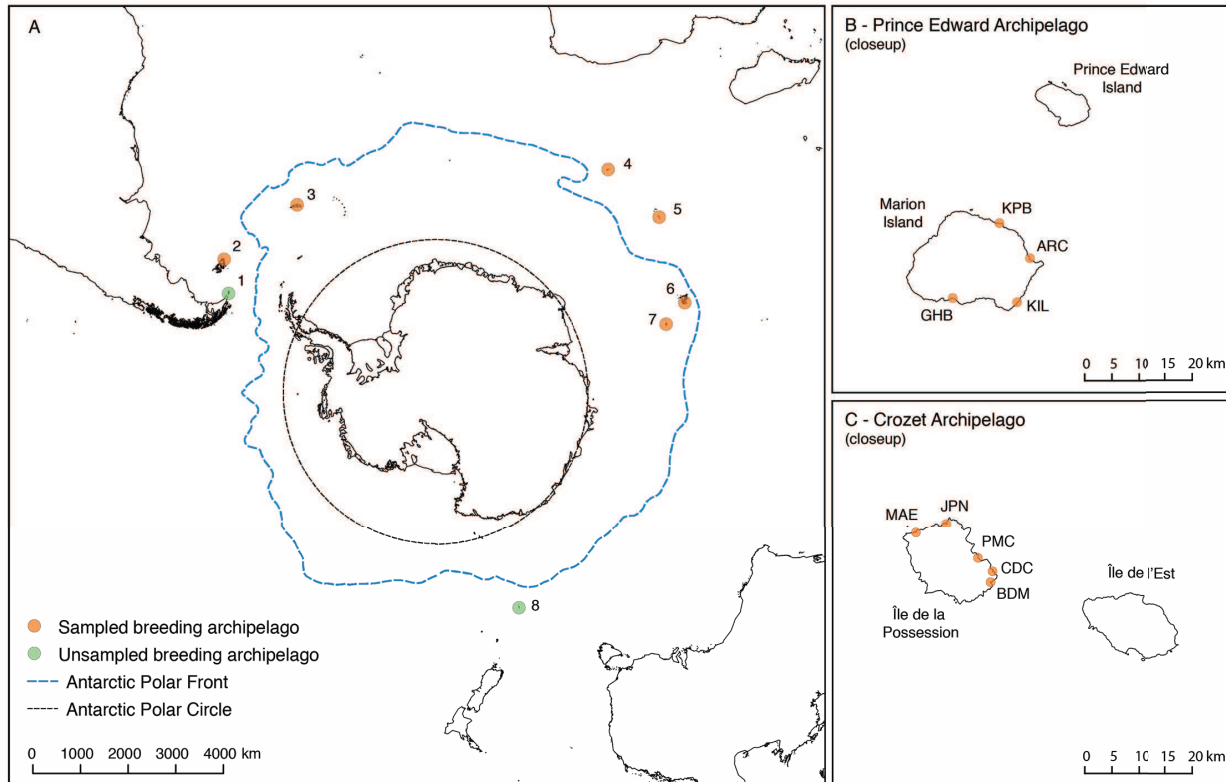
The King synnome - §84

1 RCP-4.5: only Crozet and Prince Edward are too far from the APF to sustain large breeding colonies by 2100, while
2 Kerguelen retains a favourable situation. (B) Schematic representation of the different results of climate change in the
3 Southern Ocean. Dark and light water masses: cold antarctic deep water and warmer subantarctic surface water (ma-
4 jor circulation as a black arrow). APF: Antarctic Polar Front. Dashed arrows: foraging trips in 2015 and 2100. The
5 red arrow in the egg represents the trend in breeding success.

6 The penguin example illustrates how data about habitat distribution and connectivity, dispersal
7 abilities, trophic interactions, and past range shifts need to be integrated into atmospheric and
8 oceanographic scenarios in order to capture the full extent of human impact on threatened
9 ecosystems. Beyond its implications for the Southern Ocean, the King penguin's complex step-
10 ping-stone trajectory offers a paradigmatic representation of the impact of global warming on
11 species distributions, whenever heterogeneous environmental change leads to the uncoupling of a
12 species' different critical areas (*e.g.* breeding, foraging, or overwintering grounds), and thus results
13 in mismatched strategies. As we show here, mismatching between feeding and breeding strategies
14 is strongly aggravated in fragmented habitats - a growing pressure in urban or agricultural matri-
15 ces, that has been identified as the pivotal aspect of species extinction worldwide (Haddad *et al.*
16 2015). Habitat fragmentation increases the risk of divergent trends in the different aspects of a
17 species' niche, while reducing corridors that may allow continuous niche tracking: by forcing
18 species to undergo tipping-point range-shifts, it has the double effect of aggravating the impact of
19 environmental change, but largely masking it, placing populations in a situation of climatic debt
20 well before the critical threshold is reached. Thus, the King penguin's cautionary tale should in-
21 vite us to explore further the complex and sometimes paradoxical effects of climate change, even
22 in those ecosystems that may still appear most remote and pristine.

1 *Supporting information*

2 *S-0: Supplementary methods: from sample collection to SNP typing.*



3 **Figure 22 | Sampling design.** A) The King penguin’s range and sampling: (1) Tierra del Fuego, (2) Falklands,
 4 (3) South Georgia, (4) Prince Edward archipelago, (5) Crozet archipelago, (6) Kerguelen archipelago, (7) Heard is-
 5 land, (8) Macquarie island. B) and C) local sampling on Prince Edward and Crozet archipelagos (see §184 p. 347 for
 6 details).

7 *§-85 Sample collection and DNA extraction.* A total of 163 blood samples were collected from
 8 fledged King penguin juveniles, or from breeding adults, on thirteen colonies covering most of
 9 the species’ range, between 2010 and 2014. In order to assess fine-scale patterns, we sampled all
 10 five colonies from Possession Island, on Crozet archipelago (S46°24’ E 51°45’ - Baie du Marin
 11 “BDM”, N=15, Crique de la Chaloupe “CDC”, N=16, Petite Manchotière “PMC”, N=15,
 12 Jardin Japonais “JPN”, N=16, and Mare aux Elephants “MAE”, N=16 - all samples were fledged
 13 juveniles), and all four colonies from Marion Island (S46°54’ E37°44’ - Good Hope Bay “GHB”,

The King synnome - §86

1 N=10, Kildalkey Bay, Archway Bay “ARC”, N=10, and King Penguin Bay “KPB”, N=10 - all
2 samples were fledged juveniles). We sampled one colony from Kerguelen archipelago (S49°20'
3 E69°20' “KER”, N=16 - all samples were fledged juveniles), from Falkland archipelago (S51°45'
4 W59°00' - “FLK”, N=10 - all samples were breeding adults), from South Georgia (S54°15'
5 W36°45' - “GEO”, N=12 - all samples were moulting adults), and from Heard Island (S53°00'
6 E73°30' - “HEA”, N=7 - all samples were breeding adults). See Fig. 22 for the geographical loca-
7 tion of these colonies. Blood was stored in Queen’s lysis buffer at +4°C (Crozet, Marion, Kergue-
8 len), or centrifuged, and red blood cells stored in ethanol at -20°C (Falklands, South Georgia,
9 Heard). DNA was extracted using a spin-column protocol (Qiagen DNEasy© Blood and Tissue
10 kit) with minor modifications.

11 *§-86 Genome-wide Single Nucleotide Polymorphism (SNP) typing.* SNP discovery and sequencing
12 followed a single-digest RAD-sequencing protocol (Baird *et al.* 2008). Genomic DNA was
13 checked for degradation on a 1.5% agarose gel, and only samples with consistently high molecu-
14 lar weight were retained and quantified by fluorometry (Life technologies™ Qu bit®). The 163
15 samples was retained and sequenced in 6 distinct libraries. (i) approximately 150 ng of genomic
16 DNA per sample were digested with the restriction enzyme Sbf-I-HF (NEB); (ii) each sample
17 was then ligated to a unique barcoded P1 adapter prior to pooling in a single library. The library
18 was then sheared by sonication (7 cycles 30" ON – 30" OFF); (iii) sonicated libraries were con-
19 centrated to 25 µl by DNA capture on magnetic beads (beads solution:DNA = 0.8:1), thus fur-
20 ther reducing the carry-over of non-ligated P1 adapters, and the target size range fraction
21 (350-650 bp) was then selected by automated gel electrophoresis (BluePippin®); (iv) capture on
22 magnetic beads using the same beads:DNA ratio (0.8:1) was then employed in all following
23 purification steps (after blunt-end repairing, poly-A tailing, P2 adapter ligation and library en-
24 richment by PCR). Magnetic beads were kept together with the library throughout the pre-PCR
25 steps, and DNA was re-bound to the beads for purification using a PEG-8000 binding solution;
26 (v) PCR amplification was performed in 8 x 12.5 µl aliquots pooled after the amplification in or-

The King synnome - §87

1 der to reduce amplification bias on few loci due to random drift. PCR was performed using NEB
2 Phusion[®] polymerase with the following cycles: 30" denaturation at 98°C, 18 cycles of amplifica-
3 tion (10" at 98°C, 30" at 65°C, and 30" at 72°C), and a final elongation of 5' at 72°C; (vi) the li-
4 brary was then quantified by a fluorimetry-based method (Life technologies[™] Qubit[®]), and mo-
5 larity was checked on an Agilent Bioanalyzer chip (Invitrogen[™]). A final volume of 20 µl for each
6 library was submitted for paired-end sequencing on an Illumina HiSeq2000 sequencer (V3
7 chemistry, libraries 1-3), or HiSeq2500 (V4 chemistry, libraries 4-6), at the Norwegian Sequenc-
8 ing Centre, University of Oslo, spiked with 20% PhiX control library in order to reduce low-di-
9 versity bias.

10 *§-87 Sequence alignment and genotyping.* Data processing was performed using the following
11 workflow: (i) *Sequence demultiplexing.* Read quality assessment was made in FastQC ([http://](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
12 www.bioinformatics.babraham.ac.uk/projects/fastqc/). Samples were de-multiplexed according
13 to in-line barcodes using Stacks v1.28 (Catchen *et al.* 2011; Catchen *et al.* 2013), low-quality
14 reads were discarded, and sequences trimmed to 95 bp. (ii) *Read mapping and filtering.* Demulti-
15 plexed fastq files were mapped to the published contigs of the Emperor penguin genome (Zhang
16 *et al.* 2011b) using Bowtie2 2.2.35, with standard settings, allowing only end-to-end mapping.
17 Resulting SAM files were filtered using Samtools 0.1.196, PicardTools 1.113 (picard.sourceforge.net),
18 and custom R and shell scripts (github.com/rcristofari/RAD-Scripts.git) in order to
19 discard unpaired reads and full read pairs where at least one mate has a mapping quality score be-
20 low 30. The resulting BAM files were then filtered for PCR and optical duplicates by comparing
21 mapping position and CIGAR string, using Picard MarkDuplicates. This process also allowed to
22 filter out most sequencing errors, since MarkDuplicates only retains the read with the highest av-
23 erage Phred score in each duplicate cluster. (iii) *SNP calling and genotyping.* A draft SNP-calling
24 was done in Stacks v1.28 for the general assessment of the dataset, using the “rxStacks” correction
25 algorithm, with a maximum of 5 mismatches allowed between alleles at a single locus (both with-
26 in and between individuals). For SNP-based analysis, joint SNP and genotype calling was per-

The King synnome - §88

1 formed using the GATK HaplotypeCaller pipeline (DePristo *et al.* 2011), with standard parame-
2 ters, except for population heterozygosity which was set to 0.01. We retained only SNPs
3 genotyped in at least 75% individuals, or 90% for AMOVA and PCA analyses. (iv) *Allele-fre-*
4 *quency likelihood and allele frequency spectra.* ANGSD 0.9008 (Korneliussen *et al.* 2014) was used
5 to compute per-site probability of being variable, and raw genotype likelihoods, using the Sam-
6 tools mpileup/bcftools algorithm, and the complete sample allele frequency information as a pri-
7 or. Per-site allele-frequency likelihood distribution was used to produce a maximum-likelihood
8 estimate of the derived allele frequency spectrum, either unidimensional at the population or
9 species level, or pairwise joint spectrum between pairs of populations.

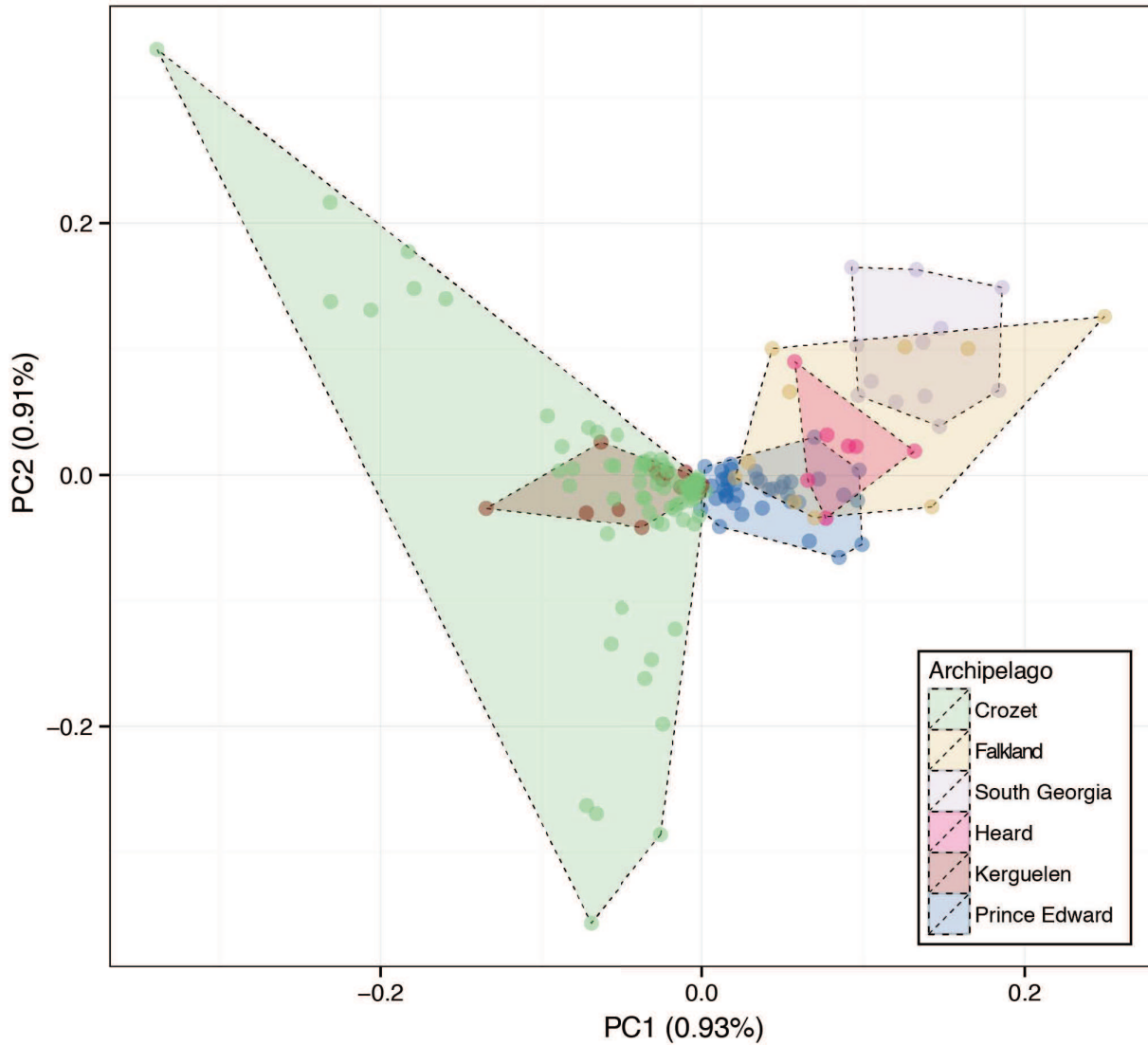
10 *§-88 Ancestral state reconstruction.* In order to polarize allele-frequency spectra, we reconstructed
11 the most likely ancestral base for all positions in the RADome. We selected 12 high-quality King
12 penguin samples covering the whole species' range, and 12 Emperor penguin samples processed
13 according to the same protocol (see *The Emperor synnome*, p. 211). We used BEDtools (Quinlan
14 & Hall 2010) and GATK's FastaAlternateReferenceMaker to update the published Emperor pen-
15 guin genome and establish a reference RADome for both the King penguin, and the Emperor
16 penguin, using only high-quality polymorphisms (phred-scale genotype quality ≥ 80). We aligned
17 this RADome to the Adélie penguin genome (*Pygoscelis adeliae*, Zhang *et al.* 2011a) using
18 Bowtie2, and extracted the corresponding regions. For each RAD locus, a maximum-likelihood
19 unrooted tree was built in PhyML (Guindon *et al.* 2010), and maximum-likelihood ancestral se-
20 quence for crown-Aptenodytes was reconstructed using PAML (Yang 2007) and Lazarus (project-
21 lazarus.googlecode.com/), using PhyML tree topology as a prior. Downstream analysis was re-
22 stricted to the sites that could be reliably polarized. All sites that were identified as belonging to
23 coding regions (Zhang *et al.* 2011b), or to sex chromosomes (Zhou *et al.* 2014), were excluded
24 from the analysis.

1 *S-1: Analysis of genetic data.*

2 *§-89 Summary statistics.* Summary statistics were calculated in Arlequin (Excoffier *et al.* 2005)
3 and with custom R scripts either from filtered SNP calls, or from short RAD haplotypes. Pairwise
4 fixation index (F_{st}), calculated using Reich's estimator (Reich *et al.* 2009), is close to zero (mean
5 pairwise F_{st} 0.0132 ± 0.00567). Nucleotide diversity π and Tajima's D were computed for full
6 RAD haplotypes. In order to avoid possible biases due to low coverage, we randomly sampled
7 one haplotype for each individual, and performed calculations on this haploid subset. Tajima's D
8 is slightly negative, and homogeneous across locations (D_{all} : -1.094 ± 0.672 , D_{HEA} : $-0.329 \pm$
9 0.925 , D_{KER} : -0.518 ± 0.899 , D_{CRO} : -0.546 ± 0.890 , D_{MAR} : -0.404 ± 0.00307 , D_{GEO} : -0.448
10 ± 0.925 , D_{FLK} : -0.312 ± 0.953), and nucleotide diversity is low (π_{ALL} : 0.00209 ± 0.00258 , π_{HEA} :
11 0.00201 ± 0.00326 , π_{KER} : 0.00215 ± 0.00304 , π_{CRO} : 0.00218 ± 0.00307 , π_{MAR} : $0.00200 \pm$
12 0.00306 , π_{GEO} : 0.00199 ± 0.00295 , π_{FLK} : 0.00182 ± 0.00294), in keeping with the prediction of
13 Romiguier *et al.* 2014 for long-lived species.

14 *§-90 Clustering analysis.* Clustering was performed in ngsAdmix (Skotte *et al.* 2013), based on
15 genotype likelihoods calculated in ANGSD with a SAMtools model, and allowing for a maxi-
16 mum of 50% missing data in order to process a site, and keeping only positions inferred as vari-
17 able with a high likelihood (p-value threshold $1e-6$). We performed 100 bootstrap replicates,
18 with K values ranging from 1 to 10. Best-fitting K was chosen using Evanno's δK method. An in-
19 dependent clustering was performed in FastStructure (Raj *et al.* 2014) using a filtered SNP
20 dataset (minimum depth of coverage of 4x, and minimum 80% individuals genotyped at each lo-
21 cus, leaving 4,784 polymorphic loci for analysis), again with 100 replicates and K ranging from 1
22 to 10. Both approaches unambiguously supported a K=1 model.

The King synnome - §91

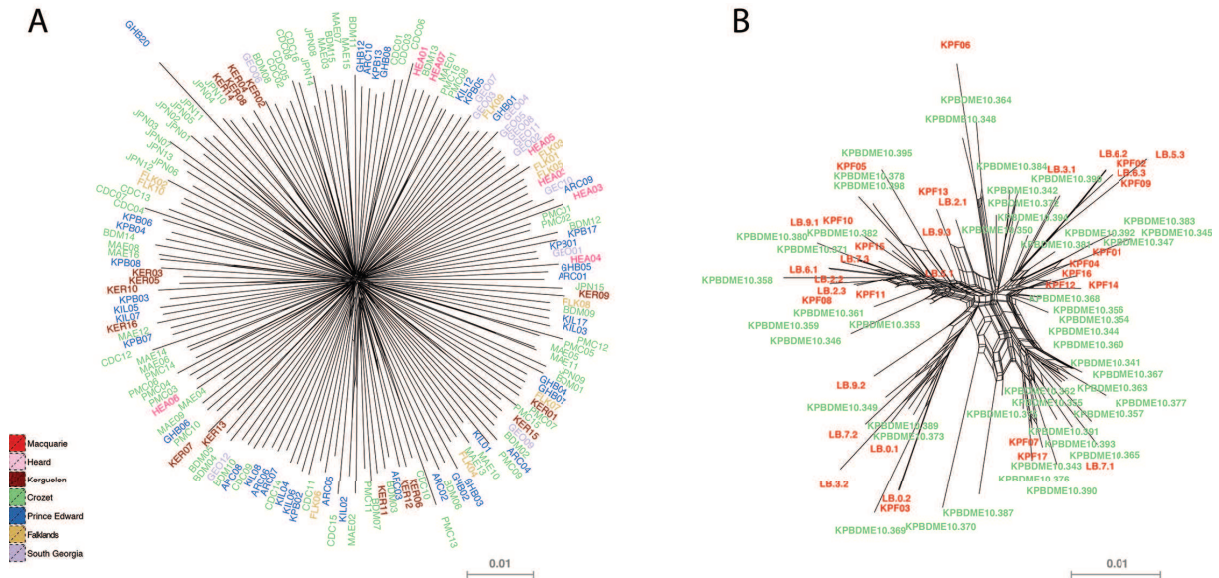


1 **Figure 23 | Principal component analysis.** As performed on genotype likelihoods in ngsAdmix (Skotte *et al.*
2 2013), retaining only variable loci. Shaded areas reflect archipelagos.

3 *§-91 Principal component analysis.* Genotype posterior probabilities calculated in ANGSD (§87 p.
4 179) were used to perform a principal component analysis (PCA) in ngsTools (Fumagalli *et al.*
5 2014), including only variable sites with a maximum-likelihood derived allele frequency at least
6 equal to $1/2N$ (with N being the number of included samples). PCA was repeated in the R pack-
7 age *adeigenet* (Jombart 2008), using a filtered SNP dataset (minimum depth of coverage of $4x$,
8 and minimum 80% individuals genotyped at each locus, leaving 4,784 polymorphic loci for
9 analysis). PCA does not resolve strong geographical structure (Fig. 23): although samples tend to

The King synnyme - §92

- 1 gather by archipelago, there is considerable overlap between locations, and no single principal
- 2 component explains more than ~0.9% of the total variation.



- 3 **Figure 24 | Neighbour-net.** Calculated A) from pairwise Hamming distances, based on genome-wide SNP
- 4 data, for 6 breeding archipelagos, and B) from the mitochondrial control region of 40 individuals from Crozet, and
- 5 39 individuals from Macquarie island.

6 *§-92 Analysis of molecular variance.* Analysis of molecular variance was performed in Arlequin

7 3.5.2.1 (Excoffier *et al.* 2005), using a filtered SNP set that included only sites genotyped in 90%

8 individuals. Amova was performed on a per-locus basis, with 10,000 permutations. We tested

9 four different grouping schemes: (i) *colonies grouped by archipelago*: ((HEA), (KER), (BDM,

10 CDC, PMC, JPN, MAE), (GHB, KIL, ARC, KPB), (GEO), (FLK)) (ii) *A. p. patagonicus vs A. p.*

11 *halli*: ((HEA, KER, BDM, CDC, PMC, JPN, MAE, GHB, KIL, ARC, KPB), (GEO, FLK)) (iii)

12 *Crozet-only*: ((BDM), (CDC), (PMC), (JPN), (MAE)) (iv) *Marion-only*: ((GHB), (KIL), (ARC),

13 (KPB)). Under all four groupings, the overwhelming majority of variance is explained at the indi-

14 vidual level: (i) 92.9% within individuals, 6.20% amongst individuals, 0.989% amongst popula-

15 tions, -0.124% amongst groups. (ii) 92.9% within individuals, 6.20% amongst individuals,

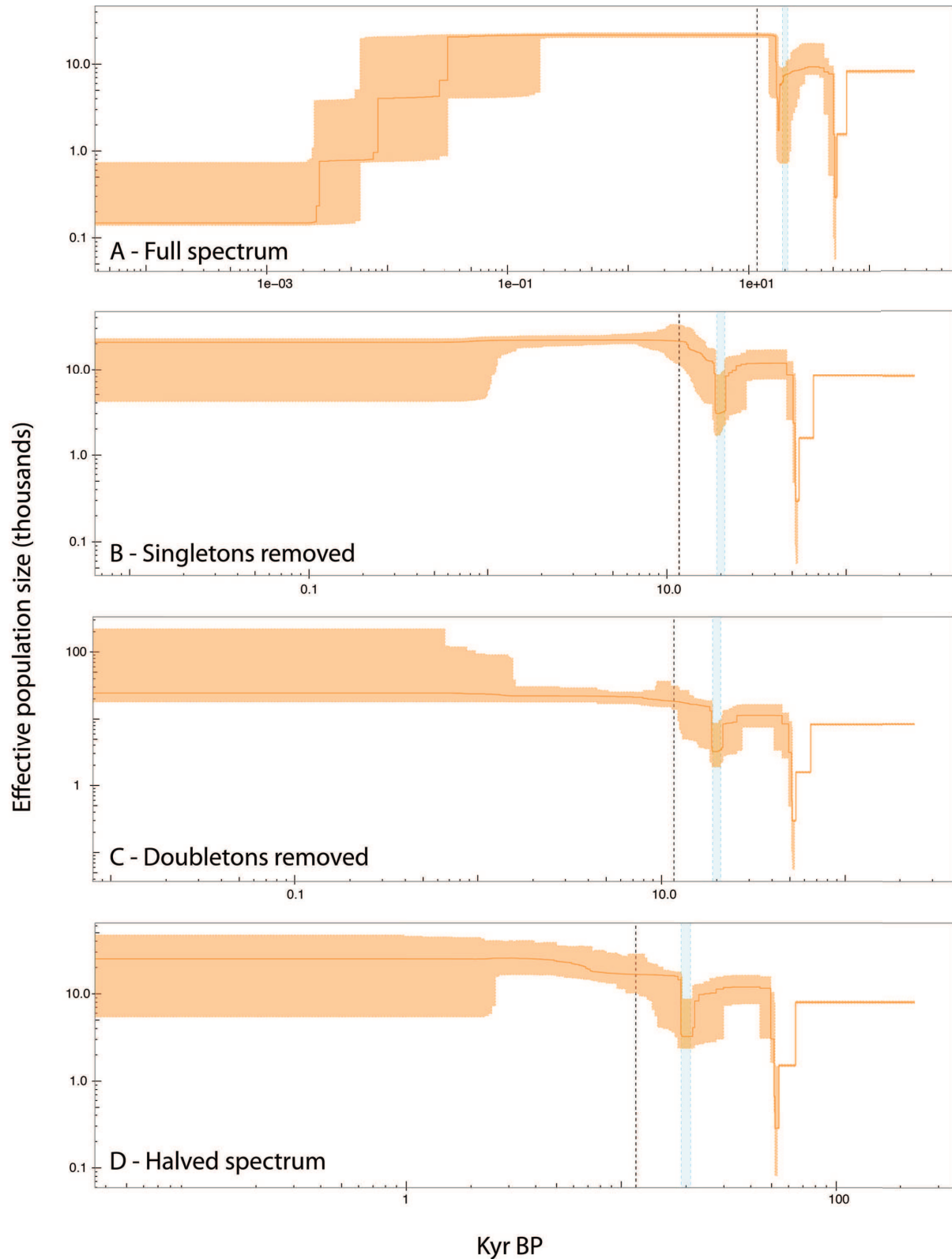
The King synnome - §94

1 0.904% amongst populations, -0.0370% amongst groups. (iii) 94.1% within individuals, 4.57%
2 amongst individuals, 1.30% amongst populations. (iv) 85.1% within individuals, 14.8%
3 amongst individuals, 0.0309% amongst populations.

4 *§-93 Pairwise Hamming distance network.* We calculated pairwise Hamming distance between in-
5 dividuals based on genotype calls using PLINK v1.9 (Purcell *et al.* 2007), and calculated the cor-
6 responding neighbour-net in SplitsTree (Huson & Bryant 2006) - see Fig. 24A. In keeping with
7 the results of AMOVA and PCA, the terminal branches explain most of the variance, and sam-
8 ples do not cluster geographically.

9 *§-94 HVR comparison with Macquarie.* Comparison of mitochondrial hypervariable control re-
10 gion (HVR) haplotypes from Crozet (Trucchi *et al.* 2014, Genbank accession number
11 KF530582-KF530621) with published sequences from Macquarie Island (Heupink *et al.* 2012,
12 Genbank accession number JQ256379-JQ256413) confirms the idea of a single, worldwide and
13 fully panmictic population. Pairwise F_{st} is low ($F_{st}=0.032$), and a haplotype network does not
14 support any population separation between the two islands (Fig. 24B).

The King synnome - §94



- 1 **Figure 25 | Robustness of the Stairway plot method.** Stairway plot reconstructions with the same set of
- 2 King penguin individuals, but based on A) the full spectrum as inferred from 140 samples, B) masking the singleton
- 3 loci, C) masking both singletons and doubletons, and D) masking singletons, and using only a random subset of 70

The King synnome - §95

1 individuals. Dashed black line: Pleistocene-Holocene boundary. Shaded blue band: last glacial maximum.

2 *§-95 Demographic reconstructions: the Stairway plot method.* The Stairway Plot method is a novel
3 method for demographic inference developed by Liu & Fu (Liu & Fu 2015). This model-flexible
4 method relies on the maximisation of the composite likelihood of the observed derived-allele fre-
5 quency spectrum, without prior hypothesis on demographic history, as opposed to previous spec-
6 trum-based demographic inference methods (*e.g.* Gutenkunst *et al.* 2009). Maximum-likelihood
7 estimation of the allele frequency spectrum was performed in ANGSD-0.901 under a SAMtools
8 model, for 140 high-quality King penguin samples, and 90 high-quality Emperor-penguin sam-
9 ples. Each spectrum was run along with 500 bootstrap replicates. Singletons were found to be the
10 least robustly estimated frequency class, due in particular to the confounding effect of sequencing
11 errors, and were consequently masked from the reconstructions - although comparison of recon-
12 structions (*i*) including all frequency categories, (*ii*) excluding singletons, or (*iii*) singletons and
13 doubletons show that only the reconstruction of the most recent demographic events are affected
14 by the low-frequency variants (Fig. 25A-C). Similarly, using only a randomly picked subset of
15 half of the individuals did not affect the reconstructions (Fig. 25D) **Generation time:** In a long-
16 lived species, generation time is not a fixed parameter, but rather a function of the demographic
17 trend. An estimator has been defined by Saether *et al.* 2005 as $\alpha + (S / (\lambda - S))$, where α is the age
18 at first breeding for females, S is the yearly adult survival rate, and λ is the yearly growth rate of
19 the population, defined as $\lambda=1$ for a stable population. Using long-term monitoring data, we ex-
20 tracted both yearly growth rate, and adult survival, from a pool of 400 adults of known age (Le
21 Bohec *pers. com.*), for the 1999-2010 period. S and λ were found to be strongly correlated over
22 that period (intercept: -0.2454, slope: 1.0936, $R^2=0.6$): therefore, we extended the empirical rela-
23 tionship between both parameters to our reconstruction. For each generation, the generation
24 time in years was therefore defined as $T_{t+1} = \alpha + (S_t / (\lambda_t - S_t))$, where λ_t and S_t are the growth

The King synnome - §96

1 rate and adult survival rate for the previous generation, defined as $\lambda_t = (N_{t+1} / N_t) \cdot e^{(1/T)t}$,
2 where N_{t+1} and N_t are the population sizes at generations $t+1$ and t , and T is the generation
3 time in years at generation t , and S_t is a linear function of λ_t , using empirically derived parame-
4 ters. This correction was applied recursively from the oldest generation in the reconstruction as-
5 suming $\lambda = 1$, and towards the present. In order to calibrate other analyses, the mean generation
6 time over the whole reconstruction $T = 10.6$ years was retained.

7 *§-96 Demographic reconstructions: the Extended Bayesian Skyline Plot (EBSP) method.* Accurate re-
8 construction of past and present population size changes requires a robust estimate of the substi-
9 tution rate. We performed a joint analysis of mitochondrial HVR and RAD data, in a multilocus
10 EBSP framework, using the robustly established substitution rate for the Adélie penguin HVR
11 (in substitutions per site per Myr: median = 0.55, 95% CI = 0.29–0.88 Millar *et al.* 2008) as a
12 calibration. Since the generation time differs widely between the Adélie penguin (6.46 years Mil-
13 lar *et al.* 2008) and the King penguin (10.48 years, see §95 p. 186), and since we are considering
14 the rate of substitution as determining the frequency of coalescence events, as opposed to the rate
15 of mutation (a purely physiological parameter - see Gibb & Hills 2013), we converted that rate
16 to reflect the difference in generation time, to 0.34 s substitutions.site⁻¹.Myr⁻¹ (95% CI = 0.18–
17 0.54).

18 We followed the protocol presented in §44 p. 112 and applied in §136 p. 252, a development of
19 the protocol of Trucchi *et al.* 2014, downsampling the data to haploid individuals, and using in-
20 dependent 50 loci with 50 haplotypes each, with 3 to 6 polymorphic sites, in addition to 50 ran-
21 domly selected HVR haplotypes. We specified one independent site model for each locus class (3,
22 4, 5 or 6 SNPs, and HVR). For each class, specified a HVR model, allowing for invariant sites for
23 the HVR, but not for the short nuclear loci, and for gamma-distributed rate heterogeneity discre-
24 tised in 4 classes. Transition-transversion ratio *kappa* was linked across nuclear models. All chains

1 were run in duplicate to check for convergence and for a sufficient length to gather ESS > 200 for
 2 all parameters, which necessitated 500,000,000 to 1,000,000,000 steps.

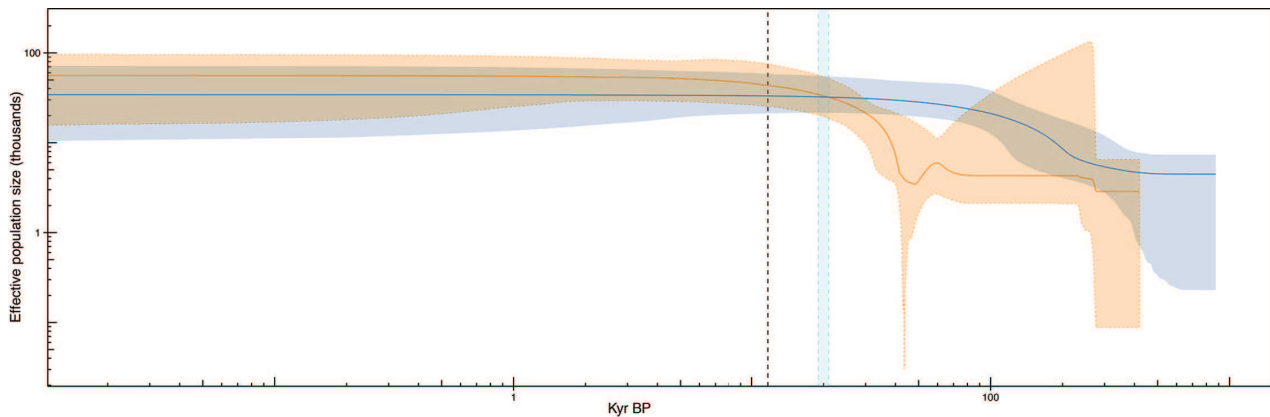
3 Since we parametrised each locus class separately, we expect our model to fit a class-specific sub-
 4 stitution rate as a function of the observed number of segregating sites, rather than a common
 5 substitution rate. However, as we focus on neutrally evolving regions of the genome, we expect
 6 the number of segregating sites to follow a Poisson distribution, of parameter λ equal to the mean
 7 number of segregating sites per RAD locus. On a large number of sequences, the expected value
 8 $E(\lambda)$ converges towards the “true” underlying constant mutation rate, multiplied by the total tree
 9 length for each locus. Thus, if we fix the tree length, λ becomes an estimator of the substitution
 10 rate μ . However, under the EBSP model, the observed number of segregating sites is taken as an
 11 estimator of λ , and consequently of the substitution rate μ . Therefore we expect the inferred value
 12 of μ for each locus class to be a posterior probability of the “true” substitution rate, conditional
 13 on the mean number of segregating sites observed for that class (Trucchi *et al.* 2014). In order to
 14 retrieve the underlying common substitution rate μ , we first fitted a log-linear model to the in-
 15 ferred substitution rates ($\mu_3= 0.0159$, $\mu_4= 0.0218$, $\mu_5= 0.0275$, $\mu_6= 0.0389$. Fitted model: inter-
 16 cept $i = -5.02$, slope $s = 0.292$, $R^2=0.997$). A Poisson model of parameter λ equal to the mean ob-
 17 served number of segregating sites was a good fit for the empirical distribution of number of
 18 segregating sites per locus ($\lambda=1.47$, chi-squared test of goodness-of-fit p-value= 0.232). Thus, we
 19 extracted μ as $e^{(s\lambda+i)} \sim 1.02e-2$ substitutions per site per Myr, or $1.08e-7$ substitutions per site per
 20 generation.

21 This rate is ca. twice slower than the one reported by Trucchi *et al.* ($2.6e-7$ subst.site⁻¹.genera-
 22 tion⁻¹, Trucchi *et al.* 2014), but much faster than the one reported by Li *et al.* ($8.11e-9$ sub-
 23 st.site⁻¹.generation⁻¹, Li *et al.* 2014). While the former was not used in Trucchi *et al.*'s analysis, but
 24 rather derived from it, Li *et al.*'s result, on the other hand, relies on two exterior and uncertain as-
 25 sumptions: 1) the divergence time between the Emperor and the Adélie penguin is set to ~23

The King synnome - §96

1 Myr, which may be a large overestimate (Gavryushkina *et al.* 2015, based on a state-of-the-art
2 total evidence Bayesian analysis, proposes ~9 Myr instead), and 2) the generation time is taken to
3 be 5 years in both species; however it has been shown to be 16 years in the Emperor penguin (Je-
4 nouvrier *et al.* 2014), and 6.46 years in the Adélie penguin (Millar *et al.* 2008) - thus 11 years
5 would be a closer (although inaccurate because assuming a single, constant rate) estimate of a
6 common generation time. Applying these corrected estimates to Li *et al.*'s findings would give a
7 rate of $\sim 4.55 \times 10^{-8}$, which is more than five times faster than proposed, and ca. half our estimate -
8 although this calculation does not take into account the possible rate heterogeneity between lin-
9 eages, and most importantly the changes in generation time between the *Aptenodytes/Pygoscelis*
10 common ancestor and the extant species, which may explain the remaining difference. Generally,
11 the rate of evolution of penguins has been a rather challenging subject, with a wide discrepancy
12 between the paleontological and molecular evidence. While fossil data has been recognised to
13 support a very recent radiation of penguins (about 10 Myr BP, see Slack *et al.* 2006; Clarke *et al.*
14 2007), molecular data has been interpreted as implying a much more ancient origin (~45 Myr for
15 Baker *et al.* 2006). This molecular-derived radiation has successively been brought to a closer
16 agreement with the fossil evidence by Subramanian *et al.* (~20 Myr, see Subramanian *et al.* 2013)
17 and Gavryushkina *et al.* (~12.5 Myr, Gavryushkina *et al.* 2015). The rate that we propose here is
18 in accordance both with the hypothesis of a very fast diversification of the spheniscids, and with
19 the findings of Trucchi *et al.* Our reconstruction supports the evidence provided both by the
20 Stairway plot analysis (§96 p. 187) and the PSMC analysis (§97 p. 190, with a fast expansion of
21 the King penguin population in the late Pleistocene, and a stable Emperor penguin population
22 throughout the period (Fig. 26).

The King synnome - §97



1 **Figure 26 | Extended Bayesian Skyline Plot.** Reconstruction of past population size changes for the King
2 penguin (orange) and the Emperor penguin (blue). Solid line: median population size; shaded area: 95% confidence
3 interval (see S2.4 for details). Dashed black line: Pleistocene-Holocene boundary. Shaded blue band: last glacial
4 maximum.

5 The EBSP demographic reconstruction shows only one bottleneck, and places it around 40 Kyr
6 BP - between the two Stairway-inferred bottlenecks. The contrast between the King and the Em-
7 peror penguin is maintained, with the Emperor experiencing only a slow and moderate expan-
8 sion before 100 Kyr BP, and the King going through more diverse demographic events in the late
9 Pleistocene. Our simulation tests (see §98 p. 192) show that, even when two bottlenecks are real-
10 ly present, the EBSP's expected behaviour is to smooth them out as one single broad population
11 depression (Fig. 28B). Thus, our reconstruction, although with a lower resolution, supports the
12 Stairway-inferred demography. The EBSP's lower resolution is not surprising, given that it only
13 includes a subset of 50 short loci (*i.e.* 250 to 300 SNPs), where the Stairway plot is using the in-
14 formation from every single genotyped SNP.

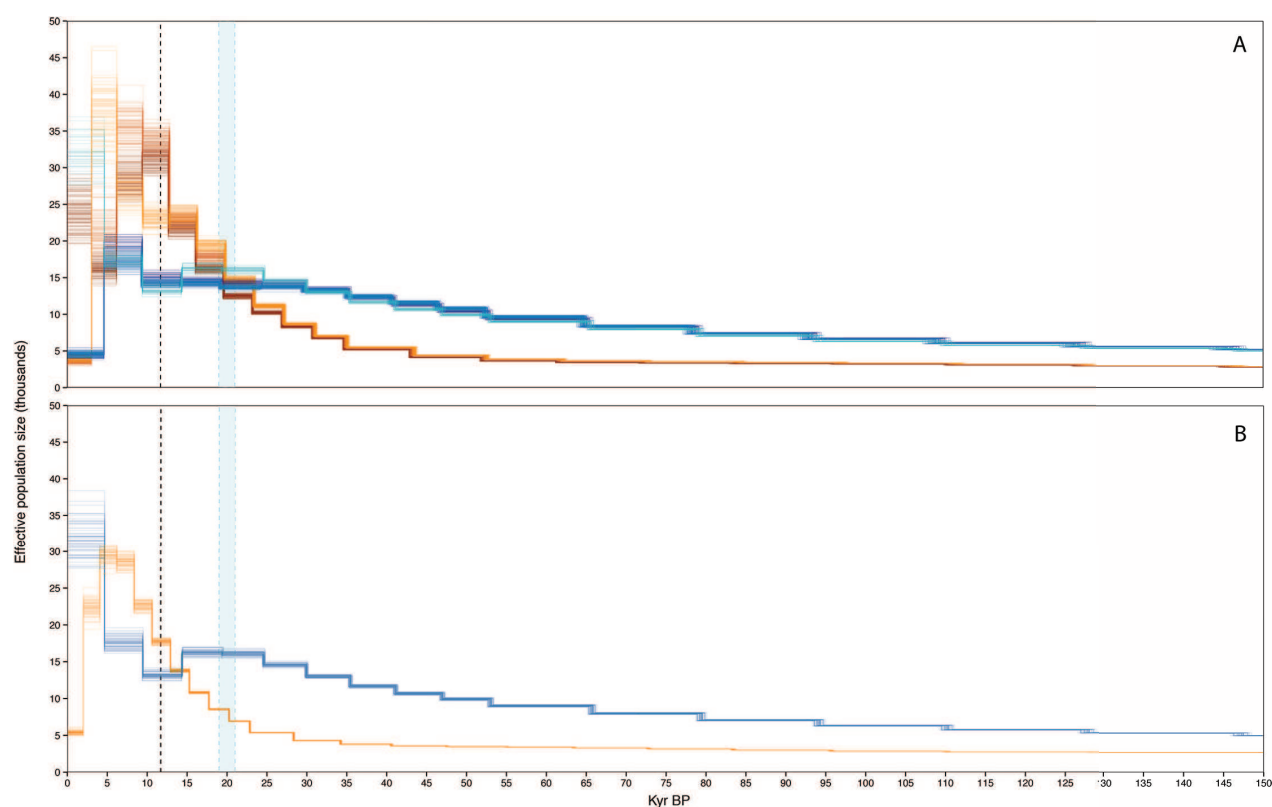
15 *§-97 Demographic reconstructions: the Pairwise Sequentially Markovian Coalescent (PSMC) method.*

16 Like the Stairway Plot and the EBSP methods, PSMC (Li & Durbin 2011; Schiffels & Durbin
17 2014) is a model-flexible method, that does not require prior specification of demographic
18 epochs or events. Instead of maximising the composite likelihood of the derived-allele frequency
19 spectrum (Liu & Fu 2015) or the full likelihood of short, non-recombining sequences (Heled &

1 Drummond 2008), the PSMC algorithm summarises the full ancestral recombination graph
2 through the depth of the most recent coalescence event (time to most recent common ancestor,
3 TMRCA) and total length of singleton branches, as a hidden Markov model in which recom-
4 bination events mark state changes. It allows for accurate reconstruction of deeper-time demo-
5 graphic events, although it lacks power for more recent time periods in its pairwise form (Li &
6 Durbin 2011; Schiffels & Durbin 2014). The full Multiple Sequentially Markovian Coalescent
7 approach (MSMC, Schiffels & Durbin 2014), which has a much improved resolution for recent
8 time periods, relies on the accurate phasing of haplotypes, which unfortunately is not possible in
9 a non-model species, in the absence of a large transmission or population dataset. In order to ex-
10 ploit unphased haplotypes, analysis must be restricted to the pairwise case, as PSMC'. However,
11 since recombination events are treated as a Markovian process along the sequence, it is still possi-
12 ble to increase the likelihood of the reconstruction by concatenating several genomes together,
13 thus increasing the independent sampling of TMRCA. We selected three high-quality samples for
14 the King penguin, and the Emperor penguin. Libraries were prepared with a standard Illumina(c)
15 TruSeq™ PCR-free protocol, and multiplexed on two lanes of a HiSeq 2500 V4 sequencer at the
16 Norwegian Sequencing Center facility, University of Oslo. Reads were mapped to the published
17 Emperor penguin genome (Zhang *et al.* 2011b) with high success (unique concordant alignment
18 rate, King penguin: ~86%, Emperor penguin: ~81%). We retained only longer scaffolds (length
19 ≥ 2 Mb, i.e. 188 scaffolds making up for ~80% of the total reference length) for the analysis.
20 Analysis was run on all three samples from each species simultaneously, with 200 bootstrap repli-
21 cates. Substitution rate and generation time were defined as above (§95 p. 186-§96 p. 187).
22 Results (Fig. 27) are very similar to the EBS analysis (§96 p. 187, Fig. 26): the King penguin
23 population grows rapidly in the late Pleistocene, while the Emperor penguin population is mostly
24 stable. However, the resolution of the PSMC' analysis is low in the recent periods, and the last 4
25 to 5 time bins exhibit considerable instability when compared across reconstructions (Fig. 27A),
26 as opposed to older time periods. Thus, the precise timing of the LGM bottleneck is not precisely

The King synnome - §98

1 retrieved for the King penguin: the two-step expansion since the mid-pleistocene (Fig. 19A) ap-
2 pears smoothed in one single growth trend. A similar behaviour can be reproduced when simu-
3 lating data with two bottlenecks in a rapid succession (see §98 p. 192): thus, our PSMC' analysis
4 is in accordance (although with much lower precision) with our general demography.



5 **Figure 27 | Pairwise Sequentially Markovian Coalescent.** Reconstruction of past population size changes
6 for the King penguin (orange) and the Emperor penguin (blue). Each individual line represents one bootstrap repli-
7 cate. Reconstruction was performed either **A)** for each individual separately (each shade represents one individual),
8 or **B)** concatenating genomic data from all three individuals for each species. Dashed black line: Pleistocene-
9 Holocene boundary. Shaded blue band: last glacial maximum.

10 *§-98 Reconstruction validation through simulation.* In order to assess the consistency of our recon-
11 struction, we simulated genetic data under the Stairway-plot demographic model for the King
12 penguin, and analysed it using all three algorithms (Stairway plot, EBSP, and PSMC'). Data was

The King synnome - §98

1 generated under a sequential Markovian coalescent model, either assuming equal substitution
2 and recombination rates (for the Stairway plot and PSMC'), or 95bp non-recombining hap-
3 lotypes (for EBSP), using *scrm* (Staab *et al.* 2015), to match the characteristics of the empirical
4 data, and was either directly converted to an allele-frequency spectrum (for the Stairway plot
5 analysis), or to sequence data, under an HKY model, using *seq-gen* (Rambaut & Grass 1997)
6 (for EBSP). Both the Stairway plot and the PSMC' approaches rely on bootstrapping, rather than
7 MCMC sampling (as EBSP does), for confidence interval estimation. Whereas the empirical data
8 was bootstrapped directly in a non-parametric way (see §96 p. 187 and §97 p. 190), here, we
9 replicated the full simulation 200 times to estimate the confidence intervals.

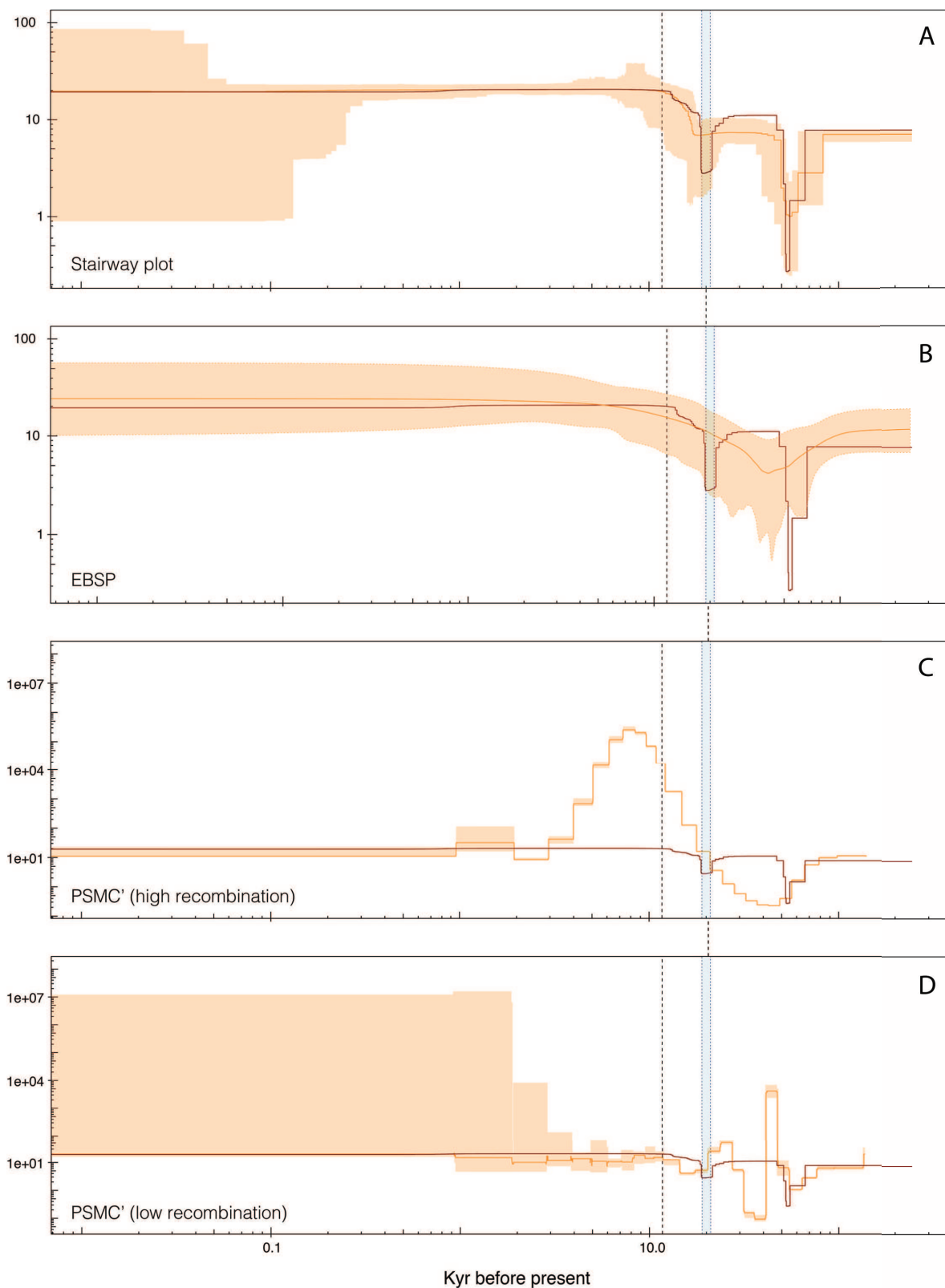
10 (i) The Stairway plot retrieves the principal events in the simulation (Fig. 28A). The main differ-
11 ence lies in the attenuation of the LGM bottleneck, that is mainly visible in the shape of the
12 95%CI. This is of importance, since it indicates that the Stairway approach may underestimate,
13 rather than overestimate, the bottleneck signal in the data: thus, the bottleneck inferred from the
14 empirical data is likely to be at least as deep as reconstructed. The demographic peak that is visi-
15 ble in the 95%CI at the beginning of the Holocene in our reconstruction from the empirical data
16 (Fig. 19A), on the other hand, although not simulated, is also present in the simulation's 95%CI.
17 Thus, that secondary peak rather appears to be entirely artefactual.

18 (ii) EBSP on simulated data globally matches the expected demographic history (Fig. 28B), with
19 the true demography nearly entirely included in the EBSP CI95% interval. However, the double
20 bottleneck in our simulated data is smoother out as one single depression in the reconstruction,
21 that matches neither bottleneck, but rather averages them - although additional complexity is vis-
22 ible in the shape of the lower CI95% interval. When comparing the empirical-data EBSP (Fig.
23 26), and the simulated reconstructions, CI95% overlap entirely although median effective popu-
24 lation size differs, and uncertainty is much larger in the empirical EBSP. Interestingly, however,
25 the empirical run exhibits some features of our simulated model that the simulated-data run fails

1 to retrieve - in particular the low population size during the Llanquihue glacial episode. Due to
2 the low number of SNPs in the loci we include in EBSP analysis, however, less resolution is ex-
3 pected for ancient time periods, so neither the observed discrepancy between simulated and em-
4 pirical runs, nor the loss of precision compared to the simulated scenario, is surprising.

5 *(iii)* PSMC' reconstruction, on the other hand, exhibits a more unexpected behaviour when ap-
6 plied to our data. When assuming equal substitution and recombination rates, none of the bot-
7 tlenecks is retrieved, but one single bottleneck is inferred instead around 40 Kyr BP, while a large
8 population size peak (absent from our simulation) is inferred in the early holocene (Fig. 28C).
9 Decreasing the recombination rate down to 1/16th of the substitution rate allows us to recover
10 both bottlenecks, yet the artefactual additional population depression remains around 40 Kyr BP,
11 as well as a sharp artefactual population peak after the most ancient bottleneck (Fig. 28D). None
12 of the reconstructions performed on simulated data matches the true demography in a satisfacto-
13 ry way: however, the very recent events on which we focus may be at the limit of the PSMC'
14 method (Li & Durbin 2011). It is noteworthy, however, that the empirical PSMC' inference fol-
15 lows the expected general demographic trend as given by both the Stairway plot analysis and the
16 EBSP analysis, smoothing out both bottlenecks in one single population increase from the early
17 Pleistocene to the late Holocene.

The King synnome - §98



1 **Figure 28 | Validation of the demographic reconstructions through simulation.** Median effective
 2 population size and confidence interval (in thousands of breeders) as a function of time, as reconstructed from simu-
 3 lated data (simulated scenario is represented by the red line on each graph). A) Stairway plot reconstruction, B)

The King synnome - §99

1 EBSP reconstruction, C-D) PSMC' reconstruction, with either high (C) or low (D) recombination rate. Dashed
2 black line: Pleistocene-Holocene boundary. Shaded blue band: last glacial maximum.

3 *S-2: Field observation data.*

4 *§-99 Capture-mark-recapture experiments.* In order to verify the hypothesis of high dispersal be-
5 tween colonies, we deployed a capture mark recapture (CMR) experiment on Possession Island,
6 Crozet archipelago, and Ratmanoff beach, Kerguelen archipelago. Ca. 9,832 king penguins were
7 equipped with passive radio-frequency identification (RFID) tags since 1990 on the BDM
8 colony, in Crozet archipelago within the framework of a long-term monitoring program (see
9 Gendner *et al.* 2005 for details). We deployed mobile detection antennas on all other colonies of
10 Possession Island, as well as on Ratmanoff beach. These antennas have a low detection distance
11 (ca. 60 cm), and are buried in the ground on paths frequented by penguins when they travel in
12 and out of the colony. Each antenna is ~5 m wide, and records the identification number of any
13 RFID-tagged individual crossing the detection zone. On average, each antenna works for ~12
14 hours. Antennas were deployed in the evening, in order to record the activity peak around sun-
15 rise. In the current state of development of this system, it is impossible to assess how many indi-
16 viduals (tagged or not) crossed the detection zone during the deployment period: thus, recaptures
17 can only be analysed as presence-absence data, and not as quantitative CMR results. Due to the
18 harsh field conditions, deployments were also in some measure opportunistic; and it is generally
19 impossible to ascertain the status of detected individuals (breeding or moulting), except when
20 their age excluded a breeding attempt. This data, however, provides us precious insights into the
21 behavioural mobility of the species, since antennas were usually located well within the target
22 colonies, and not directly at the seaside: thus, only penguins wandering into a colony distinct
23 from their birth colony were detected.

The King synnome - §100

1 We performed a total of 28 12-hours deployments during the field seasons 2011-2012,
2 2012-2013, 2013-2014, and 2014-2015 (Per-colony detections/deployments: CDC: 11/5,
3 PMC: 44/10, JPN: 9/5, MAE:1/1). Out of the 9,832 individuals marked as chicks, an average of
4 2.3 birds per 12-hour antenna deployment were detected on other colonies of the same island.
5 One anecdotic recapture, in 2014, of a tagged individual born on Crozet in 2009 (and therefore
6 reaching age of first breeding at the time of recapture) also happened on the Ratmanoff beach
7 colony, on Kerguelen archipelago. Although a single event has hardly any statistical value, the
8 Ratmanoff colony counts ~140,000 breeding pairs (Delord *et al.* 2015), and only two 5m-long
9 antennas were deployed along the beach: thus, this single recapture suggests that dispersal from
10 Crozet to Kerguelen may not be a rare event.

11 *S-3: Palaeoclimate of the Southern Ocean.*

12 *§-100 Definition and constraints of the Antarctic Polar Front.* The Southern Ocean is characterised
13 by a strong circular Westerly current that flows uninterrupted by land barriers, the Antarctic
14 Circumpolar Current (ACC). Strong westerly winds generate important northward Ekman trans-
15 port in the surface water layer, resulting in a convergence of the cold Antarctic surface waters,
16 and warmer Subantarctic surface waters, where the colder southern water masses sink below the
17 northern water mass, at the Antarctic Polar Front (APF). This convergence is compensated by a
18 divergence area, where upwelled deep water masses rise to the surface, creating an intense marine
19 productivity area (Mackintosh 1946; Peterson & Whitworth 1989; Meijers 2014). This area is
20 characterised by a steep surface temperature gradient, between 5°C and 3°C (Moore *et al.* 1999).
21 Generally, a cooling of surface waters in the Southern Ocean is reflected in a northward displace-
22 ment of the APF, while a warming brings the APF southward. However, as the APF is defined by
23 the interaction of deep and surface water masses, it is strongly constrained by the sea bottom
24 topography (Moore *et al.* 1999). Important bathymetric features, such as the Campbell Plateau,

The King synnome - §101

1 the Drake passage, or the Kerguelen plateau, may constrain the position or structure of the APF.
2 In other areas, most importantly the Southern Indian Ocean and Southern Atlantic Ocean, APF
3 displacement is mostly free from bathymetric constraints, and exhibits the largest latitudinal vari-
4 ation (Kohfeld *et al.* 2013).

5 The Campbell Plateau may be the best studied case of bathymetric constraint on the APF. Both
6 flow models and sediment core evidence showed that the APF remained south of the plateau
7 throughout the Pleistocene, despite important changes in sea surface temperature and frontal po-
8 sitions throughout the Southern Ocean. Whereas the APF is free to move south to greater de-
9 pths, it is constrained to the North by the sea floor rise (Hayward *et al.* 2008; MARGO 2009;
10 Kohfeld *et al.* 2013). Similarly, the Drake passage constrains both the northern and southern
11 boundaries of the APF (Moore *et al.* 1999; Dong *et al.* 2006; Chereskin *et al.* 2012). Finally, the
12 Kerguelen plateau has been shown to alter the subsurface structure of the front, with its deeper
13 manifestations moving North of the islands, while the surface expressions move South (Moore *et*
14 *al.* 1999; Park *et al.* 2014). These features, however, are now well modelled in the CMIP5 panel,
15 which has a much improved bathymetric resolution (Meijers 2014), and the influence of the
16 Drake Passage and Campbell Plateau on the frontal structure is accurately reproduced in our re-
17 constructions (Fig. 20).

18 *§-101 Current state of knowledge.* There are still considerable uncertainties as to the Pleistocene
19 and Holocene history of the Southern Ocean. Available evidence relies on different types of prox-
20 ies (Gersonde *et al.* 2005; Thomas & Dieckmann 2008; Kohfeld *et al.* 2013; Hodgson *et al.*
21 2014). (a) Ice core data (e.g. EPICA Dome C and Vostok) provide direct evidence for chemical
22 conditions at the core site, and indirect evidence for the oceanic source areas, provided transfer
23 models are accurate enough (Wolff *et al.* 2003; Wolff *et al.* 2006). Parameters derived from ice
24 core evidence mostly covers air temperature, sea ice extent, and marine productivity (Wolff *et al.*
25 2006) (b) Benthic sediment core provide more direct evidence for marine conditions (tempera-

1 ture, sea ice cover, productivity) at the core location (Gersonde *et al.* 2005; Martínez Garcia *et al.*
2 2009). (c) Peat cores and geological evidence on the subantarctic islands and surrounding conti-
3 nental shelf are mostly informative for land ice cover (Hodgson *et al.* 2014). Taken together, this
4 evidence allows for a general palæoclimatic reconstruction in the Southern Ocean. However,
5 there is still much progress to be done in reconciling the different sources of evidence, as variabil-
6 ity amongst core locations (especially benthic sediment cores) is high, and several land-sea cou-
7 pling mechanisms are still poorly understood (McGlone *et al.* 2010). In the current state of
8 knowledge, we can distinguish four major periods in the Southern Ocean late-Pleistocene and
9 Holocene history: (i) Quaternary conditions (59-22 Kyr BP), (ii) Last Glacial Maximum condi-
10 tions (21-18 Kyr BP), (iii) Pleistocene glacial retreat and early holocene optimum (17-9 Kyr BP),
11 (iv) Holocene hypsithermal and neoglacial conditions (8-0 Kyr BP).

12 *(i) Quaternary conditions (59-22 Kyr BP)* were mostly glacial-like, with slow onset of glaciation
13 from ~35 Kyr BP, and winter sea ice cover reaching as far as ~56°S in the Pacific. Little is known
14 of land ice throughout the period, as further glaciation obliterated most of the direct evidence.

15 *(ii) Last Glacial Maximum conditions (21-18 Kyr BP)* were characterised by extensive land and
16 sea ice cover throughout the Southern Ocean. **The Antarctic Polar Front** is thought to have
17 moved northward to 40-50°S, a movement associated with a ~5°C cooling in summer SST (Ger-
18 sonde *et al.* 2005; MARGO 2009), although frontal movement is thought to have been con-
19 strained by bathymetry south of the Campbell plateau (Neil *et al.* 2004; Kohfeld *et al.* 2013).
20 **Winter sea ice** is also thought to have reached ~50°S or further northward, or the approximate
21 position of the present-day polar front (CLIMAP 1981) (between 47°S in the Atlantic and Indi-
22 an Oceans, and 57°S in the Pacific Ocean, see Gersonde *et al.* 2005). **Marine productivity** is
23 thought to have shifted from the Antarctic to the Subantarctic region (Hodgson *et al.* 2014),
24 while not changing significantly in total biomass (Wolff *et al.* 2006; Kohfeld *et al.* 2013). **Islands**
25 **of Heard, Crozet, Marion, and the Drake Arc were entirely covered by ice**, while Kerguelen and

The King synnome - §101

1 South Georgia may have had ice-free areas. Falklands and Macquarie underwent periglacial con-
2 ditions (Hodgson *et al.* 2014). Likely faunal refugia were the currently subtropical islands of
3 Gough, Auckland and Campbell, as well as the Falklands and more generally the Patagonian shelf
4 area (Fraser *et al.* 2012).

5 *(iii) Pleistocene glacial retreat and early holocene optimum (17-9 Kyr BP)* saw a gradual thawing
6 of most land ice, with contrasting chronologies. **Antarctic and subantarctic fronts** retreated south
7 to their current location (Kohfeld *et al.* 2013). **Sea ice** retreated until the Early Holocene climatic
8 optimum (~11.5-9 Kyr BP), with an episodic increase during the Antarctic Cold Reversal around
9 14.5 Kyr BP, reaching its current position by ~10 Kyr BP. Kerguelen and South Georgia archipel-
10 ago bear signs of early deglaciation, while Crozet and Marion islands are thought to have carried
11 extensive land ice until the end of the period (Hodgson *et al.* 2014). The end of the period is
12 marked by a first cold reversal in the Antarctic waters and a short increase in sea ice cover, of un-
13 known extent (Nielsen *et al.* 2004).

14 *(iv) Holocene hypsithermal and neoglacial conditions (8-0 Kyr BP)* were characterised by a
15 warmer climate, similar to historical conditions, interrupted by minor cold reversals. The sub-
16 antarctic region is ice-free, and the northernmost islands of Gough, Auckland and Campbell are
17 located north of the Subantarctic front (Hodgson *et al.* 2014). Temperature reaches a maximum
18 around ~7.5 Kyr BP in the South Pacific (Calvo *et al.* 2007). Marine conditions are warm and ice
19 free at ~50°S until around 6-5 ka BP (Hodell *et al.* 2001; Nielsen *et al.* 2004), and temperature
20 drops slightly after ~3 Kyr BP, although with no change in the glacial landscape. **Neoglacial con-**
21 **ditions** arise after 5 Kyr in East Antarctica, and 3 Kyr in West Antarctica: open water conditions
22 are still prevalent throughout the Southern Ocean, although with possible winter sea ice episodes
23 at 53°S at some periods (~1-2 Kyr BP).

24 *S-4: Atmosphere-Ocean General Circulation Models (AOGCMs)*

1 *§-102 AOGCMs choice and multi-model ensemble approach.* We used the latest generation of
 2 AOGCMs from the IPCC Coupled Model Intercomparison Project Phase 5 (CMIP5, see Taylor
 3 *et al.* 2012), which represent a significant improvement over CMIP3 in the Southern Ocean
 4 (Meijers 2014). We applied a multi-model ensemble approach, a common improvement over sin-
 5 gle-model projections, as only the trends present in most models are retained in the final ensem-
 6 ble mean (Meijers 2014). We selected 15 AOGCMs based on the range of available outputs and
 7 their coverage of the Southern Ocean (see Table 3). All model outputs were downloaded from the
 8 ESGF nodes (pcmdi9.llnl.gov/). In our study, we used the following variables: Sea Surface Tem-
 9 perature (SST) and Sea-Ice Concentration (SIC). For each variable, we calculated the multi-mod-
 10 el ensemble mean and standard deviation using the Climate Data Operators toolset (CDO 2015,
 11 available at: <http://www.mpimet.mpg.de/cdo>).

12 Reconstructions were performed under Last Glacial Maximum, mid-Holocene, and Historical
 13 conditions, and projections according to three Representative Concentration Pathways (rcp)
 14 scenarios, the rcp2.6, rcp4.5, and rcp8.5, corresponding respectively to the strong emissions
 15 reduction scenario, a moderate emissions profile and the “business-as-usual” scenario. We exclud-
 16 ed the rcp6.0 as too few model outputs are available yet.

17 **Table 3 | Ensemble members used in habitat predictions.** Model outputs were downloaded from the
 18 IPCC archive (http://www.ipcc-data.org/sim/gcm_monthly/AR5/Reference-Archive.html). Only one ensemble
 19 member was used for each model (r1i1p1 whenever available). Not all models outputs for both 21st century and pa-
 20 leoclimate experiments, thus different ensembles were used for LGM, mid-Holocene and 21st century
 21 reconstructions.

Model	Institution	LGM	mid-Holocene	21st century
BCC-CSM1	<i>Beijing Climate Center (China)</i>	N/A	r1i1p1	r1i1p1
CanESM1	<i>Canadian Centre for Climate Modelling and Analysis (Canada)</i>	N/A	N/A	r1i1p1
CCSM4	<i>National Center for Atmospheric Research (USA)</i>	r1i1p1	r1i1p1	r1i1p1
CESM1-CAM5	<i>National Center for Atmospheric Research (USA)</i>	N/A	N/A	r1i1p1

The King synnome - §103

1	CNRM-CM5	<i>Centre National de Recherches Météorologiques, Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (France)</i>	r1i1p1	r1i1p1	r1i1p1
2	CSIRO-Mk3-6-0	<i>Australian Commonwealth Scientific and Industrial Research Organiza- tion (Australia)</i>	N/A	r1i1p1	N/A
3	EC-EARTH	<i>EC-EARTH consortium published at Irish Centre for High-End Computing (Netherlands/Ireland)</i>	N/A	N/A	r8i1p1
4	FIO-ESM	<i>The First Institute of Oceanography, SOA (China)</i>	N/A	N/A	r1i1p1
5	FGOALS-G2	<i>Institute of Atmospheric Physics, Chinese Academy of Sciences, and Ts- inghua University (China)</i>	r1i1p1	r1i1p1	N/A
6	GFDL-ESM2M	<i>Geophysical Fluid Dynamics Laboratory (USA)</i>	N/A	N/A	r1i1p1
7	GISS-E2-R	<i>NASA/GISS (Goddard Institute for Space Studies) (USA)</i>	r1i1p1	r1i1p1	r1i1p2
8	HadGEM2-ES	<i>Met Office Hadley Centre (UK)</i>	N/A	r1i1p1	r1i1p1
9	IPSL-CM5A-MR	<i>Institut Pierre Simon Laplace (France)</i>	r1i1p1	r1i1p1	r1i1p1
10	MIROC5	<i>Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (Japan)</i>	N/A	N/A	r1i1p1
11	MPI-ESM-MR	<i>Max Planck Institute for Meteorology (Germany)</i>	N/A	N/A	r1i1p1
12	MPI-ESM-P	<i>Max Planck Institute for Meteorology (Germany)</i>	r1i1p1	r1i1p1	N/A
13	MRI-CGCM3	<i>Meteorological Research Institute (Japan)</i>	r1i1p1	r1i1p1	r1i1p1
14	NorESM1-M	<i>Bjerknes Centre for Climate Research, Norwegian Meteorological Institute (Norway)</i>	N/A	N/A	r1i1p1

15 §-103 *Sea Surface Temperature (SST)*. For palæoclimate as well as 21st century projections, we fol-
16 lowed a protocol similar to that of Péron *et al.* 2012. The 5°C Sea Surface Temperature (SST)
17 isotherm was used as a diagnostic of the position of the Antarctic polar front (APF, see Moore *et*
18 *al.* 1999) where the King penguin is known to forage (Péron *et al.* 2012). The particular breeding
19 cycle of the King penguin makes the constraints on foraging behaviour especially strong during
20 the early chick rearing stage, when the juveniles have not yet reached thermal independence, and
21 need regular feeding while not being able to survive without an adult (Barrat 1976), which hap-
22 pens around the month of February. This is supported by observed foraging trips, which show a
23 much greater geographic constraint during the month of February (Péron *et al.* 2012; Bost *et al.*
24 2015). Thus, we focused our analysis to the position of the APF in February, as representative of
25 the maximum constraint on foraging trips.

The King synnome - §103

1 Before using SST outputs derived from AOGCMs, we assessed the accuracy of the representation
 2 of the Southern Ocean by comparing each model SST-output for historical runs to satellite-
 3 measured SST from december 1981 to december 2005, using the NOAA Optimal Interpolation
 4 v2 SST dataset (Reynolds *et al.* 2002). Cell-by-cell ($1^\circ \times 1^\circ$) linear correlation of SST was assessed
 5 and R^2 , slope and intercept were plotted in order to assess the spatial distribution of model de-
 6 parture from observed values.

7 As modelled SST was generally found warmer than observed SST in the APF zone over the his-
 8 torical period, we followed the correction applied by Péron *et al.* 2012. In order to maximise the
 9 fit between observed and modelled SST for each archipelago, we defined four oceanic sectors:
 10 South Atlantic Ocean (45°W to 18°E), South Indian Ocean (18°E to 80°E), Macquarie (135°E
 11 to 180°E), and Falkland region (75°W to 45°W), ranging in latitude from 45°S to 55°S , but ex-
 12 tended to 60°S in the Falkland region to account for the higher latitude of the APF around Cape
 13 Horn. For each of these sectors, we tested the linear correlation between modelled and observed
 14 SST, and we corrected the model value linearly when needed (Table 4). The 5°C SST isotherm
 15 was then calculated in GDAL (www.gdal.org), and kilometric distance between each island and
 16 the 5°C isotherm was calculated using the OGR Python library. Correctness of our model was as-
 17 sessed through 1) correlation the observed and modelled distances to the 5°C isotherm on the
 18 1981-2005 period and 2) consistency between these distances and published data on King pen-
 19 guin foraging areas.

20 **Table 4 | Correlation of observed and modelled SST in the Southern Ocean.** Slope, intercept and
 21 correlation coefficient for linear correlation of our ensemble model and observed SST data over the historical period
 22 (1981-2005), in four sectors of the Southern Ocean.

Sector	R^2	Slope	Intercept
Drake	0.903 ± 0.0291	0.683 ± 0.0472	1.84 ± 1.19
South Atlantic	0.945 ± 0.0259	0.785 ± 0.0571	1.84 ± 1.04
South Indian	0.945 ± 0.0157	0.791 ± 0.0730	2.27 ± 0.876

The King synnome - §104

1	Macquarie	0.945 ± 0.0160	0.805 ± 0.0687	1.80 ± 0.705
---	-----------	--------------------	--------------------	------------------

2 Foraging range predictions for the historical period closely matched both observed historical SST,
3 and observed foraging distances at most locations: ~380 km on Crozet (observed: 300-500 km,
4 see Péron *et al.* 2012), ~320 km on Marion (observed: 300 km three decades ago, see Adams &
5 Klages 1987), ~20 km in the Kerguelen (observed: 270 km in the APF along the 4 and 5°C
6 isotherms, see Koudil *et al.* 2000; Pütz 2002 - the APF is reached immediately, but foraging trips
7 extend further in the productivity zone), ~310 km on Heard (observed: 370 km a decade ago, see
8 Moore *et al.* 1998), ~300 km in South Georgia (observed: 300-600 km over the whole breeding
9 season, see Pütz 2002). Predicted distance for Macquarie Island (~240 km) is slightly lower than
10 the observed summer range (300-500 km, see Wienecke & Robertson 2002), however, recorded
11 foraging trajectories meet the APF in the higher-productivity areas on the edge of the Campbell
12 plateau, where upwelling is increased, rather than southward along the shortest route. Finally, the
13 predicted and observed ranges differ most strongly in the Falklands (predicted: ~640 km and ob-
14 served: 300-500 km, see Pütz & Cherel 2005), a discrepancy explained by the fact the small Falk-
15 land population frequently forages on the Patagonian Shelf break, and not directly on the APF
16 (Pütz & Cherel 2005). This different behaviour of the Falkland population makes its response to
17 APF displacement more uncertain, as other productivity areas may remain available. However, it
18 seems that the Patagonian Shelf could never sustain a large King penguin population (Pistorius *et*
19 *al.* 2012), and it is sustaining a high, and increasing, anthropogenic pressure from overfishing
20 and climate change (Halpern *et al.* 2008). It is therefore unlikely that the Falklands may sustain a
21 significant King penguin population on a centennial time scale.

22 *§-104 Winter sea-ice concentration (SIC).* Winter Sea-Ice Concentration (SIC) is known to limit
23 the southward expansion of the King penguin's breeding range, as the species overwinter breed-
24 ing cycle makes open-water conditions a requisite throughout the year (Barrat 1976). Although

The King synnome - §105

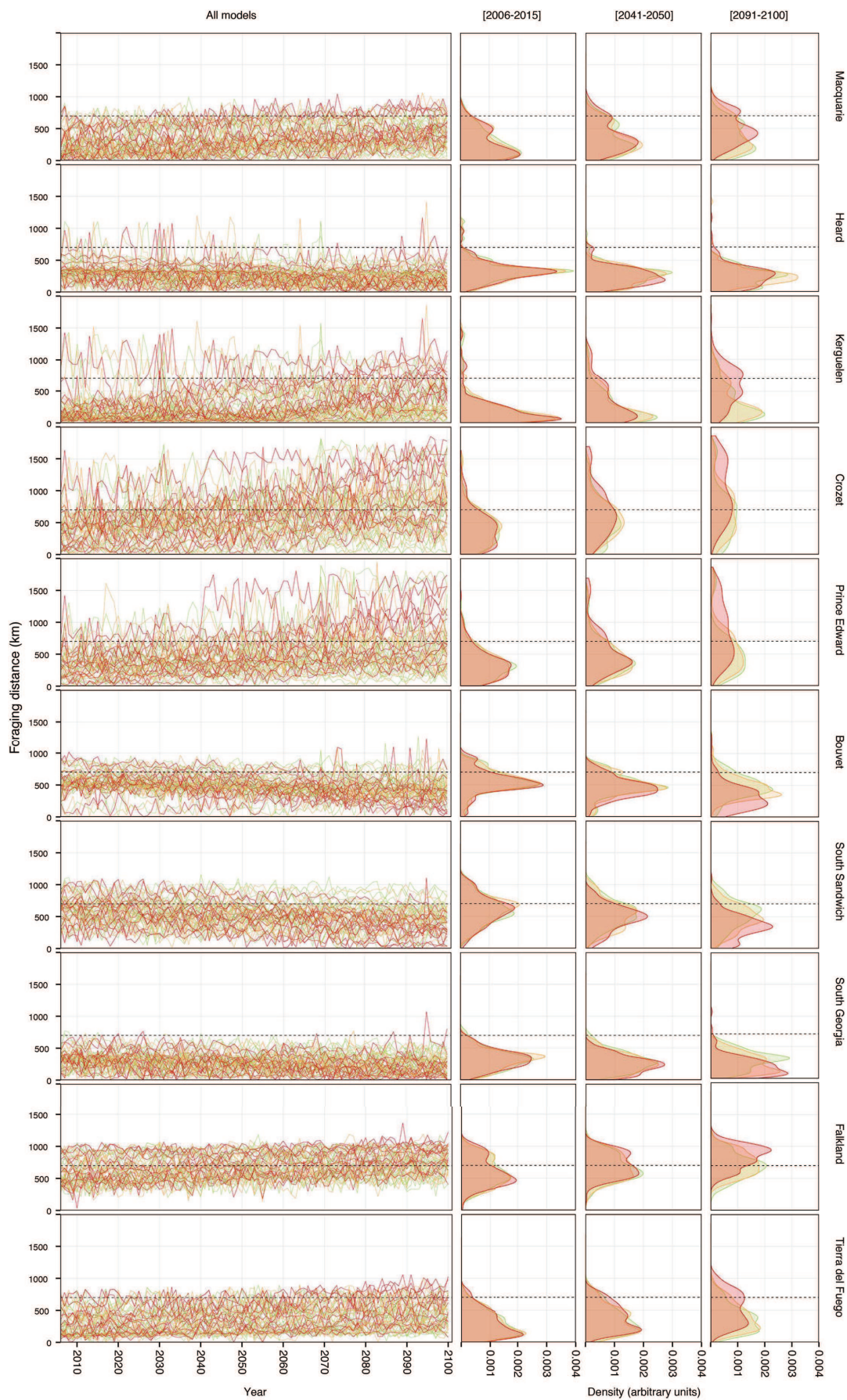
1 SIC may still be subject to biases in its representation compared to SST, it has improved since
2 CMIP3 (Turner *et al.* 2013; Xu *et al.* 2013; Shu *et al.* 2015). We take the 15% concentration iso-
3 line as being representative of the effective sea ice edge (Turner *et al.* 2013). We only consider the
4 sea-ice concentration at their maximum, during the months of august and September. Compared
5 to satellite-derived historical measures from the NOAA Optimal Interpolation dataset, ensemble
6 reconstruction gives a winter sea ice that tends to be more dense than observed values (mean den-
7 sity of sea ice above 15% concentration over the 1981-2005 period: reconstructed 85 ± 20 %;
8 observed 61 ± 22 %, t-test p-value $< 2.2e-16$), but less extended (reconstructed extent of septem-
9 ber SIC $> 15\%$ on the 1981-2005 period occupies 90% of observed SIC $> 15\%$ extent), al-
10 though correlation is strong on a per-cell basis (mean $R^2 = 0.67 \pm 0.27$). Winter sea ice extent is
11 projected to decrease in all forcing scenarios. While sea ice cover should still be relatively impor-
12 tant even at the northern tip of the South Sandwich islands during the last two decades of the
13 century (rcp2.6: 0.26 ± 0.058 , rcp4.5: 0.22 ± 0.044 , rcp8.5: 0.045 ± 0.040), Bouvet island is pro-
14 jected to become ice-free all year round by 2080 under all forcing scenarios (rcp2.6: 0.058
15 ± 0.037 , rcp4.5: 0.028 ± 0.024 , rcp8.5: 0.00041 ± 0.00053). However, sea ice projections may not
16 be quite as reliable as SST projections. Indeed, although geographical distribution is modelled
17 rather accurately, CMIP5 ensemble models fail to reproduce the increase in sea ice extent ob-
18 served in East Antarctica over the last decades, suggesting that some processes are not yet ade-
19 quately accounted for in the current models (Turner *et al.* 2013) - in particular, the impact of the
20 influx of fresh meltwater from the Antarctic ice sheet on the extent of winter sea ice may still be
21 widely underestimated. If such a bias exists, however, it underestimates the true extent of sea ice:
22 in that case, the King penguin's range reduction may be even more drastic than we forecast here,
23 as Bouvet may not be ice-free and suitable for colony establishment by the end of the century.

24 *§-105 Uncertainties assessment.* Although the use of a multi-model ensemble mean approach is
25 considered to outperform the use of a single climate model, it is also essential to assess the uncer-
26 tainties related to AOGCMs to evaluate the confidence that can be attached to our results. Out-

1 puts of the different AOGCMs may diverge across time and space because they are based on di-
2 verse parameterization of natural processes, downscaling approaches, spatial resolutions, etc... To
3 assess the uncertainties associated with our projections, we calculated, for each rcp scenario, the
4 projected foraging distance derived from each climate model separately. We followed the protocol
5 developed by Goberville *et al.* 2015 by calculating the density distribution of projected foraging
6 distance for each island (*i*) for the current period (2006-2015), (*ii*) for the middle of the century
7 (2041-2050) and (*iii*) for the end of the century (2091-2100) (Fig. 29). In addition, for the same
8 periods, we also calculated the percentage of models forecasting local King penguin population
9 collapse (February foraging distance > 700 km; Fig. 30), as proposed by Raybaud *et al.* 2013. The
10 latitude of the APF, and therefore the duration of the King penguin's foraging trips, is subject to a
11 high interannual variability, in particular under the influence of the cyclical El Niño Southern
12 Oscillation and Southern Annular Mode, with year-to-year latitudinal fluctuations of up to 200
13 km (Bost *et al.* 2015). Therefore, we considered that a location had reached its critical foraging
14 distance when foraging distance was higher than 700 km for at least 20% of a consecutive
15 decade.

16 **Figure 29 | (NEXT PAGE) Foraging distance from single models.** Projected distance between sub-
17 antarctic archipelagos and the Antarctic Polar Front in February extracted from IPCC CMIP5 models taken sepa-
18 rately. Colours correspond to rcp scenarios: rcp-2.6 (green), rcp-4.5 (orange) and rcp-8.5 (red). Dashed line repre-
19 sents the 700-km limit. **A)** yearly projection, **B)** density distribution per rcp scenario, at three different time steps.

The King synnome - §105



1 Variability between models remains relatively high, as has already been observed in previous stud-
2 ies (Cabr e *et al.* 2015; Goberville *et al.* 2015 - see Fig. 30). At all locations, predictions overlap
3 entirely between rcp scenarios for the first decade of our projections, as is expected. This is still
4 mostly the case in the middle of the century (2041-2050). Most of the divergence between
5 scenarios appear by the end of the century. This may take the form of (i) a strong divergence of
6 the rcp-8.5 projections as opposed to rcp-4.5 and rcp-2.6 (as in Kerguelen and Bouvet); (ii) an
7 increased dispersion on rcp-8.5 projections (as in Crozet and Prince Edward), or (iii) a more
8 gradual panel of possible outcomes from rcp-2.6 to rcp-8.5 (at most other locations), or (iv) no
9 strong difference between scenarios in Heard Island. This contrast between scenarios is also no-
10 ticeable when considering the proportion of individual models predicting a local extinction at
11 each time period (Fig. 30). In the last decade of the century, the ‘business-as-usual’ rcp-8.5
12 scenario stands out compared to the ‘controlled-emissions’ rcp-2.6 and rcp-4.5 in Kerguelen,
13 Crozet, Prince Edward, Bouvet and South Georgia, while all three scenarios make up a gradient
14 in Macquarie, South Sandwich, the Falklands, and Tierra del Fuego. Under rcp-8.5, more than
15 50% of the models predict extinction in Crozet, Prince Edward and the Falklands by the end of
16 the century, and the difficult position of Kerguelen and Tierra del Fuego is confirmed by the fact
17 that a large proportion of models predict extinction on these islands too. Overall, although inter-
18 model variability remains high, and alternative outcomes are possible, the strong consensus both
19 in the increasing foraging distance trend, and in the actual prediction for local extinction, stress
20 both the very likely nature of the threats upon the Southern Ocean ecosystems under the rcp-8.5
21 scenario, and the possibility of yet avoiding the most destructive effects of these threats if imme-
22 diate action allows us to bring greenhouse-gas emissions closer to the rcp-2.6 forcing scenario.

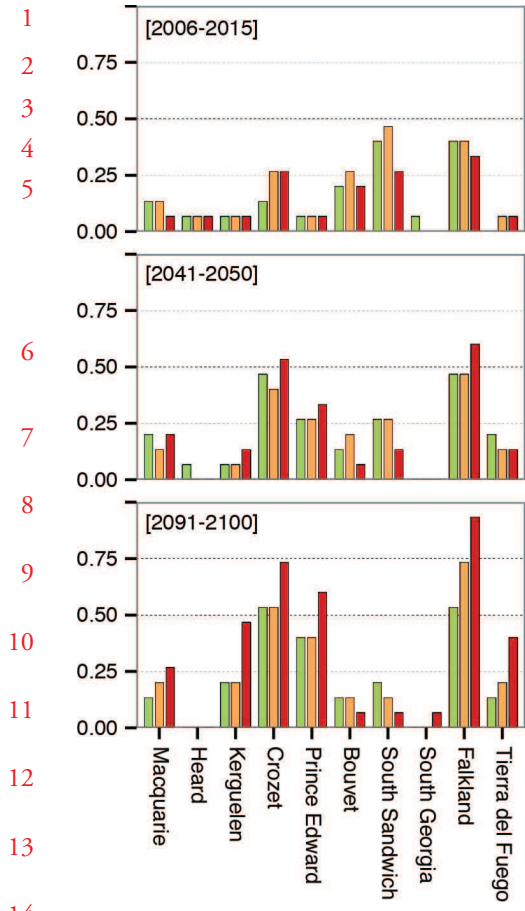


Figure 30 | Proportion of models predicting extinction of King penguin colonies. Proportion of the 15 selected models predicting a February foraging distance > 700 km for 20% of the decade, at three different time points. Colours correspond to rcp scenarios: rcp-2.6 (green), rcp-4.5 (orange) and rcp-8.5 (red).

Acknowledgements

This work was conducted within the framework of the Programme 137 of the Institut Polaire Français Paul-Emile Victor (IPEV), with additional support from the French National Research Agency (ANR) “PICASO” grant (ANR-2010-BLAN-1728-01), from Marie Curie Intra European Fellowships (FP7-PEOPLE-IEF-2008, European Commission; project no. 235962 to CLB and FP7-PEOPLE-IEF-2010, European Commission; project no.

252252 to ET), from the Centre Scientifique de Monaco through budget allocated to the Laboratoire International Associé 647 ‘BioSensib’ (CSM/CNRS-University of Strasbourg, CLB), South African National Antarctic Programme (PP) and the IPEV Programme 109 (YC). Logistic and field costs of research were supported by the IPEV Programme 137 (CLB), the South African Department of Environmental Affairs and National Research Foundation (PP). This work was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo. We are very grateful to Morten Skage, Ave Tooming-Klunderud, Marianne Selander-Hansen, and the Norwegian Sequencing Center for their very valuable help in the laboratory, as well as Lex Nederbragt and Michael Matschiner for their assistance with the Abel cluster, and Matteo Fumagalli and Thorfinn Korneliussen for their precious advice regarding ngsTools and ANGSD. Genomic data used in analyses are available the

1 Sequence Read Archive. We acknowledge the World Climate Research Programme's Working
2 Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate model-
3 ing groups (listed in Table S1 of this paper) for producing and making available their model out-
4 put. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and In-
5 tercomparison provides coordinating support and led development of software infrastructure in
6 partnership with the Global Organization for Earth System Science Portals.

1 Chapter 5: The Emperor synnome

2 Cristofari R, Bertorelle G, Ancel A, Benazzo A, Le Maho Y, Ponganis PJ, Stenseth NC, Trathan
3 PT, Whittington JD, Zanetti E, Zitterbart DP, Le Bohec C & Trucchi E, Full circumpolar mi-
4 gration ensures evolutionary unity in the Emperor penguin. (*Nature Communications - in prep*)

5 *Context*

6 *§-106 Can a species be studied from a single colony?* Most studies investigating the effects of climate
7 change on population dynamics have relied either on «anonymous» count data (*i.e.* unmarked
8 individuals treated as a group - see for example VanDerWal *et al.* 2013 or Langham *et al.* 2015),
9 or capture-mark-recapture, individual-based data (*e.g.* Le Bohec *et al.* 2008, Jenouvrier *et al.*
10 2014, or Durner *et al.* 2009). The latter approach is considerably more precise, as it is able to
11 identify some misleading compensatory effects (such as local decline masked by density-depen-
12 dent immigration) - however, fieldwork constraints typically restrict this approach to a single
13 study colony or population (as is the case in Charmantier *et al.* 2008; Le Bohec *et al.* 2008; Je-
14 nouvrier *et al.* 2014) or to a handful of populations in optimal cases (*e.g.* Dunn *et al.* 2011). Al-
15 though oftentimes highly informative, these studies all run the risk of misestimating local idio-
16 syncrasies. For example, the characteristic plastic response of Great tits to food peak advance,
17 observed in Great Britain by Charmantier *et al.* 2008, is in strong contrast with observations, al-

The Emperor synnome - §107

1 be it in very similar conditions, made in Netherlands by Visser *et al.* 1998. Similarly, the recent
2 study of Stenseth *et al.* 2015 found important differences in the response to climate change of
3 different neighbouring Belgian and Dutch Blue and Great tit (*Pares caeruleus* and *P. major*) popu-
4 lations. From a functional standpoint, the observation of Weimerskirch 2013 of fully segregated
5 foraging grounds in two nearby booby colonies suggests that different alimentary strategies may
6 lead to different exposure to environmental change. Thus, before extrapolating the results of local
7 observations to species-level predictions, it is necessary to assess the nature and extent of the cou-
8 pling between local and global processes.

9 *§-107 Linking generation-scale and coalescent-scale demography.* Amongst the hypotheses that allow
10 the extensions of local demographic models to global trends is the assumption of *demographic in-*
11 *dependence*, that Jenouvrier *et al.* 2014 proposes, in the case of the Emperor penguin, as the idea
12 that « *inter-colony dispersal is extremely unlikely [...]. Thus, neither dispersal nor demographic source-*
13 *sink dynamics will change our main conclusion about global population declines* » - census and cap-
14 ture-mark-recapture data is thus considered self-contained, and the authors « *do not consider*
15 *movement among populations; it is unlikely to be demographically important in this species* ». Al-
16 though hardly an avoidable assumption in that case, since almost no demographic data is avail-
17 able for other colonies, this postulate presents a risk: it does not acknowledge the possibility that
18 (i) the population trend in the Pointe Géologie Emperor penguin colony may be particular to
19 that colony for stochastic, or for idiosyncratic reasons (that is, not directly because of the ob-
20 served local sea ice conditions - some elements even make this a very likely possibility, see §24 p.
21 79), and (ii) part of the observed population processes may be better explained by migration than
22 by mortality or growth - that is, without assuming a direct link between local colony size and
23 global population size. Measuring the importance of dispersal in field conditions is a considerable
24 challenge (see §19 p. 68), and can only be done with massive recapture effort (see *e.g.* Barlow *et*
25 *al.* 2013). Population genetics approaches, on the other hand, allow us to access the long-term
26 migration parameter for a set of populations (see §41 p. 103), and to set, if not a point estimate

1 of instantaneous dispersal rate, at least a bracket for averaged migration ability in the species, that
2 can help us define reasonable assumptions for developing more accurate demographic models.

3 *Abstract*

4 The rate of ongoing environmental change is now thought to exceed the rate at which most
5 species (Hoffmann & Sgrò 2011), including humans (Díaz *et al.* 2006), are able to adapt, with
6 significant consequences for their resilience and for ecosystem sustainability. In a recent re-
7 view, Chown *et al.* 2015 reported that despite the pristine appearance of Antarctica, its species
8 and ecosystems are also under considerable threat. Investigating trends and risks in such a re-
9 mote area relies mostly on a handful of bio-indicator species: the Emperor penguin (*Apten-*
10 *odytes forsteri*) often fulfills this task, as it integrates the multiple effects of climate change
11 (Ainley *et al.* 2010a). A recent study suggests that populations of this flagship Antarctic
12 species may be at high risk continent-wide within the next 100 years (Jenouvrier *et al.* 2014).
13 However, predicting the dynamics of species adaptation and persistence requires a more com-
14 plete understanding of the temporally and spatially complex ecological processes shaping the
15 structure of worldwide populations, and can greatly benefit from the synergy amongst discip-
16 lines and methods. The projected decline of the Emperor penguin population (Jenouvrier *et al.*
17 2014) relies on the explicit assumption, based on behavioural observations, that inter-colony
18 dispersal is negligible. Yet, using genome-wide data from individuals belonging to several
19 colonies encompassing the whole Antarctic continent, we reveal that Emperor penguin popu-
20 lations are characterised by a high level of gene flow, with migration rates reaching up to
21 4.2% between colonies separated by more than 8,000 km of coastline, and a shared common

1 demography over the last quaternary climate change period. Thus, the population structure in
2 this polar top-predator is better explained by one single global population rather than by a
3 fragmented colonial system. By rejecting the view of the local colony as a relevant demo-
4 graphic unit, our results highlight that *i*) robust demographic projections and extinction risk
5 estimations will only be possible by including dispersal rates in the models, and *ii*) population
6 size fluctuations in single colonies should rather be taken as indicative of local stochastic
7 events, since the species' response to global environmental change will likely follow a shared
8 evolutionary trajectory.

9 *Results*

10 *§-108 Studying demography in the context of climate change.* The importance of global warming is
11 now a central subject of international concern, and an increasing number of studies seek to un-
12 derstand its impact on the world's ecosystems and to predict likely scenarios in response to cli-
13 mate projections (Thomas *et al.* 2004), either in order to set up more efficient conservation
14 strategies, or as a means to prompt urgent political action (Watson *et al.* 2013; Chown *et al.*
15 2015). One of the main difficulties of this task lies in the fact that only a handful of species have
16 been monitored for more than a few decades, and, in most cases, our knowledge of their demog-
17 raphy is limited to short-term (i.e. generation-scale) responses (Saether *et al.* 2005). Yet, in order
18 to establish reliable projections, larger-scale population parameters must be integrated into demo-
19 graphic models. Recent development of high-throughput sequencing allows the analysis of
20 genome-wide and population-scale data and provides the genomic signature of the important de-
21 mographic parameters that can be used to accurately predict species responses to global change
22 (Allendorf *et al.* 2010).

The Emperor synnome - §109

1 Extracting species-wide projections from time-series collected on a single population requires a
2 precise understanding of how local events relate to species-scale demographic processes. We need
3 for instance to establish whether the observed local extinctions or fluctuations occur as a conse-
4 quence of mortality peaks, massive dispersal events, or a combination of both. Adult mortality
5 has traditionally been proposed as the primary factor (through changes in resource availability
6 and subsequent starvation, see Barbraud & Weimerskirch 2001a, or increased predation, see Ain-
7 ley *et al.* 2010b), and several models have been built on that basis (Jenouvrier *et al.* 2014). Yet,
8 these models mostly rely on the explicit assumption that movement amongst populations is neg-
9 ligible (Jenouvrier *et al.* 2014), leaving adult and juvenile survival and breeding output as the sole
10 factors driving population dynamics. This assumption, however, is not based on direct evidence,
11 but is rather motivated by technical difficulties in discriminating emigration from mortality of
12 tagged individuals (Hunter *et al.* 2010; Ehrlén & Morris 2015).

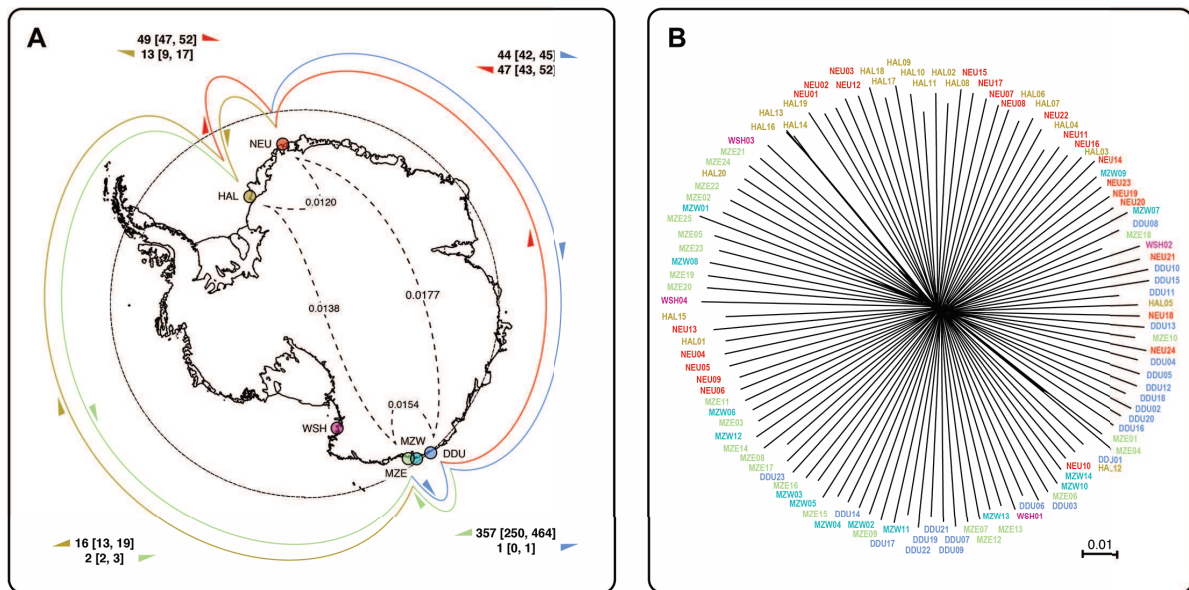
13 *§-109 Instability of Emperor penguin colonies.* The Emperor penguin, the only winter-breeding
14 top-predator species of Antarctica (Prévost 1961), stands at the forefront of the impacts of climate
15 warming (Ainley *et al.* 2010a). Emperor penguins breed nearly exclusively on sea ice: this unsta-
16 ble habitat makes the species immediately sensitive to local environmental changes. One of the
17 northernmost colonies, located on the Antarctic Peninsula, vanished during the last decades, as
18 sea ice retreated in that area (Trathan *et al.* 2011). Other colonies underwent a dramatic drop in
19 breeding success and population size shortly after a modification in local sea ice topology
20 (Kooyman & Ponganis 2014). The colony breeding on the tongue of the Mertz glacier, for in-
21 stance, disappeared after the 2010 calving of that glacier (Ancel *et al.* 2014), and the following
22 changes in local ice movements also had catastrophic consequences on the nearby Pointe Géolo-
23 gie colony, where the number of fledged chicks dropped from ~2,500 in 2010 to ~100 in 2014
24 (*personal observation*). All of these events have in common an identified proximal cause, usually
25 linked to modifications in the local sea-ice landscape that either forced adults to make longer
26 trips over the sea ice to reach open waters for foraging, with subsequent breeding failure resulting

The Emperor synnome - §110

1 from heightened energy expenditures (Massom *et al.* 2009; Kooyman & Ponganis 2014), or
2 removed the usual colony location (Ancel *et al.* 2014). On the other hand, recent empirical evi-
3 dence increasingly points to an important effect of dispersal in that species' respond to habitat
4 disruption (Forcada & Trathan 2009). The rapid recovery of the emperor penguin population in
5 Coulman Island confirms this assumption and excludes a peak in adult mortality followed by re-
6 growth (Kooyman & Ponganis 2014). Recent satellite and ground surveys have also shown that
7 whole emperor penguin colonies are able to relocate with or without an identified cause (Ancel *et*
8 *al.* 2014; LaRue *et al.* 2015). Finally, biologging experiments have emphasized the outstanding
9 distances regularly traveled by adult and juvenile emperor penguins (Thiebot *et al.* 2013). Here,
10 we demonstrate that dispersal is a fundamental component of demography in this long-lived
11 species, and propose that it does play a central role in its adaptive response at the continental
12 scale. As such, migration amongst colonies needs to be incorporated into demographic models in
13 order to achieve accurate projections.

14 *§-110 Study design.* We produced genome-wide RAD-sequencing data for 110 individuals from 6
15 Emperor penguin colonies representing the whole species' range (Fig. 31A). In order to assess
16 both fine- and large-scale processes, we sampled three colonies in a tight cluster around Terre
17 Adélie, in Eastern Antarctica: Eastern and Western Mertz colonies (Ancel *et al.* 2014 - 'MZE'
18 and 'MZW'), as well as the Pointe Géologie colony, near Dumont d'Urville research station
19 ('DDU'), all three within ~300 km. Two colonies were sampled in the Weddell Sea area, across
20 the continent: Atka Bay colony, near Neumayer research station ('NEU'), ~6,500 km away from
21 Terre Adélie, and Halley Bay colony ('HAL'), ~700 km further. Finally, one colony was sampled
22 from the Ross Sea area (Cape Washington, 'WSH'), ~1,700 km from Terre Adélie and ~8,000
23 km from Halley Bay.

The Emperor synnome - §111



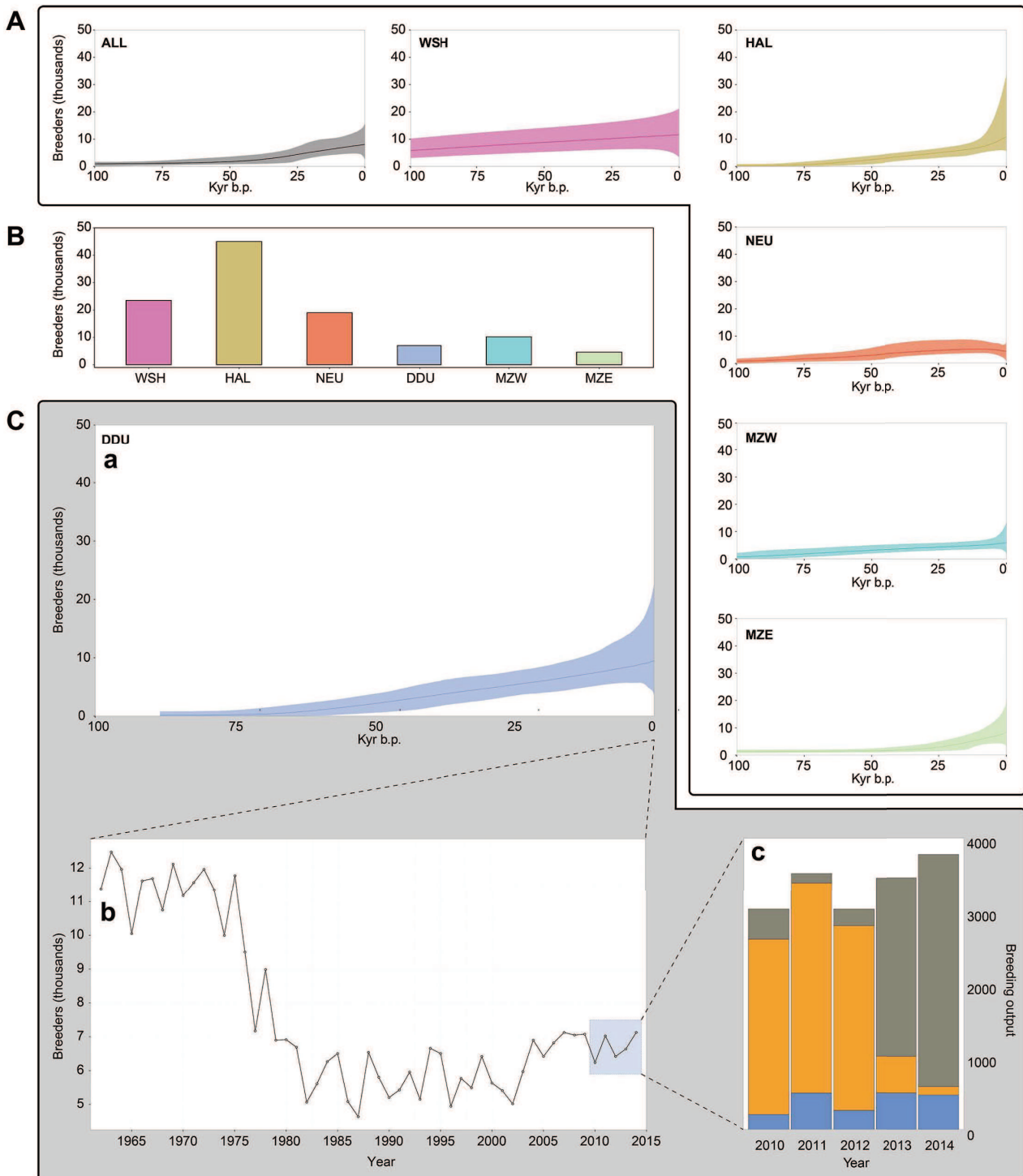
- 1 **Figure 31 | Drift and mixing concur to keep a high degree of worldwide homogeneity in the**
- 2 **Emperor penguin.** (A) Estimated pairwise migration rates (outer coloured links) and Fst (inner dashed links) be-
- 3 tween sampled colonies. Migrations as estimated from joint allele-frequency spectra. Colours refer to the receiving
- 4 populations. (B) Neighbour-net of all sampled individuals. Colours correspond to colonies on map (A).
- 5 **§-111 *A panmictic species.*** In striking contrast both with its fragmented geographical distribution
- 6 and with our current knowledge about other marine predators (Hoelzel 1998; Friesen *et al.* 2007;
- 7 Molfetti *et al.* 2013), the Emperor penguin exhibits a remarkable degree of genetic homogeneity
- 8 at the continent scale. Pairwise fixation index (Fst) values, calculated either as a function of allele
- 9 frequency covariance between populations over variable sites, or from called genotypes using Re-
- 10 ich's estimator (Reich *et al.* 2009), are very low (see details in Supplementary methods §129 p.
- 11 242 and Table 6), and only about 0.5% of total variance is explained by colony structure (as per
- 12 AMOVA, see §129 p. 242). A neighbour-net based on pairwise Hamming distances shows con-
- 13 siderable admixture between areas: individuals are roughly sorted according to geographical loca-
- 14 tion, but inter-individual variability is largely dominant (Fig. 31B). Inference of population split
- 15 topology based on allele frequency variation amongst populations (§132 p. 246) also supports
- 16 this view: neutral genetic differentiation from the ancestral population increases eastward from

The Emperor synnyme - §112

1 WSH to MZE, but with numerous migration events inferred between most colonies (Fig. 38).
2 Finally, high genetic mixing is supported by all classical descriptors of genetic variation (Hardy-
3 Weinberg equilibrium, heterozygosity, nucleotide diversity, Tajima's D - as estimated for non-
4 coding regions of the genome from downsampled haplotypes, see §129 p. 242, as well as prin-
5 cipal component analysis, see §131 p. 245). The very limited genetic drift observed between
6 colonies separated by several thousands of kilometers suggests intense gene flow in this flightless
7 seabirds along the coast of Antarctica.

8 *§-112 Demographic reconstruction.* The existence of a common, homogeneous gene pool for the
9 entire species also implies that all present-day colonies share a common demographic history. In
10 order to test that, we reconstructed past Emperor penguin population size changes in BEAST2
11 (Bouckaert *et al.* 2014) under an extended Bayesian skyline plot model. In accordance with our
12 expectation, reconstructions based either on a single colony or on haplotypes sampled randomly
13 from the whole continent converge to the same estimate of effective population size and to the
14 same demographic history. All show a moderate increase in population size over the past 100,000
15 years, as well as in the full sample (Fig. 32A, §136 p. 252), regardless of the very different
16 present-day colony size (Fig. 32B). This trend is in accordance with the findings of Li *et al.* 2014
17 based on a single-genome pairwise sequentially Markovian coalescent approach. Emperor pen-
18 guin population size does not appear to have been affected by the last glacial period. This is in
19 stark contrast with the sudden post-glacial population expansion of the Emperor's sister species,
20 the King penguin (*Aptenodytes patagonicus*) from the sub-Antarctic region (Trucchi *et al.* 2014), a
21 difference likely explained by the contrasting breeding habitats and strategies of the two species,
22 the King penguin being dependent on year-round ice-free breeding sites. The agreement of coa-
23 lescent histories at both local and global sampling scales supports the idea that all extant colonies
24 share the same genetic pool, and that the effects of neutral genetic drift between distant areas are
25 mostly counterbalanced by the intensity of gene flow.

The Emperor synnome - §112



1 **Figure 32 | Demography is a matter of scales.** (A) Demographic reconstructions (extended Bayesian skyline
 2 plots), for all samples, and per-colony. Solid line: mean population size. Shaded area: 95% confidence interval. Blue
 3 area: Last Glacial Maximum (LGM). (B) Census size for the six analysed colonies (from Fretwell *et al.* 2012 and An-
 4 cel *et al.* 2014) (C) Three different demographic times-scales for DDU colony: a. coalescent-scale (EBSP recon-
 5 struction, see 2A), b. monitoring-scale (from Jenouvrier *et al.* 2006 and unpublished data), c. current catastrophic breed-
 6 ing failure: blue, eggs lost during brooding, gray, chicks found dead, orange, successfully fledged chicks.

1 *§-113 Continental-scale migration.* In order to quantitatively assess the importance of continental
2 dispersal, we modelled the inter-colony migration rates required to generate such a level of genet-
3 ic admixture. We first co-estimated effective population size and bidirectional migration rates be-
4 tween a subset of four colonies representing our whole sampling area by simulating genetic data
5 under a continuous-time Markovian coalescent model against the observed two-dimensional al-
6 lele frequency spectra, using a composite-likelihood approach (Excoffier *et al.* 2013, see §135 p.
7 250). Analysis was calibrated using a RADome mutation rate as estimated for King penguin
8 (Trucchi *et al.* 2014) and a generation time of 16 years (Jenouvrier *et al.* 2014). Effective popula-
9 tion sizes all converge to an average of ~4,000 (from 1120 to 7640) breeding individuals (de-
10 tailed estimates in supplementary Table 7 and Fig. 31A), in keeping with the observed median
11 size of extant colonies (Fretwell *et al.* 2012). Each colony is estimated to receive, on average, be-
12 tween ~0.7% and ~4.2% of its effective population size in migrants every generation. Thus, if we
13 were to scale these results to present-day observed census size, a colony such as Pointe Géologie
14 (DDU), with a count of ~7,000 breeding adults, would exchange, on average, between ~260 and
15 ~300 migrants (3.8% to 4.2%) per generation with the rest of the continent. We further validat-
16 ed these results using a haplotype-based multilocus Bayesian approach as implemented in Mi-
17 grate-n (Beerli & Palczewski 2010, see §134 p. 247). Model ranking using Bayes factor model
18 choice gives clear support for a full-migration model with very high gene flow. Estimated migra-
19 tion rates (M) and population sizes (Θ) are highly homogeneous (mean mutation-scaled (μ) mi-
20 gration rate $M = m / \mu = 2,358 \pm 130$, mean mutation-scaled effective size $\Theta = 0.0018 \pm 8.2e^{-5}$).
21 Considering a conservative range of μ for our subset of loci (Trucchi *et al.* 2014), we can estimate
22 that each colony receives on average from 5.40% (± 0.22) to 10.00% (± 0.42) of its total effective
23 size as migrants from other colonies at each generation (see §134 p. 247). Although slightly high-
24 er, this estimate is of the same order of magnitude as the one derived from the joint allele fre-
25 quency spectra. We highlight the fact that these are estimates of the migration parameters aver-

The Emperor synnome - §115

1 aged over many generations, not of instantaneous dispersal rate (see S-0 p. 236): neither
2 population sizes, nor migration rates, should be interpreted as referring literally to the immediate
3 state of extant colonies but rather to the average coalescence-based population size and migration
4 rate over time. Dispersal rate itself may also be very heterogeneous at any given point of time,
5 with individual colonies showing positive or negative dispersal balance according to the current
6 local habitat conditions and modifications.

7 *§-114 Consequences of high dispersal.* Dispersal thus appears to be a central aspect of the Emperor
8 penguin's coloniality, and recent observations suggest that this view may be extended to other
9 seabird species (Friesen *et al.* 2007). For example, we have repeatedly recaptured King penguins
10 tagged at one colony in the Crozet archipelago at different colonies in the same archipelago, and
11 as colonies in the Iles Kerguelen, 1,500 km away (*unpublished data*). As pointed out by Mayr
12 (Mayr 1970), « *a high dispersal ability is a necessity for occupants of temporary habitats* ». Since dis-
13 persal has been the immediate response for colonies facing habitat disturbance in the recent past
14 (Forcada & Trathan 2009; Trathan *et al.* 2011; Kooyman & Ponganis 2014), it may therefore be
15 thought to play an important role by adding flexibility to an otherwise rigid philopatric system,
16 and by allowing the species to dynamically exploit the best breeding locations in rapidly changing
17 polar environments. Yet, such demographic events are of local rather than global significance.
18 Population trends extracted from a single colony reflect the immediate quality of the focal loca-
19 tion on a generation scale, rather than species-wide parameters (Fig. 32C). In this context, the
20 colony is not a truly relevant demographic unit, but rather a transient aggregation of individuals
21 at a particular point of time. In order to distinguish this structure from panmictic and metapopu-
22 lation systems, we propose the term *synnome* for this combination of an exceptionally fragmented
23 space and a very fluid gene pool (see full definition in Supplementary methods S-0 p. 236).

24 *§-115 Future prospects.* Our findings highlight the importance of adopting a cross-disciplinary ap-
25 proach, integrating population genomics and behavioural ecology, to the study of population dy-

The Emperor synnyme - §115

1 namics. Such an approach should help in the development of more robust and accurate demo-
2 graphic projections at the whole species level. These refined projections will then be more likely
3 to allow estimation of the extent of threats to vulnerable species and identifying the proximal
4 causes of their decline. Our results also highlight the absolute novelty of the Emperor penguin's
5 single fully-panmictic population. Although exceptional if compared with current knowledge
6 about other marine predators, such as turtles (Molfetti *et al.* 2013), marine mammals (Hoelzel
7 1998), and most seabird species (Friesen *et al.* 2007), intense dispersal is expected to be a com-
8 mon evolutionary strategy in unstable high-latitude environments (Friesen *et al.* 2007), especially
9 under ongoing climate warming. As a single genetic population, Emperor penguins will respond
10 to climate change through a unified evolutionary trajectory. This insight thus presents new chal-
11 lenges to our understanding of how current climate scenarios will impact upon the future of the
12 most cold-adapted species in the world, an iconic bio-indicator of the delicate Antarctic
13 ecosystem.

14 *Methods Summary*

15 Samples were collected from fledged chicks or frozen carcasses at the end of the breeding season.
16 RAD sequencing followed Baird and colleagues' original protocol (Baird *et al.* 2008) with minor
17 modifications (see Supplementary methods §126 p. 237 for details). Sequences were aligned on
18 the Emperor penguin's reference genome (Zhang *et al.* 2011b) and cleaned using standard tools.
19 Genotyping was performed using GATK's HaplotypeCaller (DePristo *et al.* 2011) pipeline,
20 Stacks (Catchen *et al.* 2011; Catchen *et al.* 2013), and ANGSD (Korneliussen *et al.* 2014).
21 Genotype likelihoods and site-allele-frequency likelihood distributions were calculated in
22 ANGSD (Korneliussen *et al.* 2014) under a samtools/mpileup model (§127 p. 238 and §128 p.
23 241). Only autosomal, high-quality loci were retained. Summary statistics were calculated in
24 ngsTools (Fumagalli *et al.* 2013; Fumagalli *et al.* 2014), *adegenet* (Jombart 2008), and custom R

The Emperor synnome - §117

1 and shell scripts (§129 p. 242). For haplotype-based analyses, each individual was randomly
2 downsampled to a haploid genome by discarding one haplotype at each locus, in order to avoid
3 low-coverage biases. Hamming distances and identity-by-descent estimation were calculated in
4 PLINK (Purcell *et al.* 2007, see §130 p. 244). For population size and migration rate estimates,
5 we used Migrate-n (Beerli & Palczewski 2010), using downsampled haplotypes (§134 p. 247),
6 and *fastsimcoal2* (Excoffier & Foll 2011; Excoffier *et al.* 2013), using empirical joint-frequency
7 spectra calculated in ANGSD (§135 p. 250). BEAST2 (Bouckaert *et al.* 2014) was used to co-es-
8 timate effective population size and population size changes through time (§136 p. 252). See on-
9 line supplementary material for more detailed information.

10 *Modelling the changes in Emperor penguin habitat.*

11 *§-116 Aim of this study.* The following part was not included in the original article « *Full circum-*
12 *polar migration ensures evolutionary unity in the Emperor penguin* ». The strong contrast between
13 the marked demographic response of the King penguin on the one hand (see *The King synnome*,
14 p. 163) and the lack of detectable response of the Emperor penguin on the other, however, raises
15 the question of the direct environmental determinants of demographic trends in the second
16 species. Here, we propose a preliminary model to explore the relationship of the Emperor pen-
17 guin to its direct abiotic environment, defined both by sea ice, and by atmospheric variables. Al-
18 though the conclusions of this study are very consistent with the demographic trends inferred
19 from genetic data, further work may still be required, outside the framework of the coarser
20 CMIP5 models, in order to understand the finer local dynamics of sea ice formation and coastal
21 polynya activity.

22 *§-117 How to model the Emperor's habitat?* Our survey of the Emperor penguin's genetic architec-
23 ture brings to light two important traits: (i) the whole species is organised as a single synnome,
24 and migration between colonies is a central demographic process, and (ii) the Emperor penguin

The Emperor synnome - §117

1 population was not affected by the late Pleistocene and Holocene climatic events (which is also
2 confirmed by our reconstructions in *The King synnome*, p. 163). A corollary of the Emperor's high
3 dispersal ability, is the possibility that colonies may relocate when local habitat conditions be-
4 come unsuitable, provided better conditions can be found anywhere in the vicinity (see *Empirical*
5 *evidence of heterogeneous dispersal*, p. 259, and LaRue *et al.* 2015).

6 These findings redefine the way the Emperor's habitat may be modelled. In a previous attempt
7 that we discussed earlier in this chapter, Jenouvrier *et al.* 2014 used the observed complex correla-
8 tion between sea ice conditions and Emperor penguin survival and breeding success on the
9 Pointe Géologie colony to define a population dynamic model, that is extended to the whole
10 continent using present-day colony locations as the focal points for assessing local environmental
11 conditions. We showed that a first shortcoming in this approach is the assumption that dispersal
12 was not interfering with local dynamics in response to icescape changes (or to other forcings such
13 as human disturbance) - which may result in a biased estimate (and hence model) of survival and
14 breeding output.

15 But a second, and equally important difficulty lies in the nature of the Emperor penguin's habi-
16 tat. Indeed, the Emperor needs both dense and firm ice for establishing its breeding colonies, and
17 open-water polynyas (see §18 p. 65) for foraging during the breeding season. Since only coastal
18 (usually *latent heat*) polynya are a requisite, the total sea ice extent around the continent is of
19 comparatively little importance, as long as offshore wind stress maintains open-water conditions
20 at regular intervals along the shore (Ainley *et al.* 2010a). However, the mobility of Emperor pen-
21 guin colonies imply that the distance between a particular present-day colony location and a
22 polynya is not a hard constraint: as opposed to the King penguin whose breeding locations are
23 fixed (see *The King synnome*, p. 163), the Emperor is free to select a colony space that will opti-
24 mise both the strength of the ice for chick rearing, and the proximity of polynyas for foraging.

The Emperor synnome - §118

1 Ultimately, this means that although the local icescape has a strong impact on short-term colony
2 demography, the species as a whole is sensitive to changes in the global state of the coastal
3 Antarctic sea-ice rather than to local conditions: as long as an equivalent breeding location is
4 available, the expected response is relocation rather than extinction (see also Forcada & Trathan
5 2009 on that point). For example, while an increasing distance between a subantarctic island and
6 the Polar Front translates as an increased foraging constraint on the King penguin (see §80 p.
7 167), an increasing distance between a given present-day Emperor penguin rookery and the near-
8 est polynya will most likely encourage the penguins to change their colony location accordingly
9 to track their optimal habitat conditions, with no predictable cost for the population. Thus, the
10 local, colony-based paradigm we adopted for the analysis of the King penguin's habitat in *The*
11 *King synnome*, p. 163, and that Jenouvrier *et al.* 2014 deployed for predicting the Emperor pen-
12 guin's future demography does not seem to be appropriate for the latter. Here, we explore an al-
13 ternative approach, based on the global sea ice conditions around the coast of Antarctica.

14 *§-118 Sea ice in the CMIP5 experiments.* An additional difficulty in predicting the Emperor pen-
15 guin's response to climate change is a technical limitation: the current CMIP5 AO-GCMs (see
16 §50 p. 130) normally provide outputs mapped on a 1° grid, which, under the higher polar lati-
17 tudes, is a rectangle of ca. 111 x 50 km - too coarse a grid to explicitly model polynyas, which are
18 often less than 50km wide. Moreover, the current representation of Antarctic sea ice in the
19 CMIP5 ensemble members still fails to capture important aspects of the observed conditions,
20 such as the significant positive trend in sea ice extent (Zhang 2007; Turner *et al.* 2013; Shu *et al.*
21 2015) - a shortcoming that may be due to the fact the Antarctic Ice Sheet is currently specified as
22 a boundary in the CMIP5 ensemble, and not explicitly modelled (Kusahara *et al.* 2015). Due to
23 these constraints, it is currently impossible to accurately determine the distribution on polynya
24 along coastal Antarctica in the CMIP5 framework.

The Emperor synnome - §119

1 However, our reconstructions of past population size changes give us an alternative viewpoint on
2 the question. Although variation in habitat quality through space can be alleviated by dispersal,
3 variations through time are rather final, and oblige the species to adapt or decline. Thus, the lack
4 of response of the Emperor penguin to Pleistocene and Holocene climate change implies that the
5 global sea ice conditions that prevailed at these periods provided suitable habitat for the species.
6 Comparing the CMIP5 predictions for sea ice conditions during the Last Glacial Maximum, at
7 midHolocene, during the historical period, and at the end of the 21st century should therefore
8 allow us to simplify the complex spatial component of the Emperor's relationship to sea ice, and
9 to consider several successive states of global sea ice distribution and characteristics. However, the
10 identified biases in sea ice representation in CMIP5 models restricts the scope of this study: we
11 expect it to be highly informative regarding how far CMIP5 models support a scenario in which
12 changes in sea ice configuration brings the Emperor penguin to large-scale decline throughout
13 the continent - but less so as to the species' actual demographic trend.

14 *§-119 Model and variable selection.* Our choice of models, within the CMIP5 ensemble, is limited
15 by the availability of outputs and the difficulty of performing objective model-choice based on
16 outputs. Only six models so far provide outputs for the PMIP3 palæoclimate experiments, the
17 historical period, and three RCP scenarios (RCP-2.6, RCP-4.5 and RCP-8.5 - see §102 p. 201):
18 the CCSM4, CNRM-CM5, the GISS-E2-R, the IPSL-CM5A-LR, the MPI-ESM-P, and the
19 MRI-CGCM3 (see Table 5 for details). Analysis was performed on six 20-years period, during
20 the LGM, at midHolocene, in 1980-1999 for historical runs, and 2080-2099 for the three RCP
21 runs. Outputs were further divided into five yearly breeding stages (based on Jenouvrier *et al.*
22 2014): interbreeding (January to March), parading and egg-laying (April-May), incubating
23 (June-July), hatching and small-chick rearing (August-September) and rearing of the thermally
24 independent chick (October to December).

1 **Table 5 | Ensemble members used in sea ice modelling.** Model outputs were downloaded from the
 2 IPCC archive (http://www.ipcc-data.org/sim/gcm_monthly/AR5/Reference-Archive.html). Only one ensemble
 3 member was used for each model (r1i1p1 whenever available).

4	Model	Institution	LGM	mid-Holocene	21st century
5	CCSM4	National Center for Atmospheric Research (USA)	r1i1p1	r1i1p1	r1i1p1
6	CNRM-CM5	Centre National de Recherches Météorologiques, Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (France)	r1i1p1	r1i1p1	r1i1p1
7	GISS-E2-R	NASA/GISS (Goddard Institute for Space Studies) (USA)	r1i1p1	r1i1p1	r1i1p1
8	IPSL-CM5A-MR	Institut Pierre Simon Laplace (France)	r1i1p1	r1i1p1	r1i1p1
9	MPI-ESM-P	Max Planck Institute for Meteorology (Germany)	r1i1p1	r1i1p1	r1i1p1
10	MRI-CGCM3	Meteorological Research Institute (Japan)	r1i1p1	r1i1p1	r1i1p1

11 **Sea Ice Concentration.** We focused primarily on Sea Ice Concentration (SIC), as proposed by Je-
 12 nouvrier *et al.* 2014. Considering that Emperor penguins are primarily affected by the local ice
 13 conditions around their coastal breeding site, and not by the total sea ice extent (Ainley *et al.*
 14 2010a), we restricted the analysis to a zone of 3° from the coastline and offshore - which corre-
 15 sponds to a distance of ~330 km, largely including the maximum length of an Emperor penguin's
 16 walking trip to the nearest polynya (~100km at most before systematic breeding failure, see Mas-
 17 som *et al.* 2009). Within that zone, we considered four global descriptors of SIC: (i) the mean
 18 concentration, (ii) the variance of concentration, (iii) Moran's global index of spatial autocorrela-
 19 tion, and (iv) the sea ice extent within the coastal zone, defined as the ratio of the total surface
 20 with a SIC > 15% (Ainley *et al.* 2010a). For each of these parameters, we calculated multi-model
 21 mean and across-model standard deviation. In order to quantify the relative importance of
 22 changes in sea ice conditions through time, compared to model uncertainties, we also calculated
 23 the coefficients of variation - either between periods (*i.e.* geological periods, and RCP scenarios)
 24 within models, or between models within periods.

25 **Sea Ice Thickness and Wind Stress.** Three other complementary model output variables were as-
 26 sessed in order to capture possible variation in the icescape that would not be reflected in SIC.

The Emperor synnyme - §120

1 First, Sea Ice Thickness (SIT) was taken as a different proxy for the presence of polynya. SIC de-
2 scribes the fragmentation of sea ice, which is normally high in polynya areas, but may be inaccu-
3 rately represented in the current models. SIT is also expected to be low in polynya areas, due to
4 the ongoing sea ice formation and offshore drift (Ainley *et al.* 2010a), and may not be subject to
5 the same biases as sea ice concentration in the current models. For SIT, multi-model mean and
6 across-model variance were calculated for the continent-wide mean thickness, standard deviation,
7 and Moran's index. Across-models and across-periods coefficients of variations were also
8 calculated.

9 Offshore wind stress is the proximal cause of polynya formation (see §18 p. 65). Since sea-ice
10 models are usually independent sub-models within the CMIP5 AO-GCMs (see *e.g.* Gent *et al.*
11 2011), we account for the fact that wind patterns themselves may be accurately modelled, but
12 that the sea-ice response at expected polynya areas may be less reliably predicted. We consider
13 two variables representative of offshore wind stress in Antarctica: the Northward surface wind
14 speed component (VAS, in $\text{m}\cdot\text{s}^{-1}$) and the Surface northward downward wind stress (TAUV, in
15 Pa). For both variables, we extracted multi-model mean and across-model variance for the conti-
16 nent-wide mean, standard deviation, and Moran's index for both VAS¹ and TAUV. Across-models
17 and across-periods coefficients of variations were also calculated. All computations relied on the
18 Climate Data Operators toolset (CDO 2015, available at: <http://www.mpimet.mpg.de/cdo>) and
19 custom R scripts.

20 *§-120 Is coastal sea ice really changing?* SIC is predicted to decrease in all 6 included models, while
21 its variance increases from LGM conditions until RCP-8.5 conditions (Fig. 33A). SIT, on the
22 other hand, shows divergent trends across models, either slightly decreasing (in CCSM4 and
23 MRI-CGCM3), or stable from LGM to 21st century (in the other models), while its variance de-
24 creases or remains unchanged (Fig. 33B). For both variables, homogeneity of spatial distribution,

1. VAS is not output by the CCSM4 model - for this variable, we therefore only use a 5-member ensemble.

The Emperor synnyme - §120

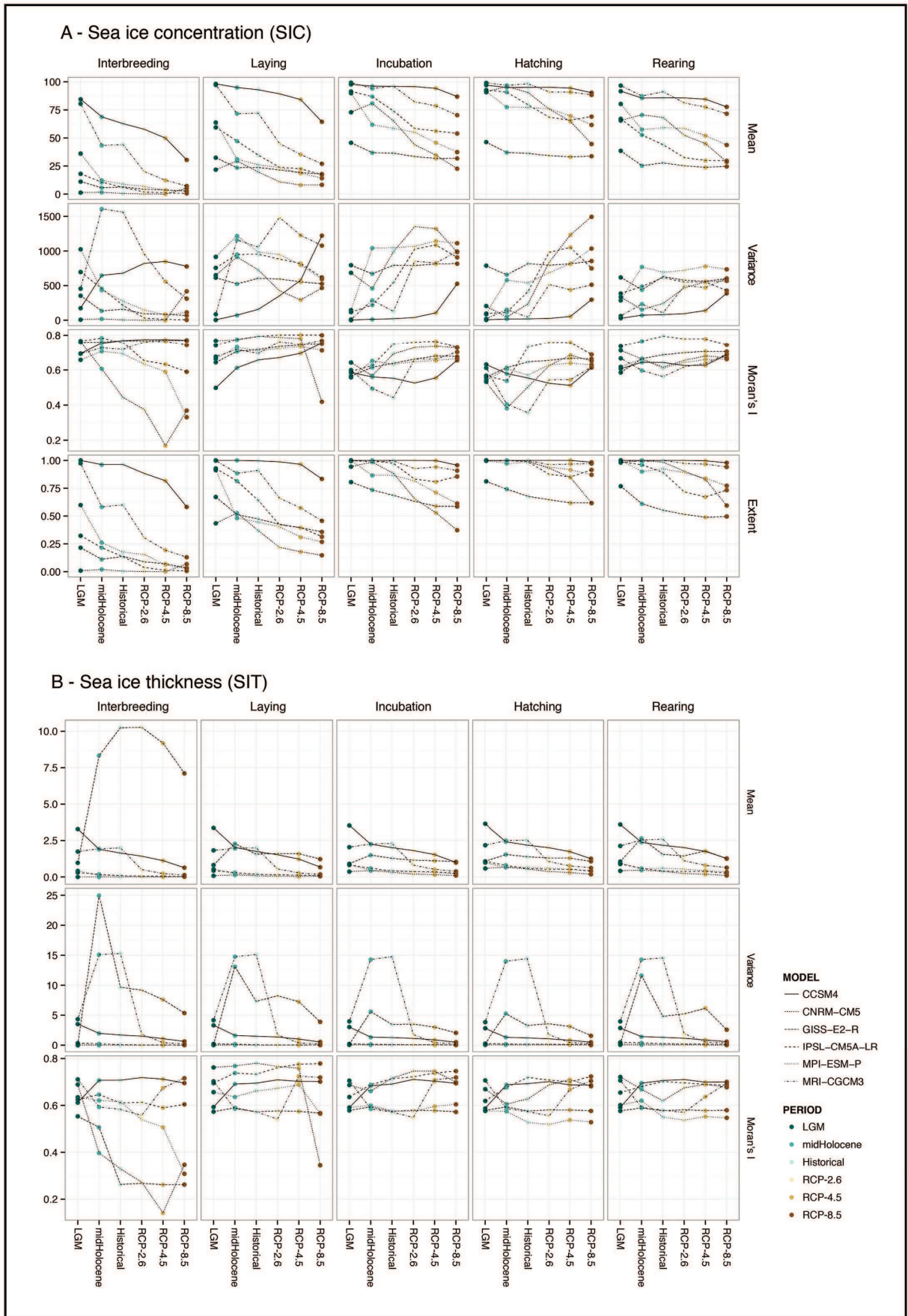
1 as summarised by Moran's global index, remains unchanged throughout Pleistocene and Anthro-
2 pocene - the rather high value ($0.5 < I < 1$) is consistent with a local clustering of low-ice areas in
3 polynya around the continent. As expected, SIC changes most during the warmer months - in
4 summer (interbreeding period), autumn (egg-laying) and spring (rearing). Overall, despite the
5 known bias towards sea ice reduction in the CMIP5 models, the *extent* of coastal sea ice only de-
6 creases during the interbreeding period, and marginally the egg-laying period, in most models.

7 Wind patterns, on the other hand, are remarkably constant on the Antarctic coastal band from
8 LGM to RCP projections. Let aside the widely divergent GISS-E2-R model, offshore surface
9 wind stress (TAUV) remains unchanged, in terms of absolute average value, spatial variance, and
10 distribution (Fig. 33C). Wind speed decreases most noticeably between at the LGM-Holocene
11 transition, but only decreases during the interbreeding period in the RCP projections (Fig. 33D).

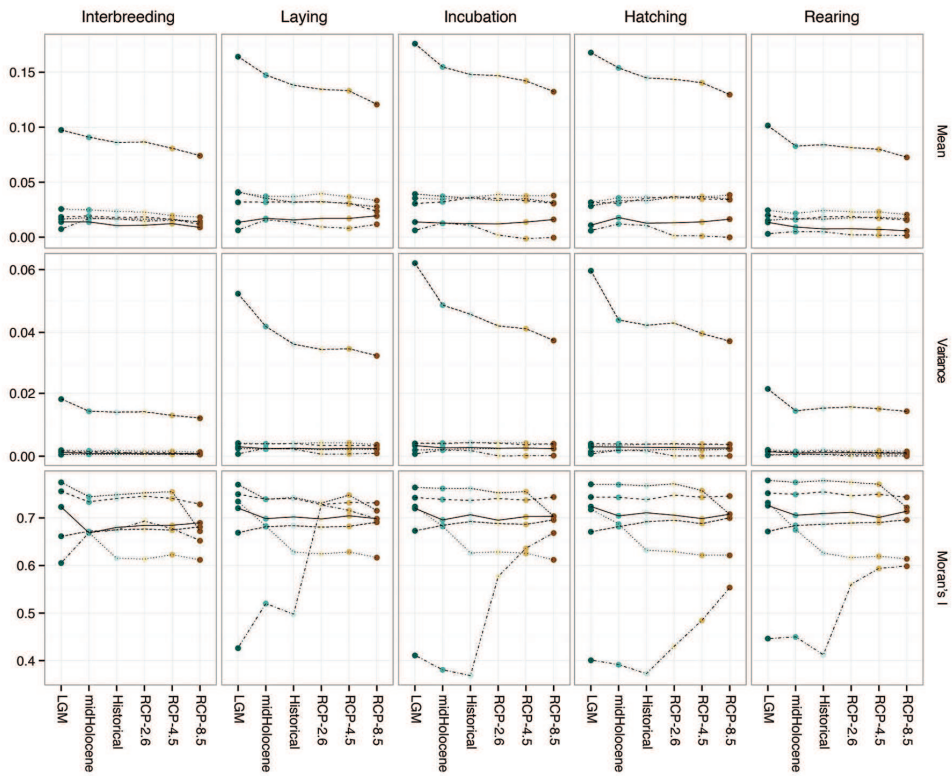
12 Since offshore wind stress is the primary cause of polynya formation, the stability of coastal off-
13 shore wind patterns may be an indication that polynya dynamics did not change much during
14 the Pleistocene and the Holocene, and may not change much in the coming decades either. This
15 is consistent with our reconstruction of Emperor penguin population size through the past mil-
16 lennia (see §82 p. 169 and §112 p. 218): the lack of demographic response to the glacial maxi-
17 mum and following deglacial period implies that the species' habitat was not significantly
18 changed, and that Emperors were able to find a sufficient number of polynya for overwinter for-
19 aging throughout the period.

20 **Figure 33 | (NEXT PAGE) Changes in sea ice configuration along coastal Antarctica from**
21 **Pleistocene to the 21st century.** Mean, variance and Moran's global index for sea ice concentration in % (A)
22 and thickness in meters (B), and offshore surface wind stress in Pa (C) and speed in $m.s^{-1}$ (D). In (A), sea ice extent is
23 additionally defined as the ratio of sea ice surface with a concentration below 15%.

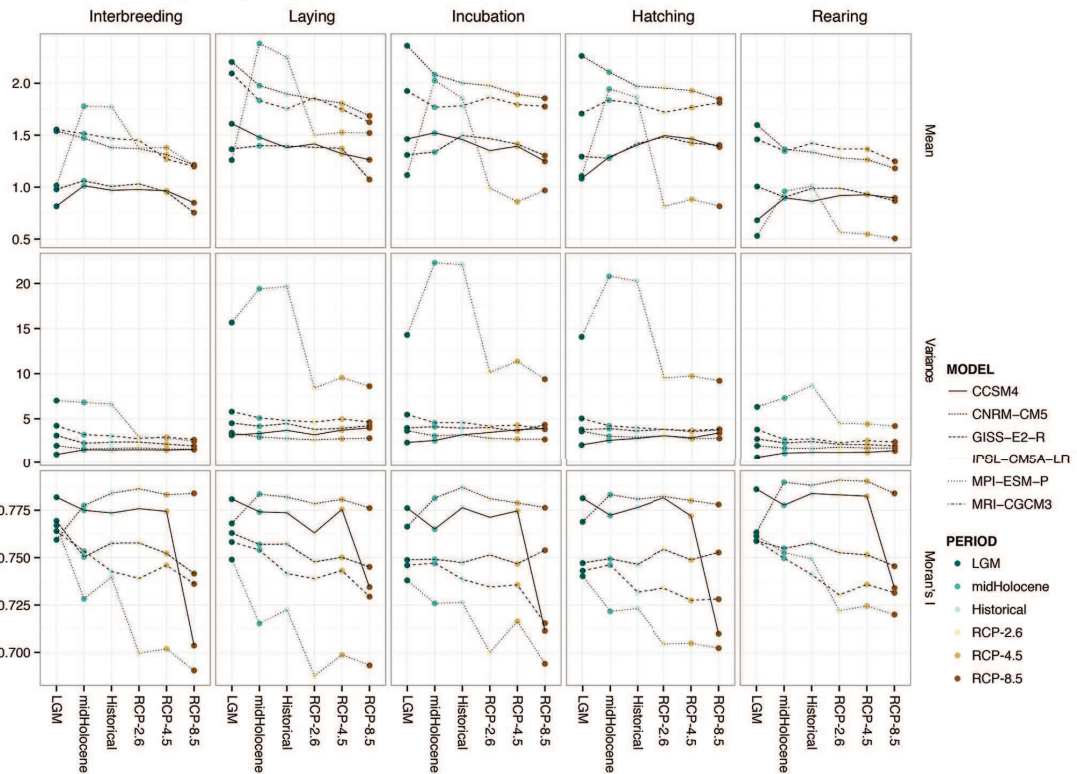
The Emperor synnome - §120



C - Surface wind stress (TAUV)

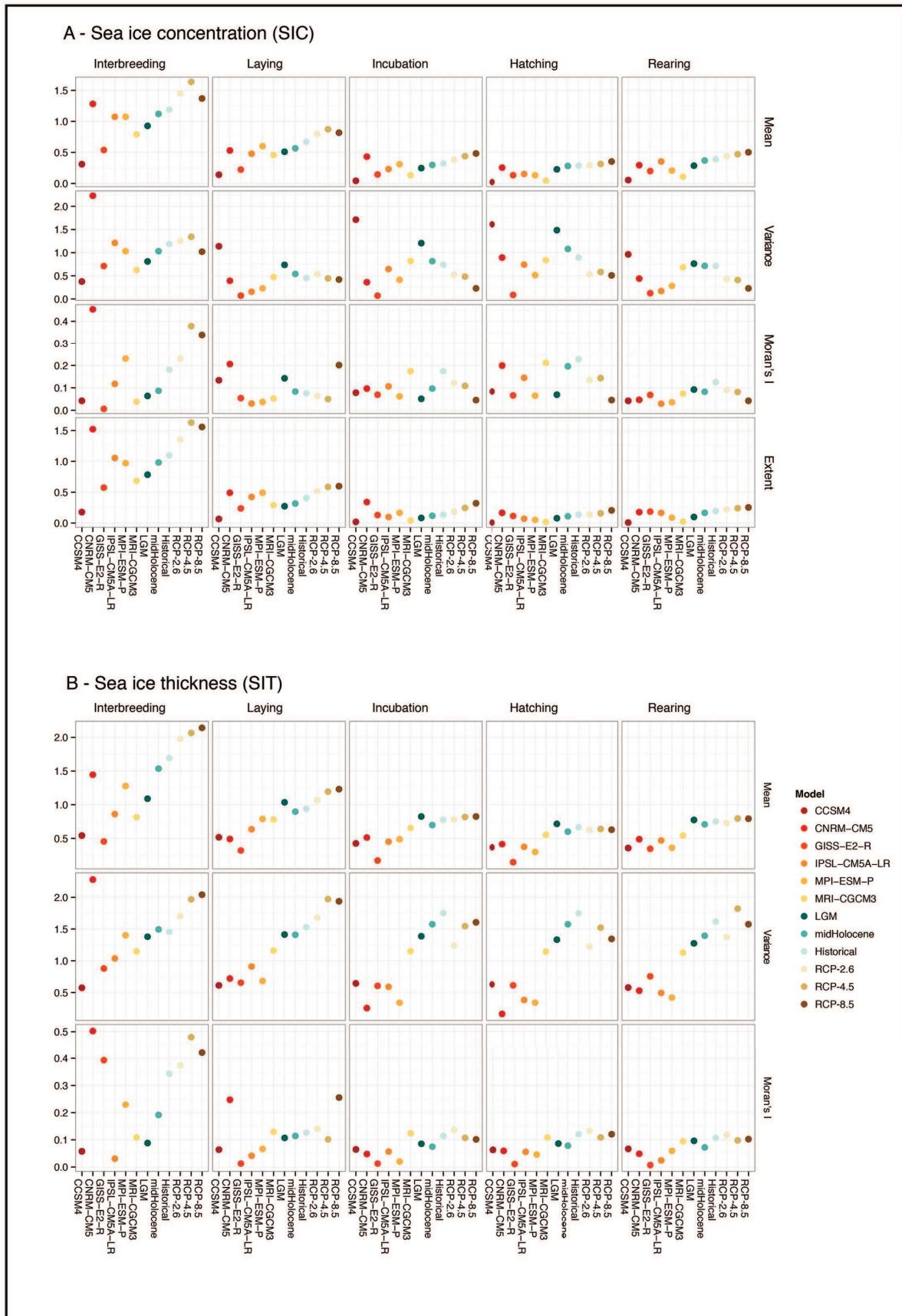


D - Surface wind speed (VAS)

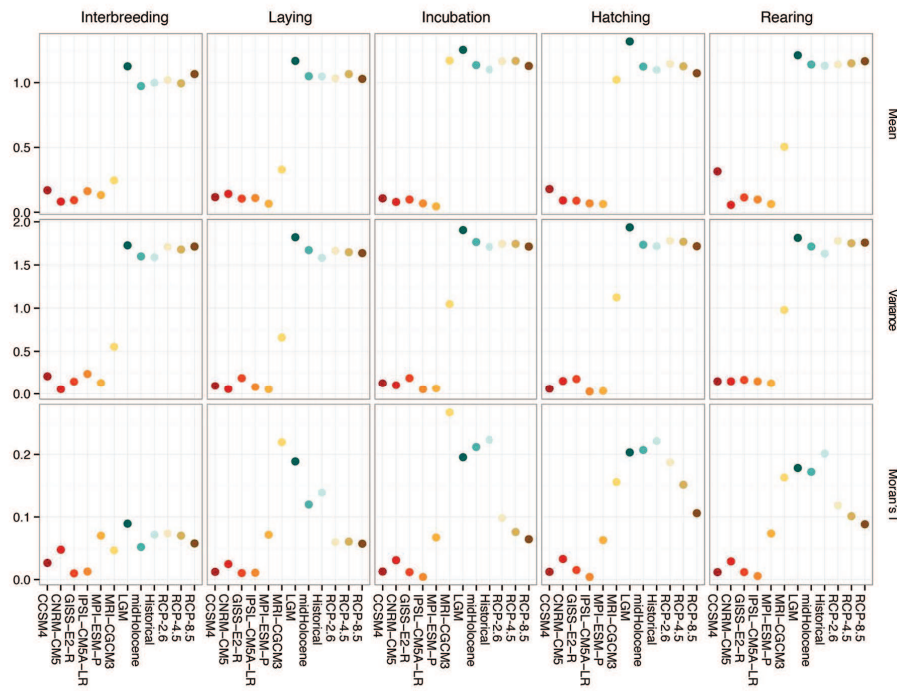


1 §-121 *Uncertainties in the CMIP5 models.* Comparison of per-model and per-period coefficients
2 of variation (CV) stresses the large uncertainty that remains in the current AO-CGMs for the
3 sea-ice zone around the Antarctic. For all four variables, CVs per period are equal or greater than
4 CVs per model (Fig. 34) - in other words, the reconstructed and predicted changes are of the
5 same order, or smaller, than the uncertainty in the models. Although such a comparison is not a
6 direct assessment of the significance of trends in the models, it means that any detected trend still
7 remains within the bounds of possible model uncertainty. In keeping with previous assessments
8 (see *e.g.* Ainley *et al.* 2010a; Turner *et al.* 2013; Shu *et al.* 2015), wind patterns exhibit far less
9 across-model variation than ice patterns, and may thus appear more robustly represented than ice
10 patterns. Taken together, these different observations indicate that (i) the current generation of
11 models disagrees strongly on the future ice patterns, and to a lesser extent wind patterns, around
12 coastal Antarctica, prompting the development of more detailed local models (integrating the dy-
13 namics of the ice sheet in particular, see Kusahara *et al.* 2015), and (ii) taking this observed un-
14 certainty into account, the evidence from the assessed models does not indicate any strong
15 change in ice and wind patterns since the LGM, even in under the more pessimistic RCP-8.5
16 scenario.

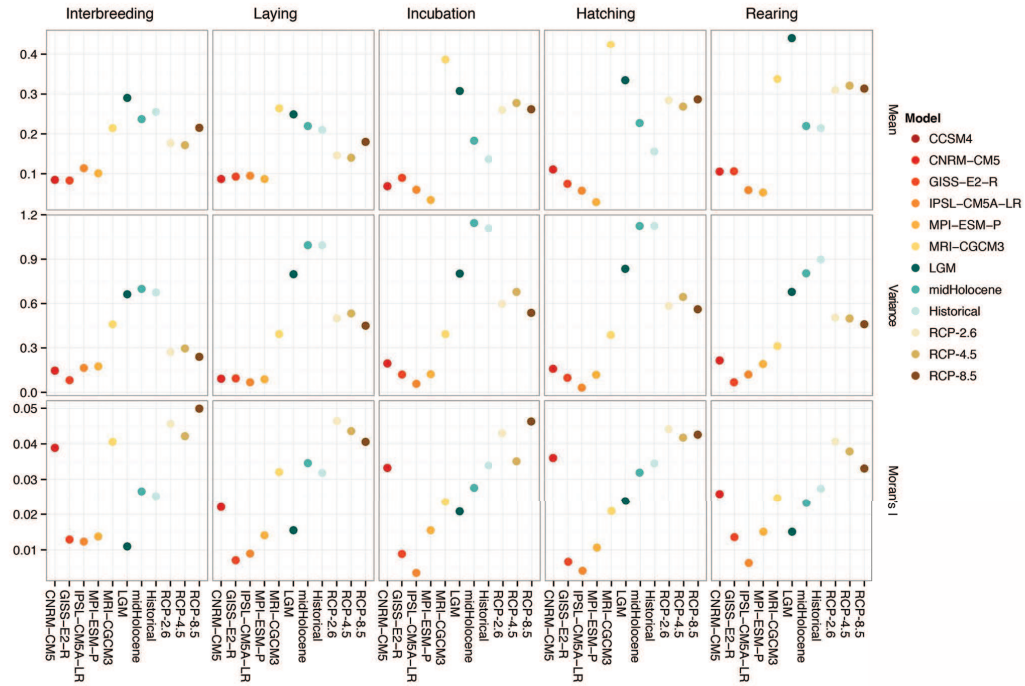
17 **Figure 34 | (NEXT PAGE) Uncertainty in the representation of sea ice in the CMIP5 models.**
18 Coefficients of variation (CV) for sea ice concentration in % (A) and thickness in meters (B), and offshore surface
19 wind stress in Pa (C) and speed in $\text{m}\cdot\text{s}^{-1}$ (D). In (A), sea ice extent is additionally defined as the ratio of sea ice sur-
20 face with a concentration below 15%. CV are computed per model, or per periods.



C - Surface wind stress (TAUV)



D - Surface wind speed (VAS)



The Emperor synnome - §122

1 §-122 *Can we predict Emperor penguin demography based on sea ice?* Our examination of the habi-
2 tat of the Emperor penguin shows the strong difference that oppose the two *Aptenodytes* species.
3 While the King penguin's fixed and fragmented breeding locations forces it to undergo drastic
4 range contractions and expansions in response to Southern Ocean chances, it appears that the
5 Emperor penguin's ability to relocate when the local icescape changes, together with the para-
6 doxical stability of the coastal Antarctic habitat, allowed it to maintain a constant population size
7 throughout the main climatic events of the Pleistocene.

8 Although the Antarctic sea-ice area as a whole underwent large changes during the Pleistocene
9 glacial-deglacial cycles (see §9 p. 42 and *The King synnome*, p. 163), the coastal zone stands out as
10 a particular region. First, as it is buffered by the large expanses of pack ice that shield it from the
11 lower latitudes, changes that occur at the edges of the ice remain mostly silent in the coastal area.:
12 this is visible in the fact that even the projected changes in SIC are mostly visible in summer,
13 when this buffer zone is reduced. Second, because the most important sea ice features, for the
14 Emperor penguin, are not directly linked to the total amount of sea ice, but rather to the dynam-
15 ics of its formation in the coastal wind-stress-driven divergence zones: thus, provided wind stress
16 remains constant, an increase or decrease in total sea ice extent may have but little impact on the
17 Emperor's ability to find accessible foraging areas. Indeed, on very short time scales (*i.e.* shorter
18 than a generation), it is possible that abrupt changes in the repartition of sea ice may result in ad-
19 verse effects for a single colony, as have been documented in Pointe Géologie (Barbraud &
20 Weimerskirch 2001a) or in Coulman Island (Kooyman & Ponganis 2014), however, relocation
21 to a more favourable location in the area seems to be rapid (see LaRue *et al.* 2015 and *Empirical*
22 *evidence of heterogeneous dispersal*, p. 259). Overall, the long-term stability of the Emperor pen-
23 guin population implies that local events have little influence on the resilience of the species as a
24 whole.

The Emperor synnome - §124

1 These results do not imply that Emperor penguins will not be threatened by global warming. As
2 we pointed out earlier (see §84 p. 174), the impacts of climate change have multiple aspects, that
3 go well beyond the local changes in icescape. Changes in the trophic web of the Southern Ocean
4 and an overall decrease in productivity (due to changes in upwelling activity on the Polar Front,
5 to acidification, and most importantly to sheer temperature effects on plankton communities, see
6 *e.g.* Beaugrand *et al.* 2002) may have a much deeper impact on the Emperor penguin population
7 than changes in polynya distribution. However, addressing these changes will require very differ-
8 ent approaches, such as ecosystem modelling - a central topic, but unfortunately beyond the
9 scope of this work.

10 *Supporting information*

11 *S-0: Main concepts in use*

12 *§-123 Choice and definition of a «synnome».* For the lack of an appropriate term that would be
13 currently in use, we chose to designate by the word *synnome* the particular colonial system that
14 can be observed in the Emperor penguin, and possibly in other species. We derive this term from
15 the Greek σύννομος, «*a common grazing of flocks*», used in particular for gathering flocks of birds
16 (*e.g.* Aristophanes' *Aves*, v. 1756-7), and which was commonly used by extension for «*reunions*»,
17 «*gatherings*», and even «*kindred*». These different meanings together convey the particularity of
18 the observed structure: *one single, nearly homogeneous pool of individuals is distributed in a highly*
19 *discrete way throughout its range. Local concentrations, or colonies, are highly consistent on the scale of*
20 *a few generations, at which scale philopatry may be the norm. Yet, migration between these areas is*
21 *high enough to maintain total homogeneity of the species' gene pool, so that, viewed on a micro-evolu-*
22 *tionary time-scale, the only relevant unit is the species as a whole.*

23 *§-124 Dispersal and migration.* **Dispersal** was originally described by Howard (Howard 1960) as
24 «*the movement the animal makes from its point of origin to the place where it reproduces*». From each

The Emperor synnome - §125

1 individual's perspective, it is «*the greatest distance its genetic characteristics are transmitted, rather*
2 *than the greatest distance the animal may have migrated or otherwise travelled away from the place it*
3 *was conceived, hatched or born*». **Migration**, on the other hand, has been described by Dingle and
4 Drake (Dingle & Drake 2007) in a biogeographical context as «*range expansions of faunas or indi-*
5 *vidual species*», such as «*the northward extension of ranges following the retreat of glaciers at the end*
6 *of the ice ages*». More specifically, in a population genetics context, the (mutation-scaled) migra-
7 tion parameter M has been defined by the same authors as «*the exchange of genes among popula-*
8 *tions by whatever means, including but not limited to migration as we consider it here*» (Dingle &
9 Drake 2007). It is used in that sense in the coalescent framework, in particular by Beerli and col-
10 leagues (Beerli 2004; Beerli & Palczewski 2010). Thus, in that context, migration is distinguished
11 from dispersal by its larger scale: whereas dispersal is an individual- and generation-centred phe-
12 nomenon that may be observed directly, migration is a time-averaged, population-centred event
13 that is only detectable through indirect methods, such as gene flow reconstruction.

14 *S-1: Supplementary methods and analysis*

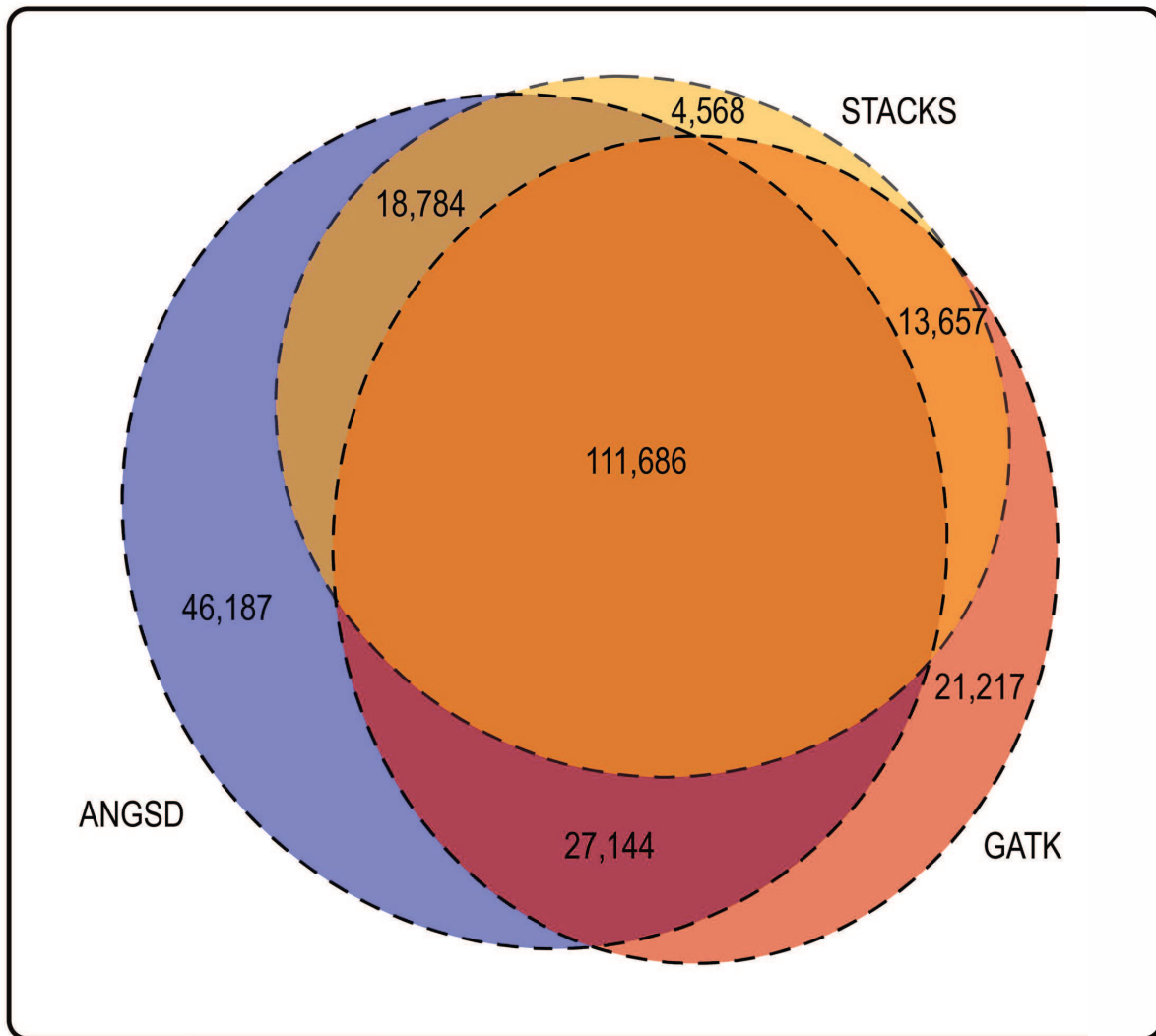
15 *§-125 Sample collection and DNA extraction.* Material was collected from six locations of East
16 Antarctica, between 2004 and 2012: Cape Washington in the Ross sea area («WSH»), Pointe
17 Géologie («DDU») in Adélie Land, Mertz East («MZE») and Mertz West («MZW») in George
18 V Land, Atka bay («NEU») in Dronning Maud Land area, and Halley Bay («HAL») in the
19 Weddell Sea area. On the DDU colony, blood samples were collected from 23 chicks prior to
20 fledging. In WSH, MZE, MZW, NEU and HAL, muscle samples were collected from frozen
21 chick carcasses collected around the colony (WSH: N=4, MZE: N=25, MZW: N=14, NEU:
22 N=24, HAL: N=20). DNA was extracted using a spin-column protocol (Qiagen DNEasy©
23 Blood and Tissue kit) with minor modifications.

The Emperor synnome - §126

1 *§-126 Genome-wide Single Nucleotide Polymorphism (SNP) typing.* SNP discovery and sequencing
2 followed a single-digest RAD-sequencing protocol (Baird *et al.* 2008). Genomic DNA was
3 checked for degradation on a 1.5% agarose gel, and only samples with consistently high molecu-
4 lar weight were retained and quantified by fluorometry (Life technologies™ Qubit®). A total of
5 110 samples was retained and sequenced in 5 distinct libraries. (i) approximately 150 ng of ge-
6 nomic DNA per sample were digested with the restriction enzyme *Sbf*-I-HF (NEB); (ii) each
7 sample was then ligated to a unique barcoded P1 adapter prior to pooling in a single library. The
8 library was then sheared by sonication (7 cycles 30" ON – 30" OFF); (iii) sonicated libraries
9 were concentrated to 25 µl by DNA capture on magnetic beads (beads solution:DNA = 0.8:1),
10 thus further reducing the carry-over of non-ligated P1 adapters, and the target size range fraction
11 (350-650 bp) was then selected by automated gel electrophoresis (BluePippin®); (iv) capture on
12 magnetic beads using the same beads:DNA ratio (0.8:1) was then employed in all following
13 purification steps (after blunt-end repairing, poly-A tailing, P2 adapter ligation and library en-
14 richment by PCR). Magnetic beads were kept together with the library throughout the pre-PCR
15 steps, and DNA was re-bound to the beads for purification using a PEG-8000 binding solution;
16 (v) PCR amplification was performed in 8 x 12.5 µl aliquots pooled after the amplification in or-
17 der to reduce amplification bias on few loci due to random drift. PCR was performed using NEB
18 Phusion® polymerase with the following cycles: 30" denaturation at 98°C, 18 cycles of amplifica-
19 tion (10" at 98°C, 30" at 65°C, and 30" at 72°C), and a final elongation of 5' at 72°C; (vi) the li-
20 brary was then quantified by a fluorimetry-based method (Life technologies™ Qubit®), and mo-
21 larity was checked on an Agilent Bioanalyzer chip (Invitrogen™). A final volume of 20 µl for each
22 library was submitted for paired-end sequencing on an Illumina HiSeq2000 sequencer (V3
23 chemistry, libraries 1-3), or HiSeq2500 (V4 chemistry, libraries 4-5), at the Norwegian Sequenc-
24 ing Centre, University of Oslo, spiked with 20% PhiX control library in order to reduce low-di-
25 versity bias.

1 *§-127 Sequence alignment and genotyping.* Data processing was performed using the following
2 workflow: (i) **Sequence demultiplexing.** Read quality assessment was made in FastQC ([http://](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
3 www.bioinformatics.babraham.ac.uk/projects/fastqc/). Samples were de-multiplexed ac-
4 cording to in-line barcodes using Stacks v1.20 (Catchen *et al.* 2011; Catchen *et al.* 2013), low-
5 quality reads were discarded, and sequences trimmed to 95 bp. (ii) **Read mapping and filtering.**
6 Demultiplexed fastq files were mapped to the published contigs of the Emperor penguin genome
7 (Zhang *et al.* 2011b) using Bowtie2 2.2.3 (Langmead & Salzberg 2012), with standard settings,
8 allowing only end-to-end mapping. Resulting SAM files were filtered using Samtools 0.1.19 (Li
9 *et al.* 2009), PicardTools 1.113 (<http://picard.sourceforge.net>), and custom R and shell scripts in
10 order to discard unpaired reads and full read pairs where at least one mate has a mapping quality
11 score below 30. The resulting BAM files were then filtered for PCR and optical duplicates by
12 comparing mapping position and CIGAR string, using Picard MarkDuplicates. This process also
13 allowed to filter out most sequencing errors, since MarkDuplicates only retains the read with the
14 highest average Phred score in each duplicate cluster. (iii) **SNP calling and genotyping.** Three in-
15 dependent algorithms were used for SNP and genotype calling, all of them built on a maximum-
16 likelihood framework. 1) First-in-pair reads were exported as BAM files for the maximum-likeli-
17 hood-based genotype caller built in the Stacks pipeline (`ref_map.pl`), with a maximum of 5 mis-
18 matches allowed between alleles at a single locus (both within and between individuals), correct-
19 ing genotype calls using the information from the whole dataset in the `rxstacks` program. 2) We
20 used the same cleaned BAM files to simultaneously call both mismatch and indel polymorphisms
21 in all samples using the GATK HaplotypeCaller pipeline (DePristo *et al.* 2011), with standard
22 parameters, except for population heterozygosity which was set to 0.01. 3) ANGSD 0.900 (Kim
23 *et al.* 2011; Nielsen *et al.* 2012) was used to call SNPs and genotypes using a maximum-likeli-
24 hood process with the Samtools `mpileup/bcftools` algorithm, using the complete sample allele fre-
25 quency information as a prior. Genotype calls were only retained for comparison purposes with
26 Stacks and GATK, however, downstream analysis was performed directly on raw genotype likeli-

1 hoods, as this approach has been shown to be much more sensitive to weak structure than classi-
2 cal SNP-calling analysis (Nielsen *et al.* 2011; Fumagalli *et al.* 2013). Genotype calls from all
3 three processes were formatted into VCF format using software-specific (Stacks and GATK) or
4 custom scripts (ANGSD); further filtering and manipulation was done using VCFtools 0.1.12
5 (Danecek *et al.* 2011). For analyses relying on SNP calls, an additional filtering step was per-
6 formed by a) extracting the list of consensus loci called as SNPs by all three independent algo-
7 rithms, b) discarding genotypes with a coverage under 3X for analyses sensitive to sequencing de-
8 pth, c) removing loci genotyped in less than 75% of all individuals, and finally d) thinning down
9 loci to keep only polymorphism distant of at least 1 kb, in order to minimise linkage between
10 markers. (iv) **Sex-linked marker analysis.** In order to check for sex-specific dispersal or structure
11 patterns, we repeated analyses for each sex separately, using either autosomal loci only or Z-
12 linked loci only. We used the published genome annotations for bird gametologs (Zhou *et al.*
13 2014) to identify potentially Z-linked scaffolds. Since females are heterogametic in birds, we ex-
14 pect non-pseudo-autosomal Z-linked regions to be fully homozygous in females, but to neutrally
15 heterozygous in males. We therefore retained only scaffolds that had a clearly bimodal distribu-
16 tion of heterozygosity, with one mode at or close to 0 (a slight tolerance was allowed to account
17 for misalignment, sequencing errors mistyped as SNPs due to low coverage, or presence of a tran-
18 sposable elements with autosomal homologs). Fifteen scaffolds were ultimately retained. Scaffolds
19 containing candidate Z-linked coding DNA sequences, but with no visible bimodal heterozygosi-
20 ty distribution, were excluded altogether from the analysis. Sex assignment was performed inde-
21 pendently from each of the non-recombining Z scaffolds, and consistency of sex calls between
22 scaffolds was checked manually.



1 **Figure 35 | Consensus SNP calling.** Venn diagram of SNP calls by three different algorithms: GATK,
 2 ANGSD with Samtools model, and Stacks.

3 *§-128 RAD data description.* Each HiSeq sequencing lane yielded an approximate 201,000,000
 4 paired-end reads ($\pm 16,000$) with a mean Phred score of 37, only part of which was dedicated to
 5 that project. After barcode demultiplexing and quality filtering, we retained an average
 6 147,000,000 read pairs per library ($\pm 13,000,000$). Concordant alignment rate was high (70.7%
 7 $\pm 9.1\%$). However, after filtering, a large proportion of the reads was identified as duplicates and
 8 removed from further analysis. On average, 479,000 read pairs were retained per individuals. We
 9 built an average 78,000 loci per individual ($\pm 17,000$). Overlap between SNP calling methods was

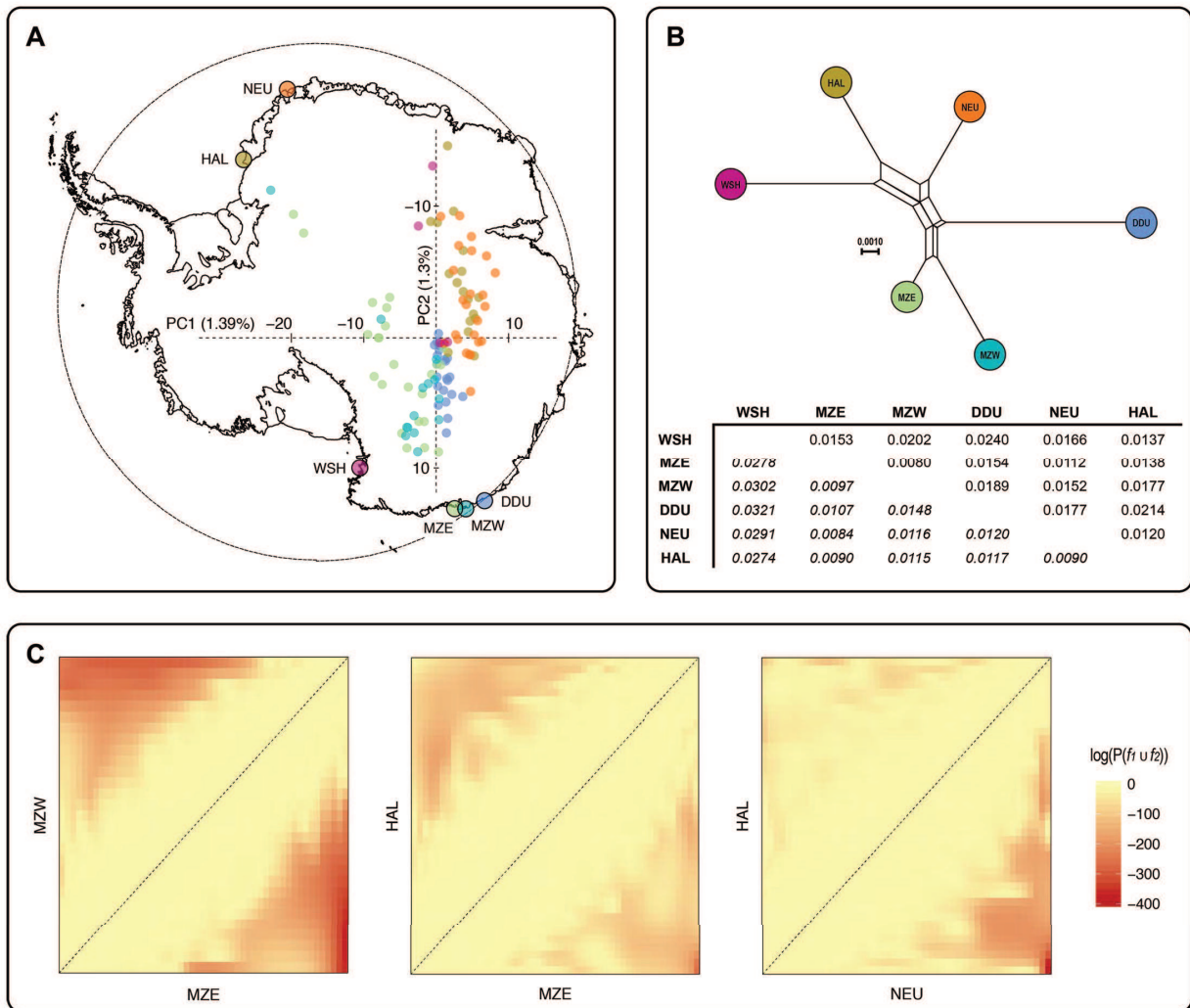
The Emperor synnome - §129

1 high: 111,686 SNPs were called by GATK (173,704 SNPs total), ANGSD (203,801 SNPs total),
2 and Stacks (148,721 SNPs total) (Fig. 35). After filtering by coverage (minimum 3x) and missing
3 data (minimum 75% representation), 59,037 highly confident SNPs were retained for analysis,
4 with a mean depth of 6.8x. Of these, 582, spread across 15 scaffolds, were identified as unam-
5 biguously belonging to the non-recombining region of the Z chromosome.

6 *§-129 Pairwise FST, AMOVA, and summary statistics.* Pairwise fixation index was estimated by
7 two independent methods. We calculated Reich's estimator (Reich *et al.* 2009 - which has been
8 shown to be most robust to small and uneven sample sizes, see Willing *et al.* 2012) on called
9 genotypes, using custom R scripts. Without calling genotypes, we calculated Reynolds' estimator
10 (Reynolds *et al.* 1983), taking into account uncertainty in site-allele-frequency, according to the
11 method implemented in ngsTools (Fumagalli *et al.* 2013; Fumagalli *et al.* 2014). Both methods
12 were performed on a per-site basis and averaged over 1-kb non-overlapping sliding windows, as
13 the ratio of the windowed sum of inter-population variance over the windowed sum of total vari-
14 ance. For the genotype-call-free method, only sites with a probability of being variant at least
15 equal to 0.95 were included in the estimation. Nucleotide diversity and Tajima's D were calculat-
16 ed on a per-locus basis. In order to avoid possible biases due to low coverage, we randomly sam-
17 pled one haplotype for each individual, and performed calculations on this haploid subset. Cal-
18 culations were made using *adegenet* and *pegas* packages, as well as custom R scripts. We found
19 that the Hardy-Weinberg equilibrium holds across 466 out of 590 scaffolds with all colonies as-
20 sessed together (out-of-equilibrium scaffolds all have less than 5 SNPs). The mean homozygosity
21 across all individuals is low ($F=0.051 \pm 0.094$) and does not exhibit any inter-colony difference.
22 The nucleotide diversity is low and highly similar across all colonies ($\pi_{DDU}=0.0026$; $\pi_{MZE}=0.0023$;
23 $\pi_{MZW}=0.0023$; $\pi_{NEU}=0.0023$, $\pi_{HAL}=0.0023$; $\pi_{WSH}=0.0026$), going along with the expectation for
24 long-lived high-investment species (Romiguier *et al.* 2014), and despite very different colony cen-
25 sus sizes. Finally, Tajima's D averaged across the whole genome does not deviate from neutral ex-
26 pectations ($D_{DDU}= -0.87$; $D_{MZE}= -0.66$; $D_{MZW}= -0.578$; $D_{NEU}= -0.64$; $D_{HAL}= -0.68$; $D_{WSH}= -0.28$).

The Emperor synnome - §129

- 1 We used SplitsTree (Huson & Bryant 2006) to build a neighbour-net based on pairwise F_{st} dis-
- 2 tances between colonies. In order to quantify the proportion of variation at each organisation lev-
- 3 el, AMOVA was performed on unlinked SNP data (§131 p. 245) using Arlequin (Excoffier *et al.*
- 4 2005), on a per-locus basis, with 1,000 permutations.



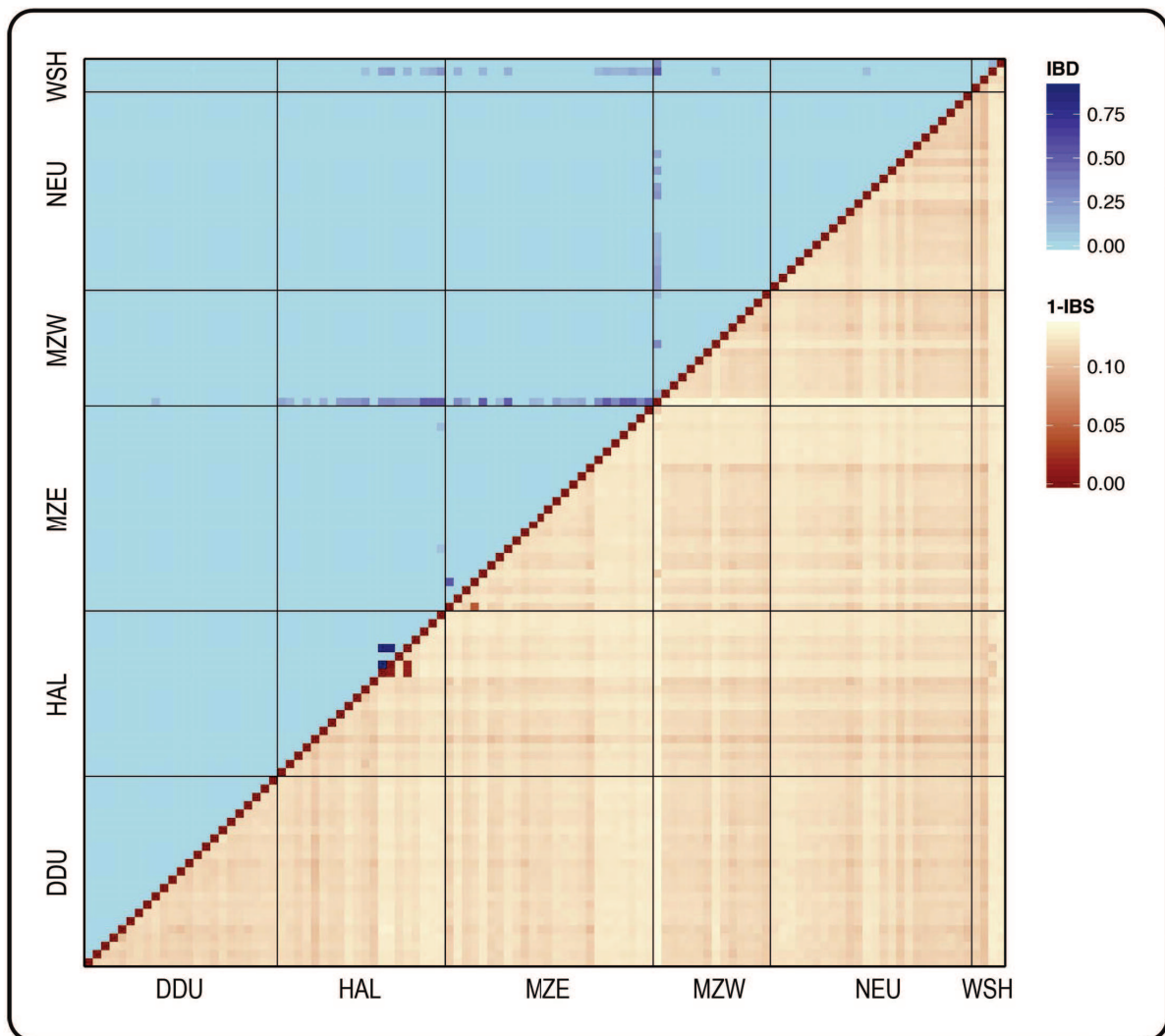
- 5 **Figure 36 | The imprint of geography on the Emperor synnome.** (A) Principal component analysis
- 6 (percentage of explained variance along the axes) - the colours on the plot refer to the colours used on the colony loca-
- 7 tions. (B) NeighbourNet built from F_{st} values according to Reich's estimator (table: upper triangle, point esti-
- 8 mates, lower triangle, standard deviation). (C) Three examples of pairwise derived-allele frequency spectra, between
- 9 two nearby South Pacific colonies (MZE and MZW), two distant colonies (HAL and MZE), and two South Atlantic
- 10 colonies (HAL and NEU).

The Emperor synnome - §130

1 **Table 6 | Pairwise Fst, according to Reich's estimator.** *Upper triangle: mean value, lower triangle: standard*
 2 *deviation.* DDU: "Dumont d'Urville", HAL: "Halley Bay", MZE: "Mertz East", MZW: "Mertz West", NEU: "Neu-
 3 mayer", WSH: "Cape Washington".

	DDU	HAL	MZE	MZW	NEU	WSH
4 DDU		0,0214	0,0154	0,0189	0,0177	0,0240
5 HAL	<i>0,0117</i>		0,0138	0,0177	0,0120	0,0137
6 MZE	<i>0,0107</i>	<i>0,0090</i>		0,0080	0,0112	0,0153
7 MZW	<i>0,0148</i>	<i>0,0115</i>	<i>0,0097</i>		0,0152	0,0202
8 NEU	<i>0,0120</i>	<i>0,0090</i>	<i>0,0084</i>	<i>0,0116</i>		0,0166
9 WSH	<i>0,0321</i>	<i>0,0274</i>	<i>0,0278</i>	<i>0,0302</i>	<i>0,0291</i>	

11 *§-130 Identity-by-state (IBS) and identity-by-descent (IBD).* Pairwise indicators of IBS and IBD
 12 were calculated in PLINK v1.9 (Purcell *et al.* 2007). Allelic distance for pairs of individuals is
 13 rather sensitive to uneven coverage (Fig. 37, lower triangle), with lower-depth individuals appear-
 14 ing more similar to all other individuals than higher-depth ones. However, IBD inference, which
 15 takes into account the total genetic variance in the whole sample, appears more robust to cover-
 16 age variation (Fig. 37, upper triangle). Three individuals in the HAL colony (HAL13, HAL14
 17 and HAL16, visible on Fig. 37 and Fig. 39), as well as two in the MZE colony (MZE01 and
 18 MZE04, visible on Fig. 37 and Fig. 39), appear markedly more related to each other than to the
 19 rest of the sample. Of these two clusters, only one individual was retained for inbreeding-sensi-
 20 tive analyses, such as principal component analysis. Pairwise IBS distances were also used to gen-
 21 erate a neighbour-net using SplitsTree (Huson & Bryant 2006).



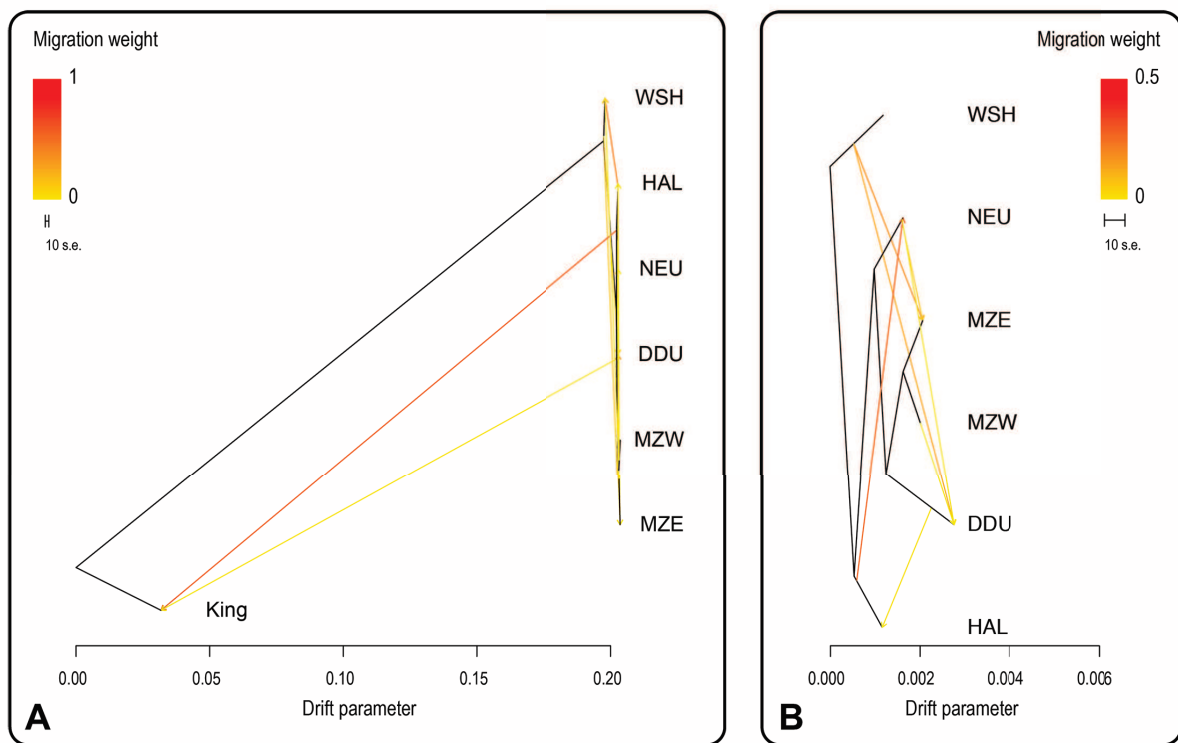
1 **Figure 37 | Pairwise genetic distance matrix.** Upper triangle: identity-by-descent inference (IBD). Lower
 2 triangle: identity-by-state pairwise (IBS).

3 *§-131 Principal Component Analysis (PCA).* PCA was performed on called genotypes using R li-
 4 brary adegenet (Jombart 2008), and on genotype likelihoods using ngsCovar (Fumagalli *et al.*
 5 2014). Analysis was performed either on all samples or keeping only one individual for each
 6 highly related cluster. When performed on all samples, analysis was mostly driven by the HAL re-
 7 latedness cluster, with that first PC accounting for 2.84% of total variance. Interestingly, these
 8 outliers are also identified as a separate group in admixture analyses (see below, § 132 p. 246),
 9 with high repeatability, although with only a slight gain in model fit compared to a single-popu-

The Emperor synnyme - §132

1 lation model (Evanno's ΔK over 10 replicates). However, discriminant analysis of principal com-
2 ponents (Jombart 2008) performed in adegenet, contrasting these outliers and the main sample
3 group, showed that single-locus contributions were very low and evenly spread across the
4 genome, thus excluding any strong directional selective phenomenon. After removal of all sam-
5 ples but one in each high-relatedness cluster, variation did not appear to be driven by outliers
6 anymore. Overall, the variance explained by geographical structure is extremely low. After outlier
7 removal, the first PCA component of variance shows a weak general pattern of isolation-by-dis-
8 tance along the coast. Mertz colonies stand on the one end, Pointe Géologie in the centre, and
9 Atka Bay and Halley Bay at the other end, yet Ross sea samples do not stand out, as would be ex-
10 pected in case of strong, systematic distance-driven drift. However this pattern remains extremely
11 marginal: the variance explained by the first component barely amounts to 1.4% of the total vari-
12 ance, and colonies can not be distinguished on the basis of PCA.

13 *§-132 Population splits and migration topology.* We used TreeMix (Pickrell & Pritchard 2012) to
14 infer the topology of population splits and migrations from allele frequency variation among
15 colonies. We produced RAD-sequencing data for the Emperor's sister species, the King penguin
16 *Aptenodytes patagonicus*, and processed it with the same protocol as for our Emperor penguin
17 samples. We genotyped 15 King penguins from the Baie du Marin colony, on Crozet archipelago,
18 that were used as an outgroup for the analysis. We restricted our dataset to ca. 15,000 highly con-
19 fident, unlinked SNPs shared between the two species. 3-population and 4-population tests did
20 not allow us to reject a tree-like topology in our data. Analysis was performed first using the King
21 penguin outgroup, and bootstrapping over blocks of 500 SNPs, in order to assess the topology of
22 relationships between our 6 Emperor penguin colonies. We used this topology to re-run TreeMix
23 without the King penguin samples, but fixing the root according to our first results, in order to
24 increase the resolution of the analysis (Fig. 38).



1 **Figure 38 | Inference of population splits.** Population tree reconstruction based on the *TreeMix* model. (A)
 2 Including polymorphism data from the King penguin as an outgroup. (B) Without the King penguin and fixing the
 3 root as the same position as inferred in A.

4 *§-133 Clustering analysis.* Clustering analyses were performed both on called genotypes, using
 5 fastStructure (Raj *et al.* 2014), and on genotype likelihoods, using ngsAdmix (Skotte *et al.*
 6 2013). For both algorithms, only consensus polymorphic loci were retained. We tested a number
 7 of components ranging from 1 to 10, with 3 independent replicates. For fastStructure, we used a
 8 simple prior for K values ranging from 1 to 10, and a logistic prior for values from 1 to 3. Most
 9 likely number of components was chosen using fastStructure’s chooseK.py script for fastStructure
 10 models, and Evanno’s ΔK method (Evanno *et al.* 2005) for ngsAdmix models. For both algo-
 11 rithms, the preferred model consistently had $K=1$, i.e. no inferred clusters in the data.

12 *§-134 Coalescent-based analysis: Migrate-n.* All methods used above provide a robust framework
 13 for identifying groups even in weakly structured populations. However, all are, to some extent,

1 functions of the covariance of allele frequencies between populations. In the hypothesis that our
2 estimate of the frequency spectrum may be biased in some way, e.g. by moderate sample size, we
3 also performed a coalescent-based structure analysis using Migrate-n (Beerli & Palczewski
4 2010). As opposed to frequency-spectrum based approaches, coalescent-based analysis relies on
5 phased polymorphisms to infer population parameters. After verifying that the number of
6 polymorphisms in each RAD locus followed a Poisson distribution of λ equal to the mean num-
7 ber of polymorphisms per locus, we selected 50 random loci comprising between 5 and 6
8 polymorphisms as an unbiased representation of the neutrally evolving part of the genome (Truc-
9 chi *et al.* 2014). In order to correct for potential over-representation of sequences in case a het-
10 erozygous individual was mis-called as homozygous, we randomly sampled one allele only for
11 each individual. For all colonies except WSH, we randomly picked a set of haplotypes in the
12 population (16 in MZE, DDU, HAL, and NEU, 14 in MZW, and 4 in WSH). We ran a cold
13 chain and 3 heated chain of 50,000,000 generations, recording every 500 generations, with a
14 5,000,000-generation burn-in. We used a static heating scheme, raising the cold chain to a power
15 of 1.5, 3 and 1e6, and proposing chain swapping every 100 steps. We used a uniform prior for
16 population sizes (Θ), bounded between 0 and 0.1 (with a δ of 0.01), and for the migration rates
17 (M), bounded at 4,000 with a δ of 400. Proper mixing under these conditions was ensured using
18 the highest parametrisation model (model 2). We compared 5 models of increasing complexity:
19 1) a panmictic model, in which all colonies were gathered in one population, estimating only a
20 general Θ , 2) a full-matrix model, in which asymmetric migrations were allowed between all pairs
21 of colonies, 3) a step-wise model, in which asymmetric migrations were allowed, but only be-
22 tween neighbouring colonies, in a closed circle, 4) a first meta-population model, in which the
23 Ross Sea (WSH), Adélie and George V Lands (MZE, MZW and DDU), and the South Atlantic
24 (NEU and HAL) were treated as three populations, with an asymmetric migrations between
25 them, 5) a second meta-population model, in which the northernmost colonies (MZE, MZW,

1 DDU, and NEU) were separated from the Weddell Sea (HAL) and the Ross Sea (WSH). Models
 2 were ordered by log Bayes factor defined by Kass and Raftery (Kass & Raftery 1995) as $\ln BF =$
 3 $2[\ln(mL(model1)) - \ln(mL(model2))]$, with $mL(model1)$ and $mL(model2)$ being the marginal
 4 likelihoods for the two compared models, as calculated by thermodynamic integration.

5 Best support goes to the full migration matrix model (model 2). Estimated migration rates M
 6 and population sizes Θ are highly homogeneous (mean mutation-scaled migration rate $m / \mu =$
 7 $2,358 \pm 130$, mutation-scaled effective sizes $\Theta_{WSH}=0.0017$, $\Theta_{MZE}=0.0018$, $\Theta_{MZW}=0.0019$,
 8 $\Theta_{DDU}=0.0019$, $\Theta_{NEU}=0.0018$, $\Theta_{HAL}=0.0019$).

9 Under this model, each migration rate can also be expressed as a proportion of the receiving pop-
 10 ulation's effective size as $m = 4.M.\mu$. Hence, the relative demographic importance of immigration
 11 for any given colony can also be expressed as $\Sigma_{i=n}(4.Mi .\mu)$, with n being the total number of
 12 populations identified as gene sources for the focal population. In order to convert mutation-
 13 scaled estimates of M and Θ , we therefore need an estimate of μ (the mean number of substitu-
 14 tions.site⁻¹.Myr⁻¹) for the set of loci used in the analysis. Since the number of sites in each locus is
 15 mainly due to the stochastic nature of the mutational process, and follows a Poisson distribution
 16 of parameter λ equal to the mean SNP density per locus (see above), we can consider that a single
 17 true mutation rate (μ) applies to the whole RADome. However, since our analysis is restricted to
 18 loci containing 5 to 6 SNPs, our estimates of Θ and M are not directly scaled by μ , but rather by
 19 posterior probabilities of μ conditional on the number of SNPs in each locus class. We used the
 20 class-specific mutation rate posterior probabilities as calculated for the King penguin (Trucchi *et*
 21 *al.* 2014). As a conservative estimate, we used a range of rates fitting the 3-SNPs to 6-SNPs class
 22 loci. Using a generation time of 16 years (Jenouvrier *et al.* 2014), these are $\mu_{3snps} = 1.14e^{-6}$ and
 23 $\mu_{6snps} = 2.16e^{-6}$. Since these were estimated from a subset of 16 haplotypes, and we included 82
 24 haplotypes in our analysis, we considered that variability was likely to be underestimated by Truc-

1 chi and colleagues (Trucchi *et al.* 2014) compared to our sample. Thus, we did not consider
2 higher estimates than those made for 6-SNPs loci.

3 *§-135 fastsimcoal2 analysis.* Joint derived-allele frequency spectra were generated in ANGSD
4 0.900 (Kim *et al.* 2011; Nielsen *et al.* 2012) for a subset of four colonies that encompass the
5 whole continent (MZE, DDU, NEU and HAL). In order to polarize these spectra, we recon-
6 structed the most likely ancestral base for all positions in the RADome. We selected 12 high-
7 quality Emperor penguin samples covering the whole species' range, as well as 12 high quality
8 King penguin samples processed according to the same protocol (see *The King synnome*, p. 163).
9 States at all positions were determined using GATK's Haplotype Caller pipeline (DePristo *et al.*
10 2011). We used BEDtools (Quinlan & Hall 2010) and VCFtools' vcf-consensus script (Danecek
11 *et al.* 2011) to update the published Emperor penguin genome and establish a reference
12 RADome for both the King penguin, and the Emperor penguin, using only high-quality
13 polymorphisms (phred-scale genotype quality ≥ 80), and including variable sites as IUPAC ambi-
14 guity codes. We aligned this RADome to the Adélie penguin genome (*Pygoscelis adeliae*, see
15 Zhang *et al.* 2011b) using Bowtie2 (Langmead & Salzberg 2012), and extracted the correspond-
16 ing regions. For each RAD locus, a maximum-likelihood unrooted tree was built in PhyML
17 (Guindon *et al.* 2010), and maximum-likelihood ancestral sequence for crown-*Aptenodytes* was
18 reconstructed using PAML (Yang 2007) and Lazarus (<https://project-lazarus.googlecode.com/>),
19 using PhyML tree topology as a prior. These ancestral states were then used to determine the an-
20 cestral and derived alleles in the Emperor penguin.

21 Reconstruction of population sizes and migration events was performed through composite-like-
22 likelihood maximisation, by simulating joint-spectra under a continuous-time Markovian coalescent
23 model in fastsimcoal2.5.11 (Excoffier & Foll 2011; Excoffier *et al.* 2013). For each run, we per-
24 formed a maximum of 80 ECM optimisation cycles over the 12 retained parameters (population
25 sizes and asymmetric migration rates between the four analysed colonies), each parameter optimi-

1 sation step requiring the generation of 100,000 simulated joint-spectra. We generated 50 non-
2 parametric bootstrap replicates for each spectrum. For each bootstrap dataset, and for the original
3 dataset, we ran 50 independent replicates, and retained the one with the highest log-likelihood.
4 We assumed a mutation rate of $2.6e-7$ substitutions.site-1.generation-1 as calculated for the King
5 penguin (Trucchi *et al.* 2014), and a generation time of 16 years (Jenouvrier *et al.* 2014). Com-
6 putation was performed on the high-performance Abel cluster at the University of Oslo, and re-
7 quired a total of ~30,000 CPU-hours.

8 We chose to restrict our analysis to a stepwise migration model for two main reasons. First, com-
9 putational load increases rapidly with the number of estimated parameters, and higher complexi-
10 ty models could not be run with the necessary amount of replication. Secondly, since our model
11 is not supposed to represent precisely the present-day state of the colonies, but rather parameters
12 averaged over a long period of time, we do not expect the intensity of the migration flow to be
13 much affected by the structure of the connectivity. However, in one particular case, the estimate
14 seems to deviate from the general pattern of migration in our analysis, and may be indicative of a
15 very recent event, still directly linked to the present-day states of the colonies. Indeed, the gene
16 flow between DDU and MZE is estimated to be much higher, and much more asymmetrical
17 than the rest of the rates (Fig. 31, Table 7), with almost all of the gene flow being directed from
18 DDU to MZE. Although this may be a circumstantial effect of our sampling, it is noteworthy
19 that the DDU colony has indeed lost nearly half its breeding population in the very recent past,
20 with no definitive explanation (Barbraud & Weimerskirch 2001a). Although it has been suggest-
21 ed that this sudden decrease may have been caused by local environmental effects (Barbraud &
22 Weimerskirch 2001a; Barbraud *et al.* 2011), the disturbance caused by the Dumont D'Urville re-
23 search station, including its research activities on the colony, and in particular the effect of flip-
24 per-banding of the Emperor penguins (Le Maho *et al.* 2011) have also been mentioned as poten-
25 tial causes for high emigration from the DDU colony. It is thus possible that the high gene flow

The Emperor synnome - §136

1 inferred from DDU to MZE indicates that this decrease was at least partly caused by emigration
 2 to the nearby Mertz glacier area in response to human-induced disturbance, rather than as by en-
 3 vironment-induced mortality.

4 **Table 7 | Migration rates and sizes as estimated from the joint allele frequency spectrum.**

	MEAN	L95	U95
Population sizes and directional gene flow			
N_HAL	4040	3990	4080
N_NEU	3060	2920	3200
N_DDU	1120	1100	1150
N_MZE	7640	6150	9130
mHALNEU	0.0166	0.0153	0.0179
mNEUHAL	0.00323	0.0023	0.00416
mNEUDDU	0.0391	0.0376	0.0407
mDDUNEU	0.016	0.014	0.018
mDDUMZE	0.0416	0.0379	0.0452
mMZEDDU	0.000868	0.000372	0.00136
mMZEHAL	0.00397	0.00327	0.00466
mHALMZE	0.000413	0.000283	0.000544
Immigration rate (ratio of receiving population)			
HAL	0.0072	0.00557	0.00882
NEU	0.0326	0.0293	0.0359
DDU	0.039968	0.037972	0.04206
MZE	0.042013	0.038183	0.045744
Immigration size per receiving population (in effective breeders)			
HAL	29	22	36
NEU	132	117	146
DDU	161	152	172
MZE	170	152	187

31 *§-136 Coalescent-based analysis: BEAST2.* An independent estimate of population size changes
 32 through time was performed in BEAST2 (Bouckaert *et al.* 2014). We proceeded as above for lo-
 33 cus selection and haplotype downsampling. In order to remain agnostic as to population struc-
 34 ture, we performed analysis on each colony separately, as well as on the whole dataset, as the lack

1 of strong population structure, evidenced by all other analyses, allowed us to sample haplotypes
2 from the whole species without violating the model's assumptions. We used an extended bayesian
3 skyline plot model (Heled & Drummond 2008) in order to co-estimate present-day Θ and possi-
4 ble past fluctuations. We followed the protocol of Trucchi and colleagues (Trucchi *et al.* 2014),
5 but reduced the parameter space by defining only one site-model per locus class (5 or 6 SNPs),
6 using HKY models with empirical base frequencies, and allowing for rate variation in 4 discrete
7 gamma categories. Kappa was linked across site models, according to our expectation for neutral
8 variation, in order to alleviate computational load (A. Drummond, *personal communication*). All
9 chains were run in duplicate to check for convergence and for a sufficient length to gather ESS >
10 200 for all parameters, which necessitated ca. 1,000,000,000 steps on all models. Reconstruction
11 for WSH colony is much less precise due to the very small number of haplotypes (N=4) sampled
12 per locus. However, present-day population size estimate converges with reconstructions based
13 on the other colonies. It is also noteworthy that our results point to a gradual increase in popula-
14 tion size over the last ~100,000 years (with some uncertainty as to the precise dating of the be-
15 ginning of the expansion, due to the difficulty of precisely calibrating mutation rates in a multilo-
16 cus approach), which confirms the findings of Li and colleagues (Li *et al.* 2014) based on a
17 single-genome pairwise sequentially Markovian coalescent approach.

18 *§-137 mtDNA analysis.* A recent study by Younger and colleagues (Younger *et al.* 2015) focused
19 on Emperor penguin mitochondrial DNA population structure. Their conclusion was that
20 colonies from the Ross sea area are significantly isolated from the rest of the continent, and had a
21 different demographic history. Our genome-wide SNP data does not support this view: however,
22 our low sampling size in the Ross Sea region does not permit any definitive conclusion. In order
23 to assess how far this result could be reproduced from mtDNA alone, we sequenced a 792 bp
24 fragment of mitochondrial cytochrome-B gene and a 414 bp-long HVR fragment for our RAD
25 samples, following the same protocol as Younger et al. Primers were the following: Cyt-b forward

The Emperor synnome - §137

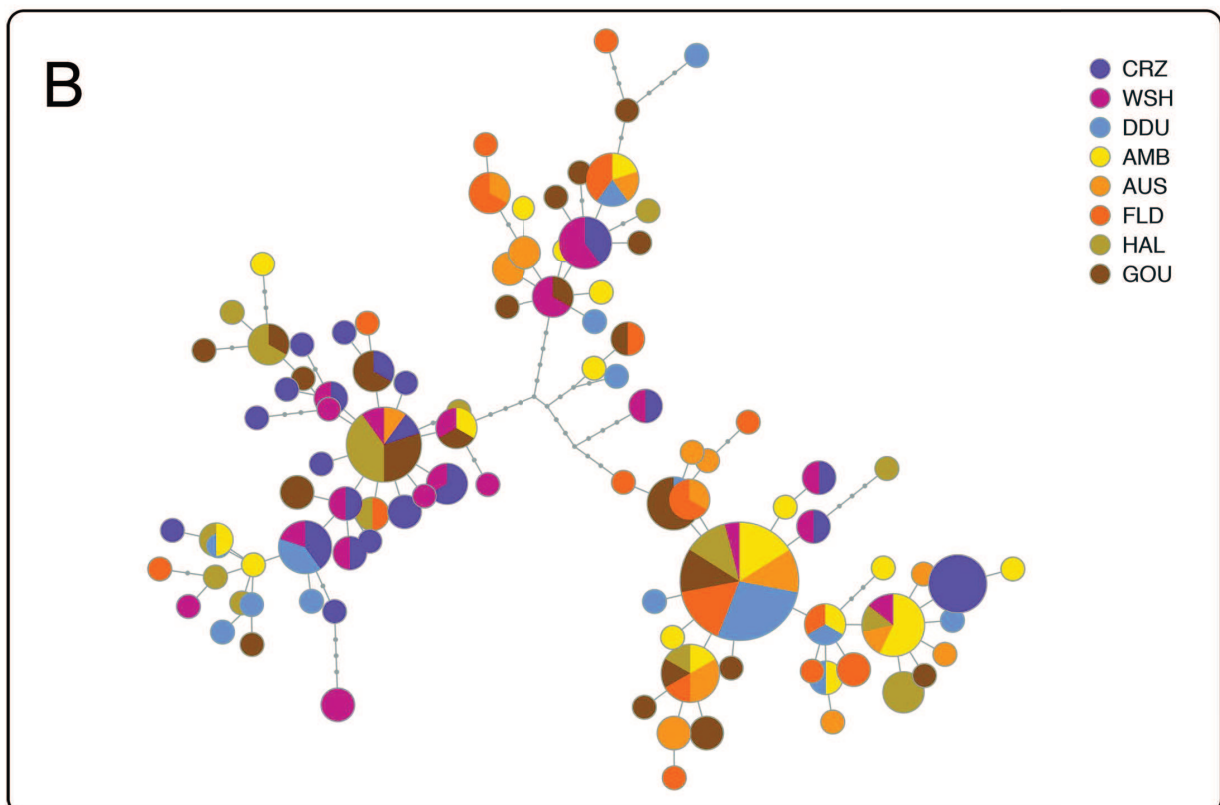
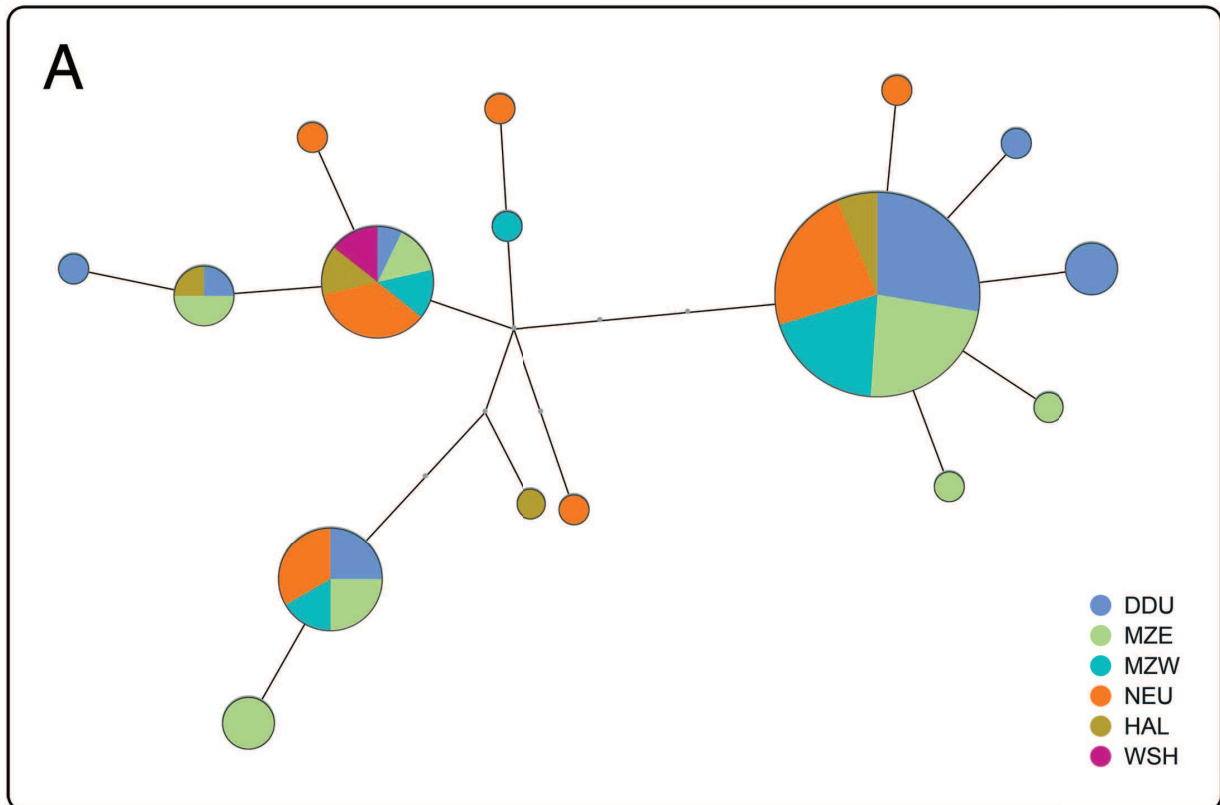
1 primer 5'-GCCCCAAACCTCCGAAAATCCCA-3' and reverse primer 5'-TGTGGAG-
2 GAGGGGGATTAGG-3'; HVR forward primer 5'-GGAACCTCCCAAAGAGTACCA-3' and
3 reverse primer 5'-CCAACCAGATGTATCGGTGA-3'. PCR conditions were thus: for Cyt-b, 5'
4 denaturation at 94°C, 35 cycles of amplification (30" denaturation at 94°C, 30" annealing at
5 57°C, and 1' elongation at 72°C), and 5' final elongation at 72°C; for HVR, 5' denaturation at
6 94°C, 35 cycles of amplification (30" denaturation at 94°C, 30" annealing at 59.5°C, and 1'
7 elongation at 72°C), and 5' final elongation at 72°C. PCR product was then purified using Illus-
8 tra™ ExoStar™, and Sanger sequencing was performed at the ABI lab of the University of Oslo.
9 Quality assessment, trimming, and manual checking were performed in Geneious® v6.1.2. Un-
10 fortunately, our HVR sequences were of consistently low quality and thus could not be used reli-
11 ably in analysis. Many double peaks were observed, possibly explained by HVR duplication (as
12 previously observed in *Thalassarche* albatrosses, see Abbott *et al.* 2005), together with possible in-
13 stances of polynucleotide repeat number variation. The uncertainties regarding base calls and site
14 phasing in the case of double peaks made it impossible to extract any reliable data from these
15 sequences.

16 We also re-analysed the data from Younger and colleagues (GenBank accession numbers
17 KP644787-KP645015 and KP640645-KP640873). Considering no reliable and controlled mod-
18 el can account for diploid sites in mitochondrial DNA, we masked all ambiguous sites from
19 analysis. We also considered that the higher reported similarity between Ross Sea mtDNA sam-
20 ples may be biased by the fact that, after masking ambiguous sites, two haplotypes are overly rep-
21 resented compared to the average haplotype diversity. Indeed, when concatenating HVR and Cy-
22 tochrome B sequences for each individual, and putting aside the two over-represented sequences,
23 each mitochondrial haplotype is present in an average 1.4 ± 0.9 copies per colony. The two over-
24 represented haplotypes, on the other hand, are found in 14 samples from Cape Washington, and
25 15 from Cape Crozier. Not surprisingly, however, these samples are the only “shed feathers [...]”
26 collected from the Ross Sea between 2010 and 2012 [...] at least 10 m apart to minimize sam-

1 pling the same bird” as opposed to blood samples for the other locations. It indeed may seem
2 rather likely that the precaution was not sufficient, and that the same birds were sampled multi-
3 ple times, as it is common for a moulting penguin to shed feathers on a great surface. Once mis-
4 called bases are masked, and potentially pseudo-replicate samples are removed, the resulting pat-
5 tern does not exhibit the clear bipartite organisation found by Younger and colleagues, but rather
6 a gradual differentiation, in keeping with the general pattern we observe in genomic SNP data
7 (Fig. 39B). As a side note, the over-representation of pseudo-replicate sequences in the Ross sea
8 region, by violating the random-sampling assumption of coalescent reconstructions, may also ac-
9 count for the differences in past demographic trends inferences found by the authors (Younger et
10 al. figure 2).

11 Cytochrome-B sequences, on the other hand, showed a standard level of variation in our dataset,
12 as assessed in DnaSP (Rozas & Rozas 1999 - 10 haplotypes, gene diversity=0.634, nucleotide di-
13 versity=0.003, Fu's F_s =-0.948, Tajima's D =-0.216, non-significant). Haplotype network was built
14 based on Fitch distances between sequences, using Fitchi (Matschiner 2015), using a maximum-
15 likelihood bifurcating tree built in RaxML (Stamatakis 2014). In keeping with the results of
16 Younger and colleagues, Cytochrome-B sequences do not reflect geographical distribution of the
17 samples in any way (Fig. 39A).

18 The particular case of the Ross Sea area may require further analysis, as mitochondrial HVR
19 alone seems too unreliably sequenced to provide positive information. The available data only
20 support the extension of a low-level isolation-by-distance model to the whole continent, in keep-
21 ing with our observations on genome-wide polymorphism. However, the intensity of the gene
22 flow between the Ross Sea region and the rest of the continent has little impact on our ability to
23 model immigration rates at the colony level. Indeed, the origin of the immigration flux is less rel-
24 evant than its intensity if we are to accurately model population dynamics from colony-level data.



1 **Figure 39 | Haplotype network for mitochondrial sequences.** Each dot represents a mutation. The areas
 2 of the discs is proportional to the number of copies found in the sample. Coloured sectors indicate the geographical

The Emperor synnome - §137

1 repartition of these haplotypes. (A) Cytochrome B sequences produced for the present study. (B) Cytochrome B and
2 HVR sequences from Younger *et al.* 2015, reanalysed for this study (S-13). Colony labels as follows: CRZ=Cape
3 Crozier, WSH=Cape Washington, DDU=Dumont d'Urville/Pointe Géologie, AMB=Amanda Bay, AUS=Auster,
4 FLD=Fold Island, HAL=Halley Bay, GOU=Gould Bay.

5 *Acknowledgements*

6 This study was undertaken within the framework of the Programme 137 of the Institut Polaire
7 Français Paul-Emile Victor (IPEV), with additional support from the French National Research
8 Agency (ANR) "PICASO" grant (ANR-2010-BLAN-1728-01), from Marie Curie Intra Euro-
9 pean Fellowships (FP7-PEOPLE-IEF-2008, European Commission; project no. 235962 to CLB
10 and FP7-PEOPLE-IEF-2010, European Commission; project no. 252252 to ET), from the Cen-
11 tre Scientifique de Monaco through budget allocated to the Laboratoire International Associé
12 647 'BioSensib' (CSM/CNRS-University of Strasbourg), and from the British Antarctic Survey
13 Ecosystems Programme, NERC (PT). Logistic and field costs of research were supported by
14 IPEV (CLB), the British Antarctic Survey Ecosystems Programme, NERC (PT), AWI (DPZ),
15 US NSF grant number NSF 0229638 (PP). This work was performed on the Abel Cluster,
16 owned by the University of Oslo and the Norwegian metacenter for High Performance Comput-
17 ing (NOTUR), and operated by the Department for Research Computing at USIT, the Universi-
18 ty of Oslo. We are very grateful to Morten Skage, Ave Tooming-Klunderud, Marianne Selander-
19 Hansen, and the Norwegian Sequencing Center for their very valuable help in the laboratory, as
20 well as Lex Nederbragt and Michael Matschiner for their assistance with the Abel cluster, Matteo
21 Fumagalli and Thorfinn Korneliussen for their precious advice regarding ngsTools and ANGSD,
22 and Alexis Drummond for his insights on BEAST2 parametrization. We thank Matthieu
23 Boureau, and the British Antarctic Survey staff at Halley Bay, for the sample collection. Last but
24 not least, we thank Romain Loriol for his help with Greek lexicology.

The Emperor synnome - §137

1 Chapter 6: Empirical evidence of heterogeneous dispersal

2 This chapter has been published as: Ancel A, Cristofari R, Fretwell PT, Trathan PN , Wienecke B,
3 Boureau M, Morinay J, Blanc S, Le Maho Y, Le Bohec C (2014) **Emperors in Hiding: When**
4 **Ice-Breakers and Satellites Complement Each Other in Antarctic Exploration.** PloS one
5 9:e100404

6 *Context*

7 *§-138 How do coalescent-scale migration translate at the ecological scale?* In the previous chapter
8 (*The Emperor synnome*, p. 211), we showed how continent-wide genomic data allows us to recon-
9 struct large-scale migration patterns in the Emperor penguin. Inferred gene flow is outstandingly
10 high, with colonies receiving between ~1% and ~5% of their effective population size as migrants
11 every generation (see §113 p. 220). However, these estimates are averaged over the period
12 spanned by the coalescent history, *i.e.* up to $4.N_e$ generations (see §41 p. 103): their distribution
13 in time may be homogeneous, with a low and constant trickle of migration, or on the other hand
14 highly heterogeneous, with periods of high or complete dispersal alternating with periods of «*sta-*
15 *sis*» and philopatric isolation. Similarly, dispersal may be homogeneous, or heterogeneous in
16 space: the determinants of the strength of dispersal may be global (*e.g.* planetary climate change
17 and modifications of the glacial or oceanographic landscape), and affect every colony equally, or

Empirical evidence of heterogeneous dispersal - §138

1 local (*e.g.* accidental changes in the ice landscape following a particular event) and affect only one
2 or a few colonies at a time.

3 Understanding the structure of migration in time and space at such scales, however, is well be-
4 yond the power of genetic approaches. The particular synnyme structure of the *Aptenodytes* pen-
5 guins (see §3 p. 28) implies that, although colonies as geographical units may persevere for long
6 periods (see §2 p. 25), particular lineages do not reside in the same colony for more than a few
7 generations. Thus, the spatial coordinates of a lineage are to some extent anecdotal: each may be
8 traced back from its current position to the previous one, but thence no further. Thus, events
9 such as the merging of two nearby and barely isolated colonies, temporary interruption of gene
10 flow, or the disappearance of a group dispersing into the neighbouring colonies are beyond the
11 resolution of our analysis methods. Understanding the dynamics of dispersal thus requires direct
12 field observations, grounded in an ecological, rather than an evolutionary framework (see §2 p.
13 25). In the present chapter, we use a direct, observational method, based on ground surveys, aer-
14 ial and satellite photography, and a long-term monitoring program (see §24 p. 79) to shed light
15 upon the possible mechanisms underlying migration in *Aptenodytes* penguins.

16 *Abstract*

17 Evaluating the demographic trends of marine top predators is critical to understanding the
18 processes involved in the ongoing rapid changes in Antarctic ecosystems. However, the remote-
19 ness and logistical complexity of operating in Antarctica, especially during winter, make such an
20 assessment difficult. Satellite imaging is increasingly recognised as a valuable method for remote
21 animal population monitoring, yet its accuracy and reliability are still to be fully evaluated. We
22 report here the first ground visit of an emperor penguin colony first discovered by satellite, but
23 also the discovery of a second one not indicated by satellite survey at that time. Several successive
24 remote surveys in this coastal region of East Antarctica, both before and after sudden local

1 changes, had indeed only identified one colony. These two colonies (with a total of ca. 7,400
2 breeding pairs) are located near the Mertz Glacier in an area that underwent tremendous habitat
3 change after the glacier tongue broke off in February 2010. Our findings therefore suggest that a
4 satellite survey, although offering a major advance since it allows a global imaging of emperor
5 penguin colonies, may miss certain colony locations when challenged by certain features of polar
6 ecosystems, such as snow cover, evolving ice topology, and rapidly changing habitat. Moreover
7 our survey shows that this large seabird has considerable potential for rapid adaptation to sudden
8 habitat loss, as the colony detected in 2009 may have moved and settled on new breeding
9 grounds. Overall, the ability of emperor penguin colonies to relocate following habitat modifica-
10 tion underlines the continued need for a mix of remote sensing and field surveys (aerial photog-
11 raphy and ground counts), especially in the less-frequented parts of Antarctica, to gain reliable
12 knowledge about the population demography and dynamics of this flagship species of the
13 Antarctic ecosystem.

14 *Introduction*

15 The emperor penguin (*Aptenodytes forsteri*), a flightless seabird endemic to Antarctica, was prob-
16 ably first seen on James Cook's second voyage, in 1773–1774. Yet the first breeding colony was
17 only discovered in 1902 during the first of Scott's Discovery Expeditions (Wienecke 2009b). For
18 the colonies that could be reached from research stations, population trends usually involved
19 ground counts (Budd 1962; Peter *et al.* 2008). The census of those not accessible by ground sur-
20 veys was usually obtained using aerial photography (Wilson *et al.* 2001). The first estimation
21 from space was carried out in 2005 and 2006 for the western Ross Sea area (Barber-Meyer *et al.*
22 2007), while the first global survey was performed in 2009 (Fretwell & Trathan 2009; Fretwell *et*
23 *al.* 2012). Today, 52 colonies of emperor penguins have been identified (Fretwell and Trathan,
24 unpublished data), all of them being distributed along the coastline of Antarctica between 64°S

1 and 77°S (Wilson 1983; Woehler 2001; Lea & Soper 2005; Ancel *et al.* 2013). However, our
2 present knowledge of the biology of the emperor penguin falls far behind the picturesque notori-
3 ety of this iconic species. The breeding behaviour and phenology of the species have been de-
4 scribed from only a few sites, e.g. from the Dion Islands (Trathan *et al.* 2011) and Pointe Géolo-
5 gie archipelago (Prévost 1961), usually settled in the vicinity of overwintering research stations.
6 Indeed, most colonies are just too remote from permanently occupied research stations, and be-
7 cause of the unique breeding cycle of this species (in which egg laying, hatching, and early chick
8 rearing all occur during the austral winter, see Prévost 1961), ground visits to most breeding
9 colonies are virtually impossible for most of the breeding season (Budd 1962). Yet, because the
10 French station in Adélie Land, Dumont d'Urville, has been located next to the Pointe Géologie
11 colony, this is the only colony which has been visited almost every day for more than 45 years
12 (Prévost 1961; Isenmann & Jouventin 1970; Isenmann 1971; Jouventin 1971a; Jouventin 1972a;
13 Jouventin 1975; Le Maho 1977; Ancel *et al.* 2009). However, it has been shown that demo-
14 graphic studies extrapolated from the colony-level can be strongly misleading as they tend to give
15 too much weight to local stochastic events (Lynch *et al.* 2012).

16 The recently developed satellite-based remote-sensing methods (Barber-Meyer *et al.* 2007;
17 Fretwell & Trathan 2009; Fretwell *et al.* 2012) are a major breakthrough. By detecting faecal de-
18 posits on the sea ice in satellite images, it is possible to locate emperor penguin colonies. Howev-
19 er, present-day satellite technology is constrained in Polar Regions as the darkness of the polar
20 winter and the frequent cloud cover during winter and early spring limit window of opportunity
21 for satellite-surveys. Further, wind-blown snow may obscure the faecal signal. Therefore, to ex-
22 pand our knowledge of this species beyond the boundaries of the few available local observations,
23 it is critical to corroborate and calibrate these remote-sensing methods through ground visits with
24 detailed and precise observations of breeding colonies.

Empirical evidence of heterogeneous dispersal - §139

1 The ground survey of the Mertz Glacier area was initiated within this context and became logisti-
2 cally possible in November 2012. The existence of an emperor penguin colony near the Mertz
3 Glacier had been suspected for almost a hundred years, and Mertz himself perished in an attempt
4 to locate the colony. More recently, during August 1999, sightings of thousands of emperor pen-
5 guins traversing the northern part of the Mertz Glacier tongue (Ancel *et al.* 1999) confirmed that
6 a colony must be located close to the glacier.

7 By using Landsat and QuickBird2 Very High Resolution Landsat imagery, confirmation of the
8 colony site at 66°54'S, 146°37'E at the tip of the Mertz glacier tongue was obtained in November
9 2009 (Fretwell & Trathan 2009; Fretwell *et al.* 2012). The high resolution of the satellite imagery
10 allowed counting 4,781 pairs at the time of image acquisition. However, before a field survey
11 could take place, the Mertz Glacier tongue calved in mid-February 2010 presumably inducing
12 the colony to relocate or merge with another colony. New satellite images obtained since then
13 suggested that the birds might attempt to breed on different sites depending on the year. Still,
14 since the exact colony location and colony size appeared to change between successive observa-
15 tions, ground-based assessments were difficult and needed to be conducted over several seasons.

16 Our objective was to validate the presence of one or more emperor penguin colonies near the
17 Mertz Glacier, to locate them, to estimate their population size (breeding adults) and breeding
18 success (number of chicks), and to obtain other biological information.

19 *Materials and Methods*

20 *§-139 Animal ethics* were not an issue as we counted the birds from aerial pictures. No manipula-
21 tion or experimentation was con- ducted on live birds and care was taken to avoid any distur-
22 bance. The French Polar Institute (IPEV) and the Terres Australes et Antarctiques Françaises
23 (TAAF) are the authorities that issued the permit to visit the colonies.

Empirical evidence of heterogeneous dispersal - §140

1 During the 2012–2013 austral summer, the French Polar Institute’s resupply ship MSS *L’Astrolabe*
2 *labe* visited the Antarctic coastline near to the Mertz Glacier. *L’Astrolabe*’s classification is ice class
3 1A super, the highest class of vessels that can operate in difficult polar conditions without the as-
4 sistance of an ice breaker. On November 1st and 2nd, using a Eurocopter Squirrel B-3 single-en-
5 gine helicopter, aerial surveys were conducted along the northern edge of the Mertz Glacier in an
6 attempt to locate the new breeding sites of the emperor penguin colonies. Aerial photographs,
7 used for recording penguin location and for counting the chicks and adults, were taken obliquely
8 in clear weather from an altitude of 300 meters, with a 35-mm Canon 40D digital camera fitted
9 with a 200 mm f/4 lens.

10 §-140 *Aerial photographs* were stitched together to form photo- mosaics in PTGui(c) software,
11 and both chicks and adults were counted separately using the count tool in Photoshop CS5©.
12 This was repeated by two different observers. The total number of individuals was estimated us-
13 ing two different methods to provide a probable range for each colony census: 1) under the hy-
14 pothesis that inter-annual climatic effects are largely dominant over fine- scale local effects, we
15 made the assumption that breeding success is approximately similar in nearby colonies for any
16 given year. We therefore estimated the number of breeding pairs from the early November chick-
17 counts using the population parameters deter- mined from the nearby Pointe Géologie colony
18 (66°40'S, 140°01'E) over the whole breeding season; 2) following Budd (Budd 1962), who pro-
19 posed that, based on a census at several different colonies, adult counts between October and No-
20 vember equate to approximately one-third (26% to 40%) of the total adult population, indepen-
21 dently of the actual breeding success. For comparison, the Pointe Géologie colony was counted at
22 different times, from photographs, using the same methodology. The total number of incubating
23 males in mid-June was taken as a proxy for the total number of breeding pairs: and chicks were
24 counted in early November, to allow comparison with the census of the Mertz colonies.

Empirical evidence of heterogeneous dispersal - §141

1 To avoid potential disturbance, no low altitude flights were made over or near to the Mertz
2 colonies, and a brief landing was made ca. 1 km from the breeding sites to minimize further dis-
3 turbance. Each colony was visited by 4 or 5 people on foot, and surveyed for approximately 3
4 hours each. No direct ground counts of chicks or adults were made during the visits due to the
5 lack of any satisfactory vantage point. Observations focused on assessing the habitat, chick abun-
6 dance and health (though no direct measures were taken in order to avoid unnecessary distur-
7 bance). Observations were also made to search for flipper-banded individuals, potentially orig-
8 inating from the Pointe Géologie colony, located at only ca. 250 km to the west.

9 To find colony locations, satellite observations were made using freely available reduced resolu-
10 tion “quicklooks” from the com- mercial satellite provider DigitalGlobe (<https://browse.digital->
11 [globe.com/imagefinder/](https://browse.digital-globe.com/imagefinder/)). These images have an approximate on-the-ground resolution of ,10 m
12 per pixel. Colonies were located visually from the images by detecting the brown signature of the
13 bird’s faecal stains on the ice. Once located, raw Very High Resolution imagery with a resolution
14 of 0.5 m per pixel was acquired. This resolution allows areas of penguins to be differentiated from
15 guano using a multivariate supervised classification method (Barber-Meyer *et al.* 2007; Fretwell
16 & Trathan 2009; Fretwell *et al.* 2012).

17 *Results*

18 *§-141 Habitat and Chick Census. First or eastern colony.* The existence of a breeding colony of em-
19 peror penguins near the Mertz Glacier was confirmed and located at 67°19'S, 145°52'E on No-
20 vember 1st, 2012 (Fig. 40). The eastern colony was breeding on fast ice amidst grounded ice-
21 bergs, 250 km east of Dumont d’Urville research station. The colony was composed of numerous
22 sub-groups, which appeared to have moved short distances, as confirmed by the clustered guano
23 deposits. The topography of the breeding site was very uneven with numerous icebergs embedded
24 in the sea ice, offering protection from the prevailing high winds. Layers of droppings were found

Empirical evidence of heterogeneous dispersal - §141

1 between layers of snow on top of the icebergs (Fig. 41), providing evidence of numerous snow-
2 fall episodes. Many icebergs, ca. 10-meter high, were also covered with guano, thus demonstrat-
3 ing that breeding birds had spent some time on the top of these icebergs (Fig. 41). This colony
4 had three former incubation areas, identified by stained sea ice, numerous egg shells and carcasses
5 of newly hatched chicks. During our 3-hour visit, no birds other than penguins, and no sign of
6 predation (e.g. predated or scavenged carcasses) were found. Two Weddell seals *Leptony-
7 chotes*
8 *weddelli* were sighted along tidal cracks in the vicinity. Our survey yielded a count of 1,750
9 downy chicks, and 1,520 adults. Approximately 300 dead chicks, ranging in age from a few days
10 (frozen) to 3-month-old (both frozen and fresh) were found. Most of them were found at the in-
11 cubation sites, and on the top of small icebergs.

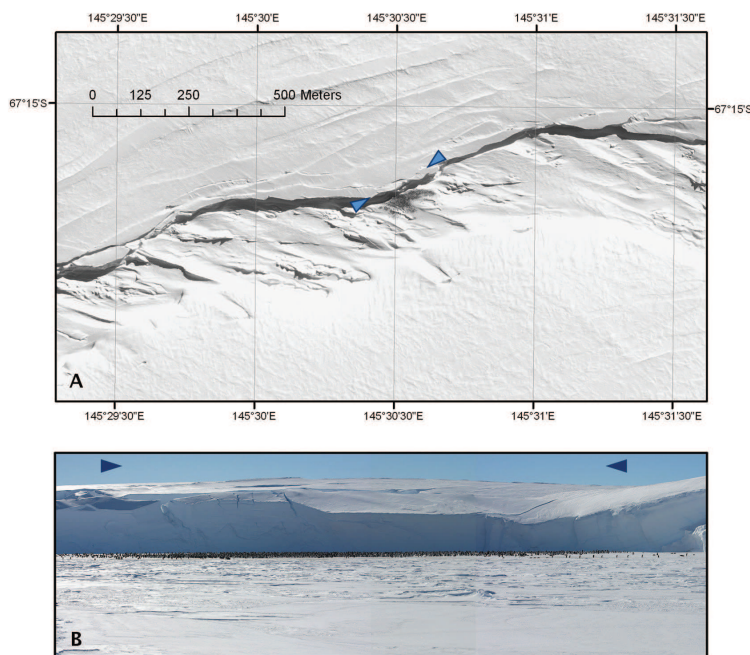


Figure 40 | Western Emperor penguin colony. A. Quickbird satellite view of the eastern emperor penguin colony (67°19'S, 145°52.9'E) showing adults and chicks present on November 1st, 2012. B. Ground view of the same (blue arrows at matching locations). Photograph by Robin Cristofari.

20 *Second or western colony.* On No-
21 vember 2nd, 2012, a second
22 colony was discovered by chance
23 from an altitude of ca. 500 m
24 during a helicopter flight in-
25 volved in resupply operations. This western colony was located on a large and flat fast ice sheet
26 extending from the Mertz glacier, ca. 20 km west of the first colony, at 67°14'S, 145°30'E (Fig.
27 #). This colony was bordered on the north-west by a high ice cliff of the Mertz glacier. Numerous
28 large tabular icebergs were grounded in the fast ice, south of the colony. This habitat was similar

Empirical evidence of heterogeneous dispersal - §142

1 to that generally found in the Ross Sea area, characterized by stable fast ice, nearby open water
2 and access to fresh snow (Kooyman 1993). The site was poorly sheltered from the wind. The
3 colony comprised one single group of ca. 3,980 downy chicks, and 2,300 adults. From empirical
4 observation, these chicks seemed bigger, and better fed, than those of the eastern colony. Three
5 Antarctic skuas *Catharacta maccormicki* and one giant petrel *Macronectes giganteus* were sighted
6 flying over the colony. Only 17 dead chicks and one abandoned egg were found during our 3-
7 hour survey, some of them partly buried. The ice surface was of heavily compacted and abraded
8 snow, and it is therefore most likely that more dead chicks were buried beneath the snow surface.
9 Guano traces were restricted to the area currently occupied by the birds, and its direct surround-
10 ings. Due to these difficult conditions, we were unable to precisely locate the incubation area. As
11 for the first colony, we did not observe any individuals bearing flipper-bands.



Figure 41 | Eastern Emperor penguin colony. A. Location of the eastern colony ($67^{\circ}199\text{S}$, $145^{\circ}529\text{E}$) on the sea-ice at the time of our visit and previous locations of the colony on top of icebergs, ca. 10 m high, covered with guano (arrow). B & C. Layers of droppings (arrows) covered by several layers of snow indicate abundant snow-covered episodes during the breeding season. Photographs by Robin Cristofari.



19 *§-142 Extrapolation to Total Colony Size.* Our average
20 2012 breeding season count for the Pointe Géologie
21 colony was ca. 3,200 breeding pairs, and we counted ca.
22 2,500 chicks in early November 2012, which gives a
23 pair-to-chick ratio of ca. 1.28. Applying this ratio to the
24 chick counts for the Mertz colonies, we estimated the
25 population size to be ca. 2,300 breeding pairs for the
26 eastern colony, and ca. 5,100 for the western colony (Table 8).

Empirical evidence of heterogeneous dispersal - §143

1 Moreover, a total of 1,864 individual adults were counted at the Pointe Géologie colony on Oc-
2 tober 31st, 2012, which implies a total count of ca. 2,330 to 3,580 breeding pairs according to
3 Budd's method. This is consistent with our average 2012 winter count of ca. 3,200 breeding
4 pairs, thus allowing us to apply this method to the Mertz Glacier census.

5 Adult counts from photographs were 1,520 for the eastern colony and 2,300 for the western
6 colony, respectively (Table 8). When applying Budd's method (Budd 1962), the estimated popu-
7 lation size was 2,300 breeding pairs for the first colony (ranging from 1,900 to 2,900) and 3,500
8 for the second (from 2,900 to 4,400), hence a total of ca. 5,800 breeding pairs for the Mertz em-
9 peror penguin population.

10 **Table 8 | 2012 census of the Mertz emperor penguin colonies.**

	Lat. (S)	Lon. (E)	Chicks	Adults	Chicks-to- adults ratio	Pairs (from chicks)*	Pairs (from adults)
11 Eastern	67°19'	145°52'	1,750	1,520	1.15	~2,300	1,900-2,900
12 Western	67°14'	145°30'	3,980	2,300	1.73	~5,100	2,900-4,400

13
14 *Breeding pairs were obtained by applying a pair-to-chick ratio of ca. 1.28 obtained from the emperor penguin
15 colony at Pointe Géologie, Adélie Land.

16 *Discussion*

17 *§-143 Strength and Shortcomings of the Remote-sensing Technique.* Our study confirmed sightings
18 made 15 years ago of thousands of emperor penguins going back and forth at the northern part
19 of the tongue of the Mertz Glacier, suggesting the presence of a colony in this area (Ancel *et al.*
20 1999). Moreover, our ground visit to the two Mertz colonies demonstrates the reliability of the
21 remote-sensing method developed by researchers (Barber-Meyer *et al.* 2007; Fretwell & Trathan
22 2009). Although the method successfully identified several previously known emperor penguin
23 colonies, this is the first time that the existence of three colonies identified solely from satellite
24 images have been accurately corroborated in the field: two at the new edge of the Mertz Glacier

Empirical evidence of heterogeneous dispersal - §143

1 in late 2012 (present study) and one on Princess Ragnhild coast (69°54'S, 27°09'E) in early 2013
2 (visit [http://www. antarcticstation.org/](http://www.antarcticstation.org/)for more information).

3 However, the locations from satellite surveys of the two colonies at the edge of the Mertz Glacier
4 were not determined during the same year, therefore suggesting the existence of only one colony
5 in this area. This fact suggests that further refinement of the present- day remote-sensing method
6 may be advantageous. Only the western colony was located in December 2011, while the eastern
7 one remained invisible. In contrast, only the eastern colony was located in 2012. Failure to
8 identify both colonies at the same time may be attributed to different factors. To locate colonies,
9 the remote-sensing method relies on guano staining on the sea-ice surface, and heavy snow cover
10 may obliterate the signal. The western colony, which could not to be detected just a few weeks
11 before our visit, was largely covered by snow during the last weeks of winter. Hence, almost no
12 guano stains, abandoned eggs, or dead chicks were found. The location of this colony along a
13 high glacier edge may also have kept it hidden from the satellite, if the image was taken at too
14 shallow an angle. In addition, a previous study (Gilbert *et al.* 2008) has shown that the density of
15 emperor penguin colony can vary considerably, i.e. from 2 to 9 birds per square meter, and this
16 within only a few hours. This means that census estimates based on colony area may be inaccurate
17 in some circumstances due to variability in density and to inter-annual variability. This suggests
18 that this colony was not the only one to be missed by the 2009 and 2012 satellite surveys,
19 and that more colonies remain to be discovered in other parts of the continent. Although not always
20 possible each season due to cloud cover and the availability of the satellite, our results suggest
21 that satellite surveys should be conducted repeatedly and combined with field surveys to ensure
22 that colonies are not missed. Our paper shows that to allow confidence in satellite
23 observations, a multi-temporal/multi-year approach has to be used to ensure that breeding sites
24 are not missed due to heavy snowfall, deep shadows or topographic features such as ice cliffs. If
25 possible, these observations should be backed up with aerial or ground counts as the limited spatial
26 resolution of satellite imagery results in large inherent variances when calculating breeding

Empirical evidence of heterogeneous dispersal - §143

1 populations. Innovative cameras combined with biologging would also be very useful to com-
2 plete the satellite survey. For instance, since the satellite survey based on 2009 imagery, four more
3 emperor penguin colonies have been discovered: two on the West Ice Shelf detected by aerial sur-
4 vey (Wienecke 2012), and two others on the West Ice Shelf and near the Jason Peninsula identi-
5 fied by satellite survey (Fretwell *et al.* 2014).

6 Due to the discrepancy between the coordinates given by satellites (66°54'S, 146°37'E, see
7 Fretwell *et al.* 2012) and our land based ones (67°19'S, 145°52'E and 67°14'S, 145°30'E), we
8 may assume that the colony initially detected from space in 2009 has been split into two
9 colonies. Following the calving of the Mertz Glacier, the birds may indeed have attempted to set-
10 tle in new favourable surroundings. The two new colonies are separated by 20 km along the new
11 northern edge of the Glacier (Fig. 42). Taking into account the proximity of the two breeding
12 sites, these colonies might however reunite in the future, especially if one or other of the sites
13 eventually proves to be more reliable for breeding and/or foraging.

14 All satellite surveys conducted up to now, before and after the Mertz calving, had concluded the
15 presence of a single emperor penguin colony, even if grouped into three close sub-colonies in
16 2009, in the Mertz Glacier area. It may also be that two nearby colonies already existed at the
17 current location(s) before the 2010 calving. This is supported by the fact that the best population
18 estimate of the Mertz Glacier colony located by Fretwell *et al.* (Fretwell *et al.* 2012) was estimated
19 to be 4,781 adults (or 5,976 pairs if we considered that 80% of the total breeding population was
20 present during the satellite survey, see Fretwell *et al.* 2012), a figure which is relatively close to
21 our assessment of 5,100 breeding pairs for the western colony alone (Table 8) but lower than the
22 sum of the two colonies we found (7,400 breeding pairs; Table 8). In such a scenario, the second
23 colony we discovered might be the one found before at the nearby Ninnis Glacier (Wienecke
24 2009a), which would have been missed in all further surveys (see Fretwell & Trathan 2009).

Empirical evidence of heterogeneous dispersal - §144

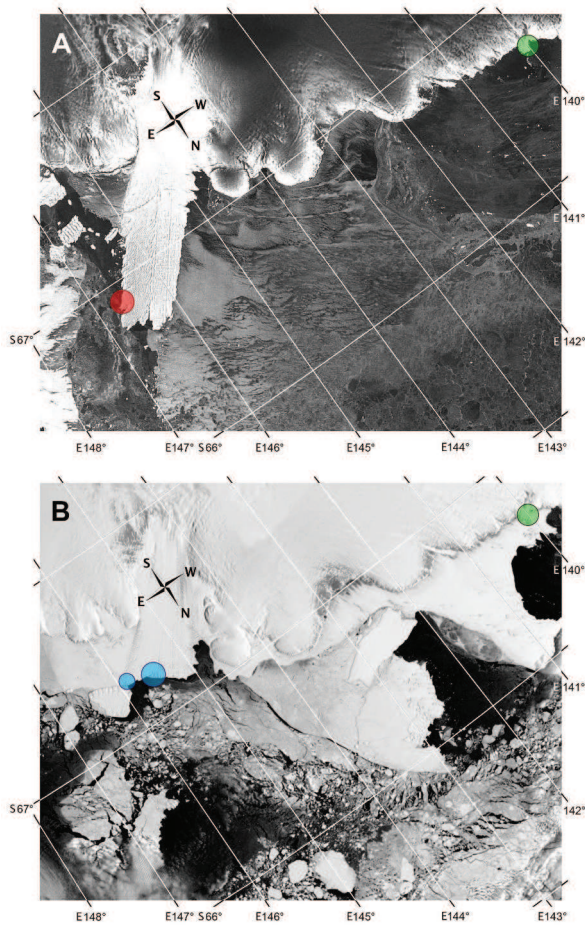


Figure 42 | Satellite images of the Mertz glacier.

Circle area is proportional to colony size. Continent is on top, open sea at the bottom. **A.** Estimated location of the Mertz emperor penguin colony on November 13th, 2009 (red circle) on the eastern flank of the Glacier tongue. A large crack in the Mertz Glacier is visible (red arrow). The green circle corresponds to Pointe Géologie colony. **B.** Location of the two new colonies of emperor penguins on December 3rd, 2012 (blue circles), 2 years after the 2010 calving of the Glacier tongue. The berg (overall length of 80 km and a width of 40 km) broke off the Mertz Glacier after being rammed by another iceberg. The green circle corresponds to Pointe Géologie colony. Images downloaded from the USGS website (<https://lta.cr.usgs.gov>).

§-144 Influence of the Breeding Location at Fine-spatial Scale. It is important to note that the two colonies appeared to differ slightly in their breeding stage. The chick-to-adult ratio was higher in the western colony than in the eastern one (1.73 vs. 1.15, respectively: Table 8). This result may be

interpreted as a difference in breeding success, with higher chick mortality in the eastern colony than in the western one. This hypothesis is supported by the much higher number of dead chicks in the eastern colony. Moreover, while the two colonies were visited on two consecutive days, the western colony appeared more advanced in its breeding cycle, as indicated by the generally better health and bigger size of the chicks than in the eastern colony. Finally, the western colony was much closer to the open water than the eastern one at the time of our visit, and a recent study (Massom *et al.* 2009) has shown that the distance between an emperor penguin colony and the open water may tend to correlate negatively with breeding success.

Budd (Budd 1961) proposed that “each rookery can be regarded as a compromise, and not always a very successful one, between the emperor’s sometimes conflicting requirements of safety from sea- ice breakouts, shelter, and access to food”. Finding a suitable trade-off may well be an

Empirical evidence of heterogeneous dispersal - §145

1 arduous matter of trial-and-error, since inter-annual variations make it difficult for the birds to
2 evaluate rapidly the consistency of ice conditions at a particular location. Therefore, after the loss
3 of their original habitat, some birds are probably breeding in a sub-optimal location, which may
4 explain the difference in breeding success between neighbouring and possibly related colonies.

5 *§-145 Status of the Emperor Penguins Census.* Emperor penguins breed almost around the entire
6 coastline of the Antarctic Continent (Melick & Bremmers 1995; Coria & Montalti 2000;
7 Woehler 2001; Todd *et al.* 2004; Fretwell *et al.* 2012). Their known breeding distribution extends
8 from Snow Hill Island (64°31'S, 57°26'W) to Gould Bay (77°43'S, 47°41'E) in the Antarctic
9 Peninsula, and Cape Crozier on Ross Sea (77°28'S, 169°19'E). The Snow Hill Island and Gould
10 Bay (Woehler 2001) colonies are the most northerly and the most southerly known emperor pen-
11 guin colonies, respectively. Colony sizes vary from more than 20,000 pairs (Coulman Island,
12 Cape Washington, Halley Bay) to just a few hundred pairs (Umbeashi Rock, Amundsen Bay,
13 Fold Island, Cape Crozier). The two largest known colonies are in the Ross Sea: Cape Washing-
14 ton with ca. 24,000 pairs and Coulman Island with ca. 28,000 pairs (Kooyman 1993).

15 The closest colonies to the Mertz Glacier are at Pointe Géologie (66°40'S, 140°01'E) and Davis
16 Bay (69°21'S, 158°29'E). While ca. 2,500 chicks of emperor penguins were raised at the Pointe
17 Géologie colony during the 2012 breeding season, the two new colonies together numbered ca.
18 5,700 chicks. Since a pair of emperor penguins may only successfully raise one chick per year, the
19 population of breeding emperor penguins in this area of Antarctica can therefore be estimated,
20 based on these chick counts, to more than ca. 16,400 breeding adults, about a fourth more than
21 previously estimated (Fretwell *et al.* 2012). This count represents a minimum estimate for the
22 breeding population, given uncertainty about the mortality rate of chicks prior to our visit.

23 Because of the persistence of the sea ice, few ships are able to reach the Antarctic coasts before
24 post-breeding adults disperse from their colonies between December and early January. Except
25 for colonies close enough to research stations, during the breeding season (from March-April to

Empirical evidence of heterogeneous dispersal - §145

1 December-January, depending on the latitude), access to colonies is difficult or impossible due to
2 adverse weather and/or extensive pack ice (Kooyman 1993). Consequently, many emperor pen-
3 guin colonies have never been counted and some have presumably not even yet been discovered.
4 Further- more, in addition to the Mertz colonies, another colony, on the Princess Ragnhild coast
5 and for which location was reported in 2009 by satellite images (Fretwell *et al.* 2012), was first
6 visited in early January 2013 by Belgian scientists, who surveyed ca. 20,000 adults, compared to
7 6,870 adults detected from satellite survey (Fretwell *et al.* 2012). The current global emperor
8 penguin population can therefore be estimated to be, at least, ca. 260,000 pairs from 52 breeding
9 colonies (Fretwell *et al.* 2012 and present study), or ca. 25% more than estimated only 15 years
10 ago (Woehler & Croxall 1997).

11 Over the past five years satellite surveys have proved a very effective method for finding new em-
12 peror penguin colonies. However, as our visit to the western colony demonstrated, the nitrogen-
13 signature of snow-covered droppings around colonies that underwent numerous snowfalls, and
14 the shade of an ice front indicate that at certain times of year, under challenging environmental
15 conditions it is virtually impossible to detect some colonies from satellite images. Importantly,
16 several colonies have not been observed since they were first reported. This problem, of missing
17 colony locations by satellite survey may be solved by taking multiple satellite images within a
18 breeding season and especially towards the end of the breeding season when the guano is more
19 apparent. Multi-temporal and multi-year satellite surveys, backed up with confirmation and fur-
20 ther ground truthing from aerial or ground based counts are essential if we wish to fully under-
21 stand the population demography and dynamics of emperor penguins. As a consequence, we can
22 still hope to discover more emperor penguin colonies in the future if further investigations are
23 conducted along the Antarctic coast.

1 *Acknowledgements*

2 We thank the crew of the MSS L'Astrolabe (resupply ship of the P&O Company, chartered by
3 the Institut Polaire Français Paul-Emile Victor (IPEV) and the Terres Australes et Antarctiques
4 Françaises) for its logistic support, the helicopter pilot Sebastien Vignoud (Groupe SAF-HELI-
5 CAP) for making the visit possible, Patrice Godon, Gaëlle Sellin, Alain Pierre and Serge Drapeau
6 (IPEV) for their assistance in the field, Stan Zamora (P&O) for providing the satellite images,
7 and Alain Hubert (International Polar Foundation) for providing the data relative to Princess
8 Ragnhild coast. We are grateful to the USGS (<https://lta.cr.usgs.gov>) for providing satellite im-
9 ages for free. This study was supported by the IPEV (Programme 137) and the Centre National
10 de la Recherche Scientifique.

1 Chapter 7: Unexpected philopatry in an insular seabird, the Pe- 2 ruvian diving-petrel

3 This chapter is based on Cristofari R, Fernandez-Zamora F, Gouin N, Le Bohec C, Plaza P, Truc-
4 chi E, Zavalaga C, Luna Jorquera G, Unexpected population isolation in a critically endangered
5 insular seabird, the Peruvian Diving Petrel (*Pelecanoides garnotii*). (in prep).

6 *Context*

7 §-146 *The Peruvian diving-petrel as a reference species.* In the previous chapters, we introduced the
8 particular case of the *synnomes* of the King and Emperor penguins - a highly original population
9 organisation, in which the whole species, despite its fragmented colonial distribution, remains
10 fully panmictic. This structure may be seen as an extreme case in the broad spectrum of seabird
11 genetic and spatial organisations (see *e.g.* a review in Friesen *et al.* 2007): in the previous chapters
12 (see *The King synnome*, p. 163 and *The Emperor synnome*, p. 211), we argue that this particular
13 form of transient philopatry is a significant behavioural trait that allows *Aptenodytes* penguins to
14 cope with the unstable polar environment, driven either by the extent of coastal upwelling and
15 polynya activity, or by the latitudinal variations in the Polar Front location. In order to test that
16 prediction, we review here the contrasting case of the Peruvian diving petrel *Pelecanoides garnotii*.
17 A procellariimorphe like the penguins (Zhang *et al.* 2014), this small auk-like procellariform bird

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §147

1 offers several convergent traits with the spheniscids. Like them, its wings are derived and adapted
2 to underwater swimming (Luna-Jorquera *et al.* 2003 - a trait it shares with *e.g.* Alcids) - albeit
3 greatly reduced, flight function is however not totally lost. This diving bird feeds mostly on fish
4 and krill (Jahncke *et al.* 1999) and may reach great depths relative to its small size (routinely 30m
5 but depths of more than 80m have been recorded, see Zavalaga & Jahncke 1997). Finally, lo-
6 calised upwellings seem to be a major determinant of foraging behaviour in this central-place for-
7 ager (Zavalaga *et al.* 2010, Fernandez-Zamora *in prep.* and Luna Jorquera, *pers. com.*). However,
8 contrary to *Aptneodytes* penguins, *P. garnotii* enjoys a mostly stable oceanographic landscape. It
9 breeds on a handful of offshore islands of central Northern Chile and Southern Peru (Hays 1989;
10 Simeone *et al.* 2003), that are part of the Humboldt Current System.

11 §-147 *The Humboldt Current System.* The Humboldt Current System (HCS) is a Northward ex-
12 tension of the Pacific subantarctic watermasses. Like the Benguela Current in the Atlantic basin,
13 it is one of the world's major Eastern Boundary Currents: North of the West Wind Drift zone (see
14 §7 p. 35), the Southern Pacific anticyclonic gyre drives nearly constant equatorwards winds along
15 the coast of South America. These winds result in a strong northward drift of cold subantarctic
16 surface waters along the coast, and in offshore Ekman transport: near the shore, the drifting sur-
17 face waters draw the upwelling of tropical nutrient-rich subsurface waters, which result in one of
18 the most highly productive oceanic systems on the planet (see Thiel *et al.* 2007 for details). Up-
19 welling activity is regulated by the seasonal displacement of the Pacific anticyclonic gyre, and is
20 constant only in Northern Central Chile and in Peru. It is also strongly influenced by the topolo-
21 gy of the South American continental shelf, where local cape and island systems promote up-
22 welling activity - *e.g.* in the Coquimbo Coastal System (Montecino *et al.* 2006). Thus, although
23 the El Niño Southern Oscillation (see §8 p. 40), just like in the Southern Ocean, has a deep im-
24 pact on local productivity, it does not influence the location of upwelling hotspots, that remain at
25 the same distance from breeding islands. This relative stability of the HCS topological and

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §147

1 oceanographic features sets a very different expectation for the population structure of seabirds¹.
2 An overview of the general behaviour of procellariimorphes makes philopatry a likely ple-
3 siomorphic trait in these species (Friesen *et al.* 2007), that may be counteracted by specific con-
4 straints, such as an excessively unstable environment: we therefore expect a « stable » species such
5 as the Peruvian diving petrel to represent the other extreme along the seabird population struc-
6 ture spectrum.

7 *Abstract*

8 Seabirds are often used as reliable sentinels of oceanic environments: however, idiosyncratic life
9 history traits strongly influence both their response to ecosystem changes and our capacity to as-
10 sess and interpret the extent of this response. In particular, the strength of philopatric behaviour
11 and dispersal ability are important determinants of each species' evolutionary response. Here, we
12 show that a representative endemic species of the Humboldt Current System, the once abundant
13 but now critically endangered Peruvian diving petrel (*Pelecanoides garnotii*), exhibits extreme
14 philopatric behaviour. Mitochondrial and genome-wide SNP data provide evidence for strong
15 isolation and low migration at very short distance. We suggest that behavioural and morphologi-
16 cal adaptations to coastal insular habitat and underwater foraging are the main drivers of the ex-
17 ceptionally strong philopatry in this auk-like species. Importantly, range fragmentation, com-
18 bined with ongoing anthropogenic pressure, are expected to put the Diving-petrel under
19 continued high risk, as the full demographic independence between colonies is likely to preclude
20 local population rescue through migration. The independence of each island's breeding popula-

1. Here, it is important to note that *stability* is only relative to a particular species' *habitat* (see §49 p. 127) within a geographic or oceanic system. For example, we stated earlier that the Emperor penguin's sea ice habitat appeared as extremely unstable (see **Empirical evidence of heterogeneous dispersal**, p. 259): yet within the precise same area, the habitat of *e.g.* bottom-dwelling notothenioid fish may be totally stable. Thus, it makes more sense to think about the stability or instability of *habitats* than of *environments*. Here, therefore, the stability of the HCS is only considered *relative to the standard seabird habitat*, that is heavily dependant on local upwellings and guano insular breeding locations.

1 tion should be a key element in the conservation strategy for this key species of the Humboldt
2 Current System coastal archipelagos.

3 *Introduction*

4 *§-148 Seabirds as terrestrial animals.* Seabirds are an important link between marine and terrestrial
5 ecosystems (Blais *et al.* 2005; Ellis 2005; McCauley *et al.* 2012): but that double bond makes
6 them highly susceptible to changes on both sides of the shoreline. Their sensitivity to changes in
7 marine ecosystems is now a well-documented subject: it involves either long-term, indirect im-
8 pact of human activities through climate change (Le Bohec *et al.* 2008; Sydeman *et al.* 2012), or
9 direct impact through fisheries which compete with seabirds for food resource (Tasker *et al.*
10 2000; Becker & Beissinger 2006; Karpouzi *et al.* 2007; Cury *et al.* 2011; Wagner & Boersma
11 2011) or even cause massive direct seabird mortality (Weimerskirch *et al.* 1999; Lewison *et al.*
12 2004). Yet the often overlooked disruption of terrestrial breeding habitats can also have a very
13 significant impact on population dynamics. Impacts may range from spread of diseases on the
14 colony areas (Barbosa & Palacios 2009) to sudden habitat disappearance (Ancel *et al.* 2014): yet
15 the most common cause of habitat loss remains the growing interaction with terrestrial predators,
16 including humans (see Croxall *et al.* 2012 for a review). Unlike marine environment modifica-
17 tions, increasing disturbance on land may be very restricted locally, and affect only one or a few
18 breeding colonies. Therefore, both environmental compartments need to be taken into account
19 when studying seabirds. This applies to the understanding of population dynamics, since locally-
20 based monitoring programs and studies may largely misjudge global tendencies by mistaking lo-
21 cal effects for general trends (Mallory *et al.* 2010; Lynch *et al.* 2012) - but also, importantly,
22 when devising conservation strategies, since marine and land pressures need to be mitigated
23 separately.

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §149

1 §-149 *Recent history of the Peruvian diving-petrel.* The Peruvian diving petrel (*Pelecanoides gar-*
2 *notii*) is a dramatic, although by no means unique, example of the combined effect of these mul-
3 tiple direct and indirect threats. According to historical records, this small insular procellariid was
4 one of the most abundant endemic seabird of the Humboldt Current System, and presented a
5 broad distribution {Coker 1919; Murphy 1936} from ~ 6°S (in Isla Lobos de Tierra, Peru) and ~
6 42°S (in Isla Chiloé, Chile) along the Pacific coast of South America. Yet the species declined
7 rapidly during the 20th century as a result of guano extraction (Murphy 1936; Hays 1989;
8 Figueroa *et al.* 2011), hunting (Hays 1989), bycatch (Hays 1989; García-Godos & Goya 2006),
9 human-introduced predators (Araya Modinger & Duffy 1987; Simeone *et al.* 2003) and habitat
10 competitors (Fernández Zamora 2011), food competition with fisheries (Hays 1989; Jahncke &
11 Goya 1998), and the increased effects of ongoing climate change on the productivity of the
12 Humboldt Current System (HCS, see Thiel *et al.* 2007). This led to the decline of their nesting
13 habitats, from thirteen colonies reported in the past to seven nesting sites at present (Araya
14 Modinger & Duffy 1987; Vilina 1992; Figueroa & Stucchi 2008; Figueroa *et al.* 2011). Its
15 breeding range has then been reduced and fragmented into two main areas, divided by a 1,300
16 km gap - one on the central Peruvian coastline, around 8°S and 14°S (Isla Corcovado, Isla San
17 Gallán and Isla La Vieja (Jahncke & Goya 1998; Valverde Romero 2006), and one in northern
18 Chile, between 26°S to 29°S (Islas Pan de Azúcar, with ~220 pairs, Isla Choros, with ~9,516
19 pairs, Isla Grande, with ~200 pairs, and Isla Pájaros II, with ~120 pairs (Mattern *et al.* 2002;
20 Simeone *et al.* 2003; Martinez Palma 2014). Historical population sizes are little known, but all
21 available information points to a critical decline: while 100,000 pairs bred on Isla Chañaral,
22 Chile, in 1938 (Araya Modinger & Duffy 1987), this population is now completely extinct
23 (Simeone *et al.* 2003); and the now recovering Peruvian population (Jahncke & Goya 1998) was
24 estimated to a total of ca. 4,000 individuals in 1989 (Hays 1989).



1 **Figure 43 | Full distribution of *Pelecanoides garnotii*.** (A) Peruvian and (B) Chilean range of the species.
2 To the left, close-up of the Chilean range of the species (dashed lines indicate discontinuity in our representation of
3 the shoreline). Isla Chañaral is currently unoccupied. Sampled islands (Pan de Azúcar and Choros) are marked by
4 coloured circles.

5 *§-150 Present-day status of the species.* Although the Peruvian diving-petrel's colonies have declined
6 and even gone extinct during recorded history, marginal range re-expansion and local population
7 regrowth have been documented in the past decades (Figueroa *et al.* 2011). Thus, small colonies
8 have been discovered on Corcovado Island, Perú (Valverde Romero 2006), and non-breeding in-

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §151

1 individuals have been sighted on Lobos de Afuera islands (Figueroa & Stucchi 2008; Figueroa *et al.*
2 2011). A rapid growth of observed census size was also documented in the past decade in Perú,
3 on La Vieja Island (IUCN Red List) and in the Chilean islands (Simeone *et al.* 2003; Fernandez
4 *et al.* 2014). However, the processes underlying this possible population recovery are poorly un-
5 derstood, and assessing true regrowth is made difficult by our very limited knowledge of the
6 species' life history traits. It has been suggested that the diving-petrel's poor flight abilities greatly
7 reduces its dispersal possibilities (Luna-Jorquera *et al.* 2003), and that the high investment re-
8 quired for burrow nest construction is likely to promote a strong philopatric behaviour (Furness
9 1978; Cortés Labra 2007). It has thus been proposed that diving petrels occupy a coastal niche
10 and are thought to be year-round residents (Murphy 1936; Figueroa *et al.* 2011), which would
11 further reduce its dispersal opportunities.

12 *§-151 Panmixia in the Humboldt Current System.* While most studies conducted in seabirds have
13 emphasized the importance of genetic or phylogeographic structure in species dynamics (see
14 Friesen *et al.* 2007 for a review), the Humboldt Current System stands out as an exception. Ge-
15 netic surveys conducted in its major seabird species have concluded in a lack of genetic structure
16 at the continental scale (*e.g.* in the Peruvian booby (*Sula variegata*, see Taylor *et al.* 2010) or the
17 Peruvian pelican (*Pelecanus thagus*, see Jeyasingham *et al.* 2013), or in very high gene flow be-
18 tween colonies in the flightless Humboldt penguin (*Spheniscus humboldti*, see Schlosser *et al.*
19 2009). High dispersal, possibly promoted by foraging concentrations around localised upwellings
20 (Jeyasingham *et al.* 2013), has therefore been proposed as a central underlying mechanism. It has
21 also been suggested that high dispersal is one of the keys to the higher resilience of these species
22 to habitat change compared to other seabirds (Jeyasingham *et al.* 2013). Thus, it is unclear
23 whether the Peruvian diving petrel follows the expectation of high dispersal and panmictic popu-
24 lation set by other seabirds of the Humboldt Current System, or rather the highly philopatric and
25 low-dispersal life history strategy that its behaviour suggests.

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §153

1 §-152 *Aims of the study.* If *P.garnotii* exhibits the same degree of panmixia as the other seabirds of
2 the Humboldt Current System, demographic independence of colonies may not be assumed. In
3 that case, observed local population growth may rather be interpreted as the result of a complex
4 process (for example, several local extinction or exodus events, anecdotally causing census increas-
5 es in a handful of refugial areas through immigration). This would imply that only range-wide
6 census may be informative as to the status of the species. Alternatively, however, if the hypothesis
7 of high philopatry and low dispersal is verified, demographic independence may be expected be-
8 tween breeding locations throughout the species' highly fragmented range. In that case, popula-
9 tion trends observed at single locations may reflect true demographic responses, *e.g.* to local con-
10 servation measures (such as reduced guano extraction in several protected islands, and creation of
11 marine protected areas), rather than range-wide and species-scale processes. In order to under-
12 stand the current demographic trends for the Peruvian diving-petrel, and the extent of the threats
13 to the remaining populations, it is therefore necessary to establish whether observed local growth
14 is a result of (i) intrinsic local population growth, or (ii) immigration-fueled refugial concentra-
15 tion. Here, we use both mitochondrial and genome-wide genetic markers to distinguish between
16 these two hypotheses, and assess the current degree of demographic independence between
17 diving-petrel colonies throughout its Chilean range.

18 *Results*

19 §-153 *Sequencing data.* Genome-wide short-locus RAD-sequencing and mitochondrial data (see
20 *Methods*) are in good agreement in *Pelecanoides garnotii*. The Illumina sequencing lane yielded
21 182,319,948 paired-end reads with a mean phred-scaled quality score of 37, of which
22 144,048,478 were retained after barcode and cut-site sequence filtering. The final database in-
23 cluded 65,582 distinct RAD loci, of which 37,170 were polymorphic, with an average of 1.3
24 SNPs per locus. After PCR-duplicate removal, SNP calling from the higher-quality reads in

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §154

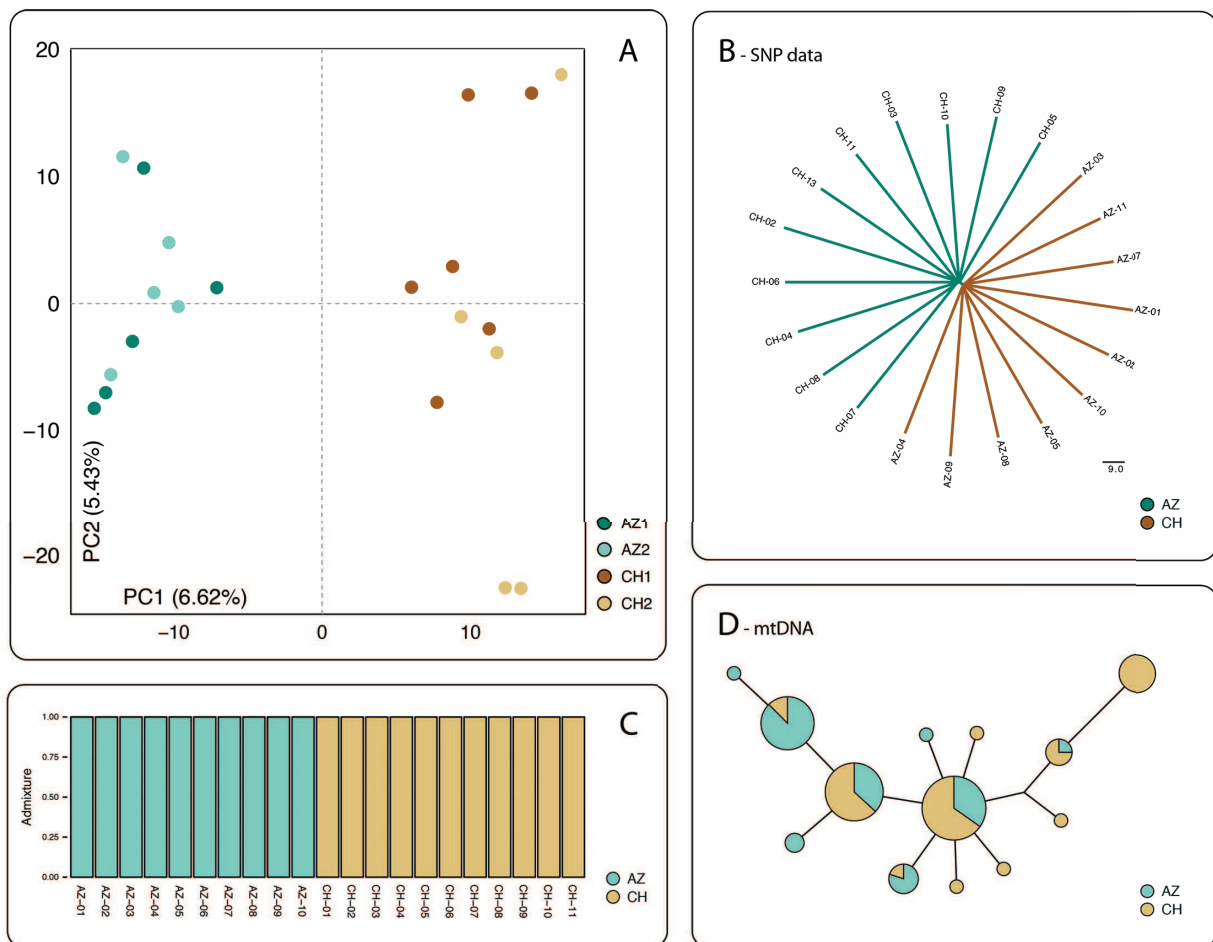
1 GATK yielded 43,175 SNPs, of which 32,964 were genotyped in at least 75% of individuals,
2 with a mean depth of 3.2X, and a median depth of 3.1X. Nucleotide diversity computed from
3 genome-wide SNP data was very similar across islands, with $\pi_{\text{CHR}} = 0.00224$ and $\pi_{\text{AZR}} = 0.00226$
4 (and pooled diversity $\pi_{\text{ALL}} = 0.00233$), at a level in accordance with the expectation for long-lived
5 seabirds (Romiguier *et al.* 2014), and Tajima's D supported neutral evolution at the selected sites
6 ($D_{\text{CHR}} = -0.364$ $D_{\text{AZR}} = -0.361$). mtDNA-based indices support this view: nucleotide diversity
7 was of the same order of magnitude for both islands, although slightly higher in Choros ($\pi_{\text{CHR}} =$
8 0.00232 and $\pi_{\text{AZR}} = 0.00161$, and pooled diversity $\pi_{\text{ALL}} = 0.00218$). Both Tajima's D and Fu &
9 Li's F indices supported neutral evolution of the polymorphic sites ($D_{\text{CHR}} = -0.463$ and $D_{\text{AZR}} =$
10 -0.649 , $P > 0.10$; $F_{\text{CHR}} = -1.26$ and $F_{\text{AZR}} = -1.34$, $P > 0.10$). The sequenced cytochrome-b frag-
11 ment contained 13 polymorphic sites, only 6 of which were polymorphic at both locations.

12 *§-154 Population structure.* All analyses support an outstandingly high level of genetic divergence
13 at such a short geographical distance. Fixation index is high for genome-wide SNP data ($F_{\text{st}} =$
14 0.049 ± 0.0082 , averaged over 500-SNPs windows), and higher from mtDNA data ($F_{\text{st}}=0.158$).
15 Analysis of molecular variance also supports significant differentiation between Isla Choros and
16 Isla Pan de Azúcar. Based on SNP data, difference between islands account for 2.51 % of the
17 total variation, while difference between colonies within each island accounts for -0.099 %. Ac-
18 cording to mtDNA data, these components account respectively for 14.28 % and 0.95 % of
19 total variation. Differentiation between islands is significant according to both datasets ($\Phi_{\text{ST}_{\text{mtD}}}$
20 $\text{NA} = 0.152$ and $\Phi_{\text{ST}_{\text{SNP}}} = 0.0251$, both p-values < 0.05), but differentiation between colonies
21 within islands is not ($\Phi_{\text{SC}_{\text{mtDNA}}} = 0.0111$, p-value = 0.297; and $\Phi_{\text{SC}_{\text{SNP}}} = -0.00101$, p-value =
22 0.795).

23 Principal component analysis clearly supports a two-population structure (Fig. 44A). Although
24 the first component only accounts for 6.62% of total variance, it discriminates widely between
25 individuals from Isla Choros and Isla Pan de Azúcar - while the second and next components

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §154

1 only outline inter-individual variance. The idea of a low-level but highly consistent separation is
 2 also supported by a Hamming-distance-based neighbour net (Fig. 44B): terminal edges are much
 3 longer than structural inner edges, yet sorting is complete between the two islands (a structure
 4 supported, although to a lesser extent, by a mtDNA haplotype network, see Fig. 44D). Finally,
 5 clustering analyses, whether genotype likelihood based (Fig. 44C) or K-means based strongly
 6 supports a two-population model, both over panmixia, and over higher complexity models.



7 **Figure 44 | Consistent genetic structure within the Chilean range.** (A) Principal component analysis
 8 performed on genome-wide genotype likelihood data; (B) NeighbourNet built from nuclear SNP data; (C) Indi-
 9 vidual components of admixture as inferred from genome-wide genotype likelihood data (median values from 500
 10 bootstrap replicates); (D) Fitch-distance haplotype graph for the mitochondrial cytochrome-b gene (disc area pro-
 11 portional to the number of haplotype copies). In all panels, “AZ” stands for Isla Pan de Azúcar and “CH” for Isla
 12 Choros. AZ1 and AZ2, and CH1 and CH2 respectively stand for two colonies of each island.

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §155

1 *§-155 Isolation and migration analysis.* Migrate-n models of connectivity support a structure with
2 low migration between separated populations. Bayes factor (BF) model comparison across three
3 independent replicates allows unambiguous rejection of both panmixia, and full isolation hy-
4 potheses. BF does not allow for a clear choice between symmetric, and asymmetric migrations -
5 although BF does not take into account the number of estimated parameters: thus, the lower-
6 parametrised symmetric migration model should be preferred. Under this model, $\Theta_{\text{CHR}} = 0.00697$
7 $\pm 8.0829e^{-5}$, $\Theta_{\text{AZR}} = 0.00747 \pm 0.000212$ and $M = 2075 \pm 10.6$. Using the reconstructed substitu-
8 tion rate for highly polymorphic RAD loci $\mu = 3.55 \cdot e^{-8}$ substitutions.site⁻¹.generation⁻¹, inferred
9 effective population sizes scale to $N_{\text{CHR}} \approx 58,600$ and $N_{\text{AZR}} \approx 62,800$, with a symmetric migration
10 rate of $M \approx 3\%$ effective migrants per generation.

11 Joint minor-allele frequency spectrum based inferences are globally consistent with this model.
12 AIC supports our fifth model - a model with a recent change in the migration regime (median:
13 207 generations ago, CI 95% [82; 375]), but no population size change. Population sizes are
14 widely asymmetric (median sizes and CI95%; Choros: 91,350 [87,977 ; 95,509]; Pan de Azúcar
15 1,013 [1,010 ; 1,018]). Migration from Choros to Pan de Azúcar is inferred stable (before the
16 change: $1.146e^{-5}$ [$1.121e^{-5}$; $1.209e^{-5}$]; after the change: $1.170e^{-5}$ [$1.124e^{-5}$; $1.240e^{-5}$]), but increases
17 at least ten-fold from Pan de Azúcar to Choros (before the change: $2.756e^{-4}$ [$2.397e^{-4}$; $3.025e^{-4}$],
18 after the change: $4.268e^{-3}$ [$2.553e^{-3}$; $1.041e^{-2}$]). The second-best AIC score supports a model with
19 changes in migration regime and in population size: inferred migration patterns are similar
20 (Choros to Pan de Azúcar, before change: $1.135e^{-5}$ [$1.018e^{-5}$; $1.207e^{-5}$], after change $1.224e^{-5}$
21 [$1.122e^{-5}$; $2.226e^{-5}$]; Pan de Azúcar to Choros, before change: $2.853e^{-4}$ [$2.466e^{-4}$; $3.128e^{-4}$], after
22 change: $1.103e^{-2}$ [$3.724e^{-3}$; $1.116e^{-1}$]), and population size in Pan de Azúcar is stable or increases
23 slightly (before change: 1,014 [1,010; 1,019], after change: 1,027 [1,011; 17,702]), but popula-
24 tion is reduced strongly in Choros (before change: 95,542 [88,815; 101,124], after change:
25 23,832 [10,068; 36,104]).

1 *Discussion*

2 §-156 *A highly structured population.* The Chilean populations of *Pelecanoides garnotii* stand out as
3 remarkably isolated, despite short geographical distances and a tumultuous recent history. All
4 available evidence supports the idea of a distinct population per island, with limited gene flow.
5 Genome-wide polymorphism based analyses concur on a complete sorting of individuals between
6 Isla Choros and Isla Pan de Azúcar, albeit with limited differentiation: the variance accounted for
7 by the first component of PCA (6.62%), although highly consistent with the geographical struc-
8 ture, is only slightly above the neutral expectation. This is fully in line with the relative lengths of
9 internal and terminal edges, and the location sorting, in the neighbour-net (Fig. 44B). There is a
10 good agreement between the genome-wide nuclear and the mitochondrial signal: some aspects,
11 however, deserve consideration. The clearer sorting of individuals observed from nuclear data
12 compared to mtDNA (see *e.g.* Fig. 44B *vs* Fig. 44D) is most likely a consequence of the much
13 higher resolution of the large neutral nuclear marker dataset, compared to the relatively short,
14 and non-recombining, functional cytochrome-b gene sequence. For F-statistics based analysis
15 (pairwise F_{st} , or analysis of molecular variance), on the other hand, mtDNA gives a stronger sig-
16 nal than nuclear markers. Although higher dispersal in males than in females has often been pro-
17 posed as an explanation in similar cases, this interpretation of discrepancies between nuclear and
18 mitochondrial signals has been questioned (Zink & Barrowclough 2008), and the difference in
19 coalescence rate of haploid and diploid markers has been convincingly proposed instead. In our
20 case, the much larger sample size used for mtDNA analysis (83 individuals) compared to nuclear
21 analysis (21 individuals) may also account for a large part of this apparent difference. The resolu-
22 tion and the strength of the signal, however, should not be confused: the weaker, but clearer
23 nuclear signal may characterise the true isolation processes more accurately than the larger and
24 more contrasted mitochondrial dataset (Hoban *et al.* 2013b).

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §157

1 *§-157 Population size and gene flow.* Reconstructed population sizes and gene flow between sam-
2 pled locations are globally consistent between haplotype-based and spectrum-based approaches,
3 although with a considerably higher precision for the spectrum-based inferences. Migrate-n and
4 fastsimcoal2 models allow us to reject panmixia as well as full isolation between colonies, and
5 support partial isolation, with ongoing gene flow. However, the models differ on two important
6 points. First, Migrate-n suggests equal population sizes, with $N_e \approx 60,000$ for each population,
7 while fastsimcoal2 infers unequal (yet stable) population sizes, with $N_{CHR} \approx 90,000$ and N_{AZR}
8 $\approx 1,000$. Despite these contrasting results, however, both models propose an effective population
9 size that is considerably higher than the present-day census: the cumulated population of
10 $\sim 90,000$ to $\sim 120,000$ effective breeders is in stark opposition with the currently observed popula-
11 tion of $\sim 19,000$ breeders in Isla Choros (Martinez Palma 2014) and ~ 220 in Isla Pan de Azúcar
12 (Mattern *et al.* 2002), but more in tune with the historical populations (*e.g.* 100,000 breeders on
13 the major breeding location of Isla Chañaral in 1938, and similarly large colonies elsewhere,
14 Araya Modinger & Duffy 1987): a dire testimony to the rapidity and extent of the population
15 collapse in the past decades (Roman & Palumbi 2003).

16 Second, Migrate-n infers symmetric, and relatively high gene flow ($\sim 3\%$ per generation), whereas
17 fastsimcoal2 supports much lower, asymmetric gene flow: northward migration (from Isla
18 Choros to Isla Pan de Azúcar) is stable and very low ($\sim 0.001\%$ per generation), while southward
19 migration is higher, and increases from $\sim 0.03\%$ to $\sim 0.4\%$ in the recent past. Although the histor-
20 ical records are too scarce to draw any definitive conclusion, it is noteworthy that on Isla Pan de
21 Azúcar, a thousands of decaying, unoccupied burrows can still be observed (Fernandez *in prep.*) -
22 a possible sign of one or several recent massive emigration events. Thus, generally, the haplotype-
23 based reconstruction suggests a “homogenised” system, with population sizes averaged by higher
24 gene flow: in contrast, the spectrum-based approach offers a more nuanced reconstruction, that
25 better reflects the observed state of the populations. The higher resolution of the spectrum-based
26 approach, however, is expected. First, it is able to use the information from the full RADome,

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §158

1 and not only from a random subset of loci. Second, haplotype-based methods chiefly capture
2 events occurring at a substitution (*i.e.* millennial) scale, while spectrum-based approaches also
3 capture drift in allele frequencies that may occur in a few generations, thus making them more
4 suitable for the recovery of recent population history. Therefore, we may more appropriately con-
5 sider Migrate-n reconstructions as a general framework, that does not contradict the more precise
6 reconstruction given by our fastsimcoal2 model.

7 *§-158 Evolution of philopatry.* The high differentiation between two assessed sites at such close
8 range is remarkable for seabirds in general (Friesen *et al.* 2007), but all the more so for the guano-
9 seabirds (“*aves guaneras*”) of the Humboldt Current System. The observed structure is in strong
10 contrast with the HCS-wide panmixia of the Peruvian booby (*Sula variegata*, see Taylor *et al.*
11 2010) or the Peruvian pelican (*Pelecanus thagus*, see Jeyasingham *et al.* 2013). Weak genetic struc-
12 ture has indeed also been detected in the Humboldt penguin (*Spheniscus humboldti*), but only at
13 a range of ~2,000 km (Schlosser *et al.* 2009), as opposed to the ~350 km separating the two as-
14 sessed diving-petrel breeding islands. Since both the Peruvian pelican and the Humboldt penguin
15 share with the Peruvian diving-petrel a recent history of dramatic population decline, the contrast
16 in population structure can hardly be explained by their recent collapse. Morphological and
17 metabolic aptitude to flight may, on the other hand, influence dispersal patterns - indeed, it has
18 been proposed that the diving-petrel’s adaptation to underwater wing-propelled swimming, and
19 the necessary trade-off of poor flying abilities (Luna-Jorquera *et al.* 2003), reduced the species’ ac-
20 tivity radius. Although no data is available for the Peruvian diving petrel, its two sister species,
21 the Common diving petrel (*P. urinatrix*) and the South Georgian diving petrel (*P. georgicus*), have
22 been shown to forage more locally than similar-sized, sympatric “gliding” Procellariiformes, both
23 during the breeding season (Navarro *et al.* 2013), and in the inter-breeding period (Navarro *et al.*
24 2015). Further studies, however, will be necessary to decide how far the lack of dispersal ability is
25 involved in the Peruvian diving petrel’s genetic fragmentation. An alternative explanation may be
26 purely behavioural: while philopatric behaviour is the norm in Procellariiformes (see *e.g.* Milot *et*

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §159

1 *al.* 2008), the higher investment needed for burrow construction (Furness 1978; Cortés Labra
2 2007), together with the scarcity of available habitat, may reinforce that behaviour in the diving
3 petrel.

4 *§-159 Implications for conservation.* The observed level of population isolation has important im-
5 plications for the conservation of the species. Whereas the panmictic character of the HCS popu-
6 lation of Peruvian pelican or Peruvian booby has been suggested to increase their robustness in
7 the face of environmental change (Jeyasingham *et al.* 2013, see also *The Emperor synnome*, p.
8 211), the high fragmentation of the *P.garnotii* population, on the other hand, is expected to add
9 to their already great vulnerability to land habitat destruction and marine environment degrada-
10 tion. Indeed, high philopatry and low dispersal imply that the loss of a combined safe breeding
11 island and productive foraging area can not be mitigated by mass dispersal, but will rather lead to
12 a local extinction, and the irremediable loss of the associated genetic diversity. An associated pre-
13 diction is that current populations are demographically independent (Waples & Gaggiotti 2006).
14 Therefore, the observed local trajectories are unlikely to be an artifact of labile migratory fluxes,
15 but rather represent the true local demographic trend: thus, the local census increases that are ob-
16 served (in particular in Isla Choros, Martínez Palma 2014) may be interpreted as the positive
17 outcome of successful conservation measures, such as the creation of the Humboldt Penguin Na-
18 tional Reserve (that includes Isla Choros and the surrounding marine area) - an encouragement
19 to persevere in such endeavours.

20 Our results make it especially clear that any conservation strategy in the Humboldt Current Sys-
21 tem should focus on each island (and the surrounding marine areas, Luna-Jorquera *et al.* 2012) as
22 a significant unit. Within islands, colonies do not stand out as genetically separate entities, al-
23 though philopatric behaviour may still apply at that scale for stretches of a few generations. But
24 the distinctness of each island's gene pool, as well as their inferred demographic independence,
25 makes the conservation of each single one of the four remaining Chilean breeding islands an ab-

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §161

1 solute necessity for the survival of this once extremely abundant, and now critically endangered
2 seabird species. Indeed, despite the positive trend observed in the last decade on Isla Choros, im-
3 portant threats remain both for the nesting sites and for the marine habitats of diving petrels.
4 One of the four islands where petrels nests (Isla Pájaros II, currently with no legal protection) is
5 still under a valid permit of guano extraction (Ministerio de Minería 1988), and ongoing extrac-
6 tion of guano was confirmed on the island as recently as 2003 (Simeone *et al.* 2003). Additional-
7 ly, future plans for mining projects include the building of a large harbour in the close vicinity of
8 the two central breeding locations of the species, Isla Choros and Isla Grande de Atacama (Servi-
9 cio de Evaluación Ambiental 2015). This could significantly affect both the coastal foraging areas
10 of the species (through increased marine transit) and nesting areas (through direct effects of in-
11 creased human intrusion). However, our results clearly show that the conservation of each single
12 breeding location is essential for preserving the full remaining genetic diversity of this already sev-
13 erely reduced species, and maximizing the chances of a recovery after its massive historical
14 decline.

15 *Material and Methods*

16 *§-160 Sample collection and DNA extraction.* Access to the protected areas and sampling were per-
17 formed under permit N°38/2012, delivered by the CONAF authority (Atacama region). Sam-
18 pling was performed between april and may, outside the reproductive peak season (Fernández
19 com. pers.) to minimise disturbance.

20 *§-161 Sampling and DNA extraction.* Four colonies of *Pelecanoides garnotii* were sampled
21 throughout its Chilean breeding range: two in Isla Choros (Fig. 43 - Ch1 and Ch2: 32 and 11
22 individuals respectively), and two in Isla Pan de Azúcar (Fig. 43 - Az1 and Az2: 27 and 11 indi-
23 viduals respectively). Adults were captured using a mist net placed at the entrance of the colony
24 at dusk. Blood was sampled from the *retia* in the interdigital membrane of the foot, using a he-

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §163

1 parinized microcapillary, dried, and frozen. DNA extraction was performed using a standard
2 spin-column protocol (Qiagen DNeasy® Blood & Tissue kit).

3 *§-162 Mitochondrial marker sequencing.* A fragment of the mitochondrial cytochrome-b gene was
4 amplified for all samples using either two or four specific primers designed from the published se-
5 quence (Nunn & Stanley 1998 - as 5'-3', yun1f: GCCCCAAACCTCCGAAAATCCCA and
6 yun2r: GGTGATGGAGGCTAGTTGGCCG or with internal primers yun1r: GCCT-
7 GATTCGTGAAGGAAGGTGAGG and yun2f: CCACCCTAACCCGATTCTTCGCC). Am-
8 plification required the following conditions: 5' at 94°C, followed by 35 cycles of 30" at 94°C,
9 30" at 57°C, and 1' at 72°C, and finally 5' at 72°C. PCR product was then purified using Illus-
10 tra™ ExoStar™, and Sanger sequencing was performed at the ABI lab of the University of Oslo.
11 Quality assessment, trimming and manual checking were performed in Geneious® v6.1.2.

12 *§-163 RAD library preparation and SNP typing.* SNP discovery and sequencing was performed on
13 a subset of individuals (11 from Isla Choros and 10 from Isla Pan de Azúcar) following a single-
14 digest RAD-sequencing protocol (Baird *et al.* 2008). High quality samples were selected using a
15 1.5% agarose gel, and quantified by fluorometry (Life technologies™ Qubit®). (i) for each sample,
16 125ng of genomic DNA were digested with the restriction enzyme SbfI (NEB); (ii) each sample
17 was ligated to a unique barcoded P1 adapter prior to pooling in a single library. The library was
18 then sheared by sonication (7 cycles 30" ON – 30" OFF); (iii) sonicated libraries were concen-
19 trated to 25 µl by DNA capture on magnetic beads (beads solution:DNA = 0.8:1), and the target
20 size range fraction (300-500 bp) was then selected by gel electrophoresis and manual excision;
21 (iv) capture on magnetic beads using the same beads:DNA ratio (0.8:1) was then performed in
22 all following purification steps (after blunt-end repairing, poly-A tailing, P2 adapter ligation and
23 library enrichment by PCR); (vi) PCR amplification was performed in 8 x 1 2.5 µl aliquots
24 pooled after the amplification in order to reduce amplification bias on few loci due to random
25 drift. PCR was performed using NEB Phusion® polymerase with the following cycles: 30" denat-

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §163

1 uration at 98°C, 21 cycles of amplification (10" at 98°C, 30" at 65°C and 30" at 72°C), and a fi-
2 nal elongation of 5' at 72°C; (vii) the library was then quantified by fluorimetry (Life technolo-
3 gies™ Qubit®) and molarity was checked on an Agilent Bioanalyzer chip (Invitrogen™). A final
4 volume of 20 µl was submitted for paired-end sequencing on an ILLUMINA HiSeq2000 se-
5 quencer (V3 chemistry) at the Norwegian Sequencing Centre, University of Oslo, and spiked
6 with 20% PhiX control library in order to reduce low-diversity bias.

7 As no reference genome is available for *P. garnotii*, data processing followed a two-step protocol.
8 First, loci were built *de novo*, using the Stacks v1.28 pipeline (Catchen *et al.* 2011; Catchen *et al.*
9 2013), and a synthetic reference RADome was built. Second, this reference RADome was used to
10 perform alignment-based SNP typing. The workflow was as follows: (i) Sequence de-multiplex-
11 ing and *de novo* locus assembly was done according to in-line barcodes using the Stacks *deno-*
12 *vo_map.pl* pipeline, with a maximum of 5 mismatches allowed between alleles at a single locus
13 (both within and between individuals). The corresponding paired-end contigs were assembled us-
14 ing Velvet (Zerbino & Birney 2008) and Stacks, setting a minimum contig length of 200bp in
15 order to filter out multiple non-overlapping short paired-end reads. (ii) Resulting loci were con-
16 catenated, together with their respective paired-end contigs, to form the scaffolds of a reduced
17 reference genome, or RADome. At that point, loci comprising multiple paired-end contigs (a po-
18 tential sign of collapsed paralogous loci) were removed. The raw paired-end fastq files were then
19 mapped onto this RADome using Bowtie2 (Langmead & Salzberg 2012). The resulting align-
20 ments were processed using Samtools 0.1.19 (Li & Durbin 2009), PicardTools 1.113 ([http://pi-](http://picard.sourceforge.net)
21 [card.sourceforge.net](http://picard.sourceforge.net)), and custom R and shell scripts in order to discard orphaned reads and low-
22 quality pairs. These cleaned alignments were further filtered for PCR and optical duplicates using
23 Picard MarkDuplicates. (iii) SNP and genotype calling was restricted to the first-in-par reads. We
24 then used GATK's HaplotypeCaller pipeline (DePristo *et al.* 2011) to simultaneously call SNPs
25 and genotypes in all samples, using first reads only, with standard parameters, except for popula-
26 tion heterozygosity which was set to 0.01, and retaining only di-allelic, non-indel sites sequenced

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §165

1 in at least 75% of samples. We repeated key analyses using ANGSD 0.900 (Kim *et al.* 2011; Korneliussen *et al.* 2014) to estimate allele frequency and genotype likelihoods, without hard-calling
2 genotypes, an approach that has been shown to be much more robust for low-coverage data
3 (Nielsen *et al.* 2012; Korneliussen *et al.* 2014).

5 *§-164 Analysis of population structure.* Summary statistics were calculated in DnaSP (Rozas &
6 Rozas 1999) for mtDNA, and R package *adegenet* (Jombart 2008) and custom scripts for SNP
7 data. Analysis of Molecular variance was performed for both datasets using Arlequin v3.5 (Ex-
8 coffier *et al.* 2013) (in a locus-by-locus rationale for SNP data, and haplotypic for mtDNA, and
9 15,000 permutations to assess significance levels). A neighbour-net was built in SplitsTree (Hu-
10 son & Bryant 2006) either directly from the mtDNA sequences, or from a matrix of pairwise
11 Hamming distances, as calculated in PLINK v1.9 (Purcell *et al.* 2007) from SNP data. Principal
12 component analysis was performed on genome-wide data either using called genotypes in *ade-*
13 *genet*, or raw genotype likelihoods in ngsTools (Fumagalli *et al.* 2014). In order to understand the
14 contribution of individual loci to the overall structure, we performed a discriminant analysis of
15 principal components (Jombart *et al.* 2010), as implemented in *adegenet*. Admixture between
16 populations was estimated using two different approaches. First, we used a classical model-based
17 clustering approach relying on the hypothesis of Hardy-Weinberg equilibrium (HWE) (Pritchard
18 *et al.* 2000), as implemented with an empirical Bayes algorithm either in fastStructure (Raj *et al.*
19 2014) for called genotypes, or in ngsAdmix (Skotte *et al.* 2013) for raw genotype likelihoods, re-
20 taining only sites that did not violate HWE, with a number components ranging from 1 to 4,
21 and 100 bootstrap replicates. Model complexity was chosen using Evanno's ΔK method (Evanno
22 *et al.* 2005). Second, in order to assess the impact of the HWE assumption, we used a model-free
23 approach, as implemented in the *adegenet* k-means clustering (Liu & Zhao 2006).

24 *§-165 Model-based estimation of population history.* We relied on two separate coalescent-based ap-
25 proaches to estimate the amount of gene flow between the North and the South of the species

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - §165

1 Chilean range. First, population sizes and migration rates were co-estimated based on multilocus
2 haplotype data in a Bayesian framework, as implemented in Migrate-n (Beerli & Palczewski
3 2010). We selected three sets of 50 random nuclear loci with 4 to 6 polymorphisms, as an unbi-
4 ased representation of the neutrally evolving part of the genome (Trucchi *et al.* 2014). In order to
5 correct for potential allele dropout, we randomly sampled one haplotype only for each indi-
6 vidual, at each locus. We ran a cold chain and 3 heated chain under a static heating scheme, rais-
7 ing the cold chain to a power of 1.5, 3 and 1e6, and proposing chain swapping every 100 steps.
8 Chains were run for 50,000,000 generations, recording every 500 generations, with a 5,000,000-
9 generation burn-in, specifying uniform priors both for population sizes (Θ , bounded between 0
10 and 0.1 with a δ of 0.01), and for the migration rates (M , bounded at 4,000 with a δ of 400).
11 Proper mixing under these conditions was ensured using the highest parametrisation model, and
12 checking convergence between the three independent random datasets. Models were ordered by
13 log Bayes factors, as defined by Kass and Raftery (Kass & Raftery 1995), u sing thermodynamic
14 integration approximation of marginal likelihoods. Four models were tested: (1) the two islands
15 are fully isolated populations, (2) the two islands are independent population exchanging
16 migrants with symmetric gene flow, (3) the two islands are independent population exchanging
17 migrants with directional gene flow, and (4) the two islands belong to a single, panmictic
18 population.

19 Co-estimation of population size, gene flow, and population history was performed through ex-
20 plicit model testing in fastsimcoal2 (Excoffier *et al.* 2013), through Akaike Information Criterion
21 (AIC) comparison of the optimised composite likelihood of the two-dimensional folded allele
22 frequency spectrum under six different models. We first tested two simple models: (1) a full isola-
23 tion model with two stable, independent populations, and (2) a classical isolation-with-migration
24 model, with population sizes, migration rates and divergence time as free parameters. We then
25 tested four more complex scenarios (3-6), in which the two population diverge after the last
26 glacial maximum, ca. 7,000 generations ago (a conservative biogeographic estimate for most

Unexpected philopatry in an insular seabird, the Peruvian diving-petrel - \$165

1 taxa). Under model (3), human activities have no impact on the populations, and constant gene
2 flow is maintained between the islands until the present, with no population size changes. Under
3 models (4), population size do not change, but gene flow is disrupted at some point during the
4 past, presumably by the expansion of guano exploitation. Under models (5) and (6), population
5 size changes during the past, and either gene flow remains constant (model 5) or gene flow also
6 changes (model 6). We generated 50 nonparametric spectrum bootstrap replicates, and per-
7 formed 50 parallel runs for each model and each bootstrap replicate, keeping only the one with
8 the highest composite log-likelihood. Each run necessitated a maximum of 80 ECM cycles, sim-
9 ulating 100,000 spectra at each step.

10 Models were calibrated using the general background genomic mutation rate for waterbirds,
11 $1.6e^{-3}$ substitutions.site⁻¹.Myr⁻¹, as established from a wide panel of genomic data by Zhang *et al.*
12 (Zhang *et al.* 2014), or $\mu = 9.6 e^{-9}$ substitutions.site⁻¹.generation⁻¹ considering a generation time
13 of 6 years, as suggested by Welch *et al.* 2011. Migrate-n models, however, were based on loci with
14 4 to 6 polymorphic sites: these can not be calibrated directly using the general background sub-
15 stitution rate. Considering that the RADome is assumed to evolve neutrally, the number of
16 polymorphic sites per locus is expected to follow a Poisson distribution, and the background sub-
17 stitution rate μ applies for the expected value λ of that distribution. Therefore, after verifying the
18 goodness-of-fit of that the Poisson model, we extracted the expected substitution rates for loci
19 with $S=4$, $S=5$ or $S=6$ SNPs, as $(S/(\lambda\mu))$, and used their weighted average to calibrate our
20 reconstructions.

1 *Acknowledgements*

2 We thank Diego Miranda-Urbina and Nicole Licuime Castillo for their invaluable help on the
3 field, as well as Rasmé Heremé at the CEAZA, Morten Skage, Marianne Selander Hansen and
4 Ave Tooming-Klunderud of the Norwegian Sequencing Center for their assistance in the the lab-
5 oratory. This work was conducted with the support of the French National Research Agency
6 (ANR) “PICASO” grant (ANR-2010-BLAN-1728-01), from Marie Curie Intra European Fel-
7 lowships (FP7-PEOPLE-IEF-2008, European Commission; project no. 235962 to CLB and
8 FP7-PEOPLE-IEF-2010, European Commission; project no. 252252 to ET), from the Centre
9 Scientifique de Monaco through budget allocated to the Laboratoire International Associé 647
10 ‘BioSensib’ (CSM/CNRS-University of Strasbourg). This work was performed on the Abel Clus-
11 ter, owned by the University of Oslo and the Norwegian metacenter for High Performance Com-
12 puting (NOTUR), and operated by the Department for Research Computing at USIT, the Uni-
13 versity of Oslo.

1 Chapter 8: Discussion

2 *Population structure in colonial seabirds*

3 *§-166 Aptenodytes penguins are synnome species.* The most consistent and unexpected aspect of our
4 results is the similar, and original structure of the worldwide Emperor and King penguin popula-
5 tions. Both species have circumpolar, zonal distributions, with large and well defined colonies,
6 that are either homogeneously distributed along the coast of Antarctica, for the Emperor pen-
7 guin, or clustered together on the subantarctic archipelagos, for the King penguin. As exposed in
8 *The King synnome*, p. 163 and *The Emperor synnome*, p. 211, within these circumpolar distribu-
9 tions, each species is made up of one single, range-wide panmictic population, with all evidence
10 consistently and unambiguously converging to that conclusion. Interestingly, however, in both
11 species, principal component analysis does show a hint of sorting by distance. In the Emperor
12 penguin (Fig. 36A), a rough gradient is visible between the colonies of the Adélie Land area and
13 those of the Dronning Maud Land area, that may correspond to an underlying isolation-by-dis-
14 tance pattern. In the King penguin (Fig. 23), although no firm structure can be extracted, indi-
15 viduals from the same archipelago are consistently grouped together - even though groups over-
16 lap largely. Yet in both species, the variance explained by the first p rincipal c omponents is
17 extremely low: the first component explains 1.39% of the total variation in the Emperor pen-
18 guin, and 0.93% in the King penguin - in both cases, this is barely above the expectation for full
19 homogeneity. This observation, instead of supporting the idea of structured metapopulations,

Discussion - §166

1 makes a case for the importance of taking the full evidence into account when describing a
2 species' genetic architecture. Restricting the analysis to the sole observations that support some
3 kind of genetic structure could allow us to suggest that an isolation by distance process is at work
4 in both species: however, extensive model-based testing clearly shows that this is not the case.

5 In the Emperor penguin, the slightly more gradual geographical pattern does retain information
6 as to the intensity and directionality of gene flow between colonies: both haplotype-based and
7 spectrum-based approaches converge on similar estimates. In the King penguin, however, the
8 even higher extent of panmixia obliterates any quantitative information as to the intensity of
9 gene flow: model-based approaches hardly converge, but propose very high and seemingly ran-
10 dom values for all migration parameters - an expected behaviour, since full mixing can inform a
11 lower bound, but not an upper bound on gene flow parameters. The slight difference in intensity
12 of genetic mixing between the two species may be interpreted in several ways. First, it may be an
13 artifact of our sampling design: our geographical representation of the King penguin being far
14 more complete than that of the Emperor penguin, it is possible that our view of the latter is over-
15 ly contrasted by the artificial break in the sampling distribution (between the Adélie Land and
16 the Dronning Maud Land groups), and that sampling intermediate colonies would bring the en-
17 semble closer to a mixed whole. Indeed, the four sampled Ross Sea individuals hint to that direc-
18 tion, as they do not seem to fit in even a slight isolation-by-distance pattern (see Fig. 36A). Al-
19 ternatively, the contrasting demographic histories of the two species may have played a role in
20 this difference. Whether the Pleistocene bottleneck in the King penguin (see *The King synnome*,
21 p. 163) «reset» a state of full panmixia 20,000 years ago that light drift could not yet disrupt, or
22 rather (a more likely hypothesis in our opinion) the 20th century history of massacres and the
23 subsequent recovery (see §22 p. 75) involved a very recent increase in gene flow for the recoloni-
24 sation of the most severely affected archipelagos, it seems certain that the King penguin had a
25 more tumultuous history than the Emperor - which could justify even slightly higher genetic
26 mixing. However, we can not insist enough on the fact that, for both species, any kind of struc-

Discussion - §167

1 ture signal is far below significance - and that interpreting differences at that level may be perfect-
2 ly vain.

3 Nevertheless, field observations and genetic data agree at the local, and the global scales. In the
4 King penguin, equally high return rates have been evidenced at the level of the nest (Barrat 1976;
5 Bried & Jouventin 2001) or the colony (Le Bohec *et al.* 2008; Saraux *et al.* 2011b): yet genetic
6 structure is equally absent at both levels. A simple explanation seems, as we proposed in §71 p.
7 157, the difference in standpoint when considering return rate from an ecological or a genetic
8 perspective. Indeed, if 90% precise local return will doubtlessly be considered a case of extreme
9 philopatry by any field ecologist, because of the allowance made for mortality at sea, and of the
10 one-sided character of the implicit test, this observation may be reversed as up to 10% of yearly
11 dispersal: a flow that is high enough to obliterate any hint of populations structure. This bias, the
12 «*seabird paradox*» of Milot *et al.* 2008, is in fact only a paradox as long as we do not really exam-
13 ine the bridge between the generation-scale and the evolutionary scale: a phenomenon that seems
14 negligible when considering the lifespan of a few individuals becomes a major process on the evo-
15 lutionary level. Moreover, the empirical case of the Emperor penguin (see *Empirical evidence of*
16 *heterogeneous dispersal*, p. 259) provides us with additional indications on how heterogeneity in
17 dispersal probability can be organised in time: the likely splitting of the Mertz colony following
18 changes in the local icescape, and the repercussions on the neighbouring colony of Pointe Géolo-
19 gie, lift the veil on the events of catastrophic dispersal that may punctuate the history of penguin
20 colonies, and maintain genetic mixing even in the case of temporarily reduced gene flow.

21 *§-167 Genetic landscapes in the Southern Ocean.* As exposed in introduction (see §19 p. 68), most
22 seabird species exhibit a certain degree of phylogeographic structuring, and species with a world-
23 wide distribution are all the more likely to have marked demes - at least one per ocean basin
24 (Friesen *et al.* 2007). In the Southern Hemisphere, and especially at higher latitudes where inter-
25 basin circulation is high (see §6 p. 33 and §7 p. 35) and biogeographic zones often have a strong

Discussion - §167

1 zonal component (see §10 p. 48), most species retain a colonial circumpolar distribution similar
2 to that of the *Aptenodytes* penguins. Yet a marked genetic structure normally evolves, and may
3 result in zonally separated populations that sometimes reach the subspecies level - as in the Giant
4 petrel *Macronectes giganteus*, in which a Northern and a Southern form have been identified (as
5 *giganteus* and *halli*, see Techow *et al.* 2010), or most commonly location-specific populations, as
6 in also the case in the Giant petrel (Techow *et al.* 2010), the Snow petrel *Pagodroma nivea* (Bar-
7 braud & Jouventin 1998), or the lesser Sheathbill *Chionis minor* (Viot *et al.* 1993). More remark-
8 able is the evolution of fully separate species along the circumpolar range. This is for example the
9 case of the Wandering Albatross *Diomedea sp.* species complex, in which at least three different
10 clades can be distinguished, each with a full species status - *D. exulans* in Crozet, Kerguelen, Mar-
11 ion & Prince Edward, and South Georgia archipelagos and Macquarie Island, *D. dabbeni* on
12 Tristan da Cunha, and *D. antipodensis* on Antipodes, Adams and Campbell islands (Burg &
13 Croxall 2004): a structure that is made more surprising by the fact that Wandering albatross easi-
14 ly circumnavigate the Southern Ocean during one foraging season (Croxall *et al.* 2005). The
15 Southern Skua *Catharacta sp.* species complex offers a similar picture, with both zonal and regio-
16 nal differentiation between species (*C. maccormicki* in Antarctica, *C. antarctica lonnbergi* in the
17 subantarctic belt, *C. a. antarctica* in the Falklands and the Patagonian shelf area, *C. a. hamiltoni*
18 in Gough and Tristan da Cunha, and *C. chilensis* in Tierra del Fuego and Patagonia), despite
19 ocean-wide seasonal migration (Weimerskirch *et al.* 2015), and ongoing gene flow at the contact
20 zones of species (Ritz *et al.* 2008). The Diving-petrel *Pelecanoides sp.* species complex, with *P.*
21 *urinatrix* and *P. georgicus* breeding in sympatry throughout the subantarctic belt, *P. magellani* in
22 Tierra del Fuego, and *P. garnotii* along the Humboldt Current (Shirihai & Kirwan 2008; Onley
23 & Scofield 2013), also offers a clear example of local isolation both through distance, and
24 through foraging niche specialisation (Navarro *et al.* 2013). The Common diving petrel *P. urinatrix*
25 *trix* is itself separated in six subspecies that follow geographical areas (*P. u. urinatrix* in Australia
26 and Northern New Zealand, *P. u. chathamensis* in the southern New Zealand archipelagos, *P. u.*

1 *exsul* in South Georgia and the Southern Indian and Pacific archipelagos, *P.u. dacunhae* in Tris-
2 tan da Cunha and Gough, *P.u. berard* in the Falklands, and *P.u. coppingeri* in Southern Chile,
3 see Onley & Scofield 2013). In the present study on the Peruvian diving-petrel *P.garnotii* (see
4 *Unexpected philopatry in an insular seabird, the Peruvian diving-petrel*, p. 275), we demonstrate
5 how this oceanic-scale structure in diving-petrels is likely the consequence of very strong local
6 philopatry: even at the scale of a few hundred kilometers, populations may evolve near-complete
7 isolation. In *Pelecanoides* petrels, genetic structure therefore appears to organise itself in a complex
8 hierarchical system, with very closely related and partly sympatric species subdivided in sub-
9 species (as in *P. urinatrix ssp.*), and locally in strongly differentiated demes.

10 Penguins are no exception here, and phylogeographic structure has been identified in several
11 species across oceanic boundaries, often reaching subspecies or even speciation levels. The Little
12 penguin *Eudyptula minor*, for example, exhibits strong isolation between populations, and has
13 been divided six subspecies (Kinsky & Falla 1976), at least three of which are genetically mono-
14 phyletic (Peucker *et al.* 2009), and exhibit local isolation-by-distance patterns (Burrige *et al.*
15 2015). The Rockhopper penguin *Eudyptes chrysocome* has similarly been separated in a Northern
16 and a Southern group by De Dinechin *et al.* 2008, and the same applies for the Gentoo penguin
17 *Pygoscelis papua* (De Dinechin *et al.* 2012). Finally, two groups have been proposed in the Adélie
18 penguin, in the Ross Sea and in the rest of East Antarctica (Roeder *et al.* 2001). In the *Spheniscus*
19 *sp.* group, the geographically isolated *S. demersus* in southern Africa, *S. magellanicus* in Tierra del
20 Fuego and Patagonia, and *S. humboldti* in the northern Humboldt Current are identified as fully
21 separate species with traces of ongoing gene flow between the latter two species at their contact
22 point (Simeone *et al.* 2009). Similarly, the *Eudyptes* species complex may be considered as mostly
23 structured by allopatry between breeding islands, as species have very well defined, and usually
24 non-overlapping breeding range, with *E. pachyrhynchus* along the Fjorland coast of New Zealand,
25 *E. robustus* on the Snares islands, *E. sclateri* on Bounty and Antipodes islands, the extinct *E.*
26 *chathamensis* on Chatham island, *E. schlegeli* on Macquarie island, *E. chrysolophus* in the Southern

Discussion - §167

1 Atlantic Ocean and Crozet archipelago, *E. chrysocome chrysocome* around the Patagonian shelf
2 area, *E. c. moseleyi* on Gough, Inaccessible, Tristan da Cunha, Saint-Paul and Amsterdam islands,
3 and *E. c. filholi* on Marion & Prince Edward, Crozet, Kerguelen, Heard, Macquarie and the
4 Campbell Plateau area (Williams 1995; Borboroglu & Boersma 2013).

5 Considering this general picture, it appears that Southern Ocean seabirds (penguins included)
6 have a marked geographic structure of genetic diversity, where the intensity of gene flow will de-
7 termine the level of structure. Higher gene flow keeps groups on a purely intraspecific phyloge-
8 graphic level (as in *Eudyptula minor*), while a decreasing level of admixture gradually leads to sub-
9 species systems (as in the *Eudyptes chrysocome ssp.* or the *Pelecanoides urinatrix ssp.* complexes) and
10 full species systems (as in the *Diomedea sp.* or the *Spheniscus sp.* complexes). As suggested in our
11 introduction (see §6 p. 33), the balance between philopatry, dispersal, and at larger scale incipi-
12 ent local adaptations, appears to vary along a gradient of increasingly strongly structured bioge-
13 ographies. The continuous nature of this gradient, however, does raise the question of the signifi-
14 cance of observed structures. While full speciation may be rather easy to observe when
15 morphological, behavioural, and genetic cues converge, the lower reaches of phylogeographic
16 structure are more unclear (Waples 1998; Waples & Gaggiotti 2006), especially as genetic and
17 ecological separation entertain a rather blurred relationship in the lower differentiation areas (Es-
18 ler *et al.* 2006; Waples & Gaggiotti 2006). In those lower ranges, it may sometimes appear that
19 very subjective factors - such as publication pressure - influence the interpretation of observed
20 structures. Publication bias is a recognised confounding factor in several domains, including ge-
21 netics (Munafò *et al.* 2004; Pan *et al.* 2005), and usually takes the form of a bias towards report-
22 ing significant results over nonsignificant ones - or, more often, of reporting barely significant
23 results as indicative of a strong process. In the particular field of phylogeography and ecological
24 population genetics, it also often leads to the overinterpretation of barely significant results,

Discussion - §167

1 where otherwise insignificant signal is presented as having a tendency towards significance¹. In
2 seabirds, we may mention the study of Bouzat *et al.* 2009, in which colonies of Magellanic pen-
3 guin (*S. magellanicus*) amongst which 99% of variation was explained by inter-individual differ-
4 ences, with F_{st} levels below 0.01, are presented as a «metapopulation system»; or the study of
5 Freer *et al.* 2015, in which non-significant F_{st} values are presented as «*consistent with female bias*
6 *dispersal*». In these cases, the lack of a strong conceptual and hypothesis-testing framework often
7 results in misleading interpretations. The probability of observing similar values because of ran-
8 dom sampling variation is rarely examined, and a selection of statistically significant indicators
9 over statistically nonsignificant ones is common. All of these choices often result in overly con-
10 strated population structures - often easier to publish, but at the cost of losing the more nu-
11 anced aspects of the interspecific variations in demographic organisations.

12 In §137 p. 253, we discussed such a case, in a previous analysis of population structure in the
13 Emperor penguin, using mtDNA, by Younger *et al.* 2015, where a combination of biased sam-
14 pling and not fully reproducible analysis led to the conclusion that Emperor penguins were sepa-
15 rated in two different clades. Similarly, a pioneer study by Viot *et al.* 1993, using protein gel elec-
16 trophoresis and a short fragment of the cytochrome-b gene², concluded in a strong isolation
17 between Crozet and Kerguelen king penguins. Thus, prior findings using « classical » (*sensu* §28
18 p. 84) markers set the expectation of significant population structure in both species, on the tem-

1. This often comes down to a subjective decision as to the *side* from which results will be presented. For example, even if significant, an F_{st} value or 0.01 is strongly indicative of a high gene flow: but it is more often cited as «significant evidence for separation». This statement is not wrong in itself, since any given state of population connectivity is a balance between some degree of isolation, and some degree of mixing: but it is misleading, as it focusses on the minor isolation instead of the more meaningful high migration process.

2. As we discussed in §156 p. 286, the characteristics of « classical » mitochondrial markers also makes them likely to overestimate population structure, due to their halved effective population size, and consequently shallower coalescence time, compared to nuclear data. If this distinction is not kept in mind, mtDNA results can yield largely overestimated population structure.

Discussion - §168

1 plate of other penguins, or seabirds in general¹. Our results (see §166 p. 297) bring clear evidence
2 against both mtDNA-based studies.

3 *§-168 Synnomes and speciation.* Amongst all Southern Ocean seabirds for which genetic data has
4 been published, the two *Aptenodytes* species are thus, for the time being, the only two for which
5 no relevant population structure could be identified. In terms of biogeography, with their full
6 circumpolar distribution (Fig. 7), each of the *Aptenodytes* species should be more appropriately
7 compared to the full *Diomedea* species complex: in the latter, the small population sizes and high
8 philopatry of Wandering albatrosses has lead to full speciation between oceanic areas, whereas
9 both large penguin species have maintained a high connectivity and full homogeneity at the same
10 geographical scale. Indeed, preliminary analyses even emphasize the relatively low degree of diver-
11 gence between the King and the Emperor penguin: inter-individual intraspecific variance may be
12 almost of the same order of magnitude as interspecific variance, and shared polymorphism is im-
13 portant compared to fixed differences. Although the question of speciation and possibly incom-
14 plete lineage sorting between *A. forsteri* and *A. patagonicus* will require further investigation,
15 based amongst other things on whole-genome resequencing data², the present study support the
16 view of Gavryushkina *et al.* 2015, who place the divergence of the two *Aptenodytes* lineages in the
17 relatively recent past (1.52 million years ago, but possibly as recently as 730,000 years ago), at a
18 slightly more recent period than the divergence of the still-hybridising *Spheniscus magellanicus*
19 and *S. humboldti*. The question of the extent of prezygotic isolation between the two *Aptenodytes*

1. As a side note, the example of the evolution of laboratory and analysis methods over the past two decades should act as a twofold caution for present studies. First, conclusions drawn from less reliable material, or analysed using less robust approaches, should not systematically be taken for granted: re-analysis using more advanced methods may yield very different results. Second, it should also serve as a reminder to the fact that contemporary approaches may also become outdated soon enough, and that our conclusions only hold until then. We mentioned earlier how theoretical frameworks like the ARG are yet in their early development, both from a conceptual point of view, and from sheer technical limitations (*e.g.* the lack of an accurate and very-high-throughput sequencing platform, or the limits of contemporary computational power). It may well happen that further methodological and technical developments will lead us to reconsider our results - a somewhat tiresome, but absolutely unavoidable possibility.

2. Which we already produced, but still remains to be analysed, see §30 p. 86.

Discussion - §169

1 species if ever their ranges were to overlap, in particular, is currently unresolved, and will be of
2 particular interest for understanding the future of those species if ever climate change and subse-
3 quent range contraction brought them into close contact again.

4 The *synnome* structure of the two *Aptenodytes* species also raises questions as to the mechanisms
5 involved in their recent speciation. Indeed, reduced gene flow and local adaptation under allopa-
6 try is often invoked in the case of sister species currently occurring in allopatry, as is the case of
7 the King and Emperor penguins. However, the *synnome* structure that appears to be ple-
8 siomorphic in large penguins makes allopatric speciation a rather counter-intuitive event. A
9 strong disruption of this structure may have been necessary at the time of speciation: either
10 through colonisation of new habitats on the Antarctic continent, or through a sudden interrup-
11 tion of a previously much broader range (for example through changes in the structure of the
12 Southern Ocean water masses, although no such event has been documented yet in the strati-
13 graphic record).

14 *Seabirds on the evolutionary scale.*

15 *§-169 Inferring demography in synnome species.* The *synnome* structure of both species offers a
16 considerable advantage for reconstructing the demographic history of *Aptenodytes* penguins. In-
17 deed, population structure is a major determinant of genetic diversity: with a constant combined
18 census population size, fragmented populations undergo considerably more genetic drift than
19 one single panmictic population (see §39 p. 99), with an increased loss of rare alleles and a bias
20 towards intermediate allele frequencies in the full sample as a result. Thus, the subdivision of a
21 large population in two subpopulations, with low gene flow, even with no loss of individuals, has
22 essentially the same effect as a reduction in the total effective population size (Chikhi *et al.* 2010).
23 This is an important *caveat* for the inference of population history, since population size and
24 metapopulation dynamics are by no means independent parameters. Migration, for one, is often

Discussion - §169

1 a density-dependent parameter: a common prediction is that migration increases with density, *i.e.*
2 usually with population size (Sæther *et al.* 1999; Taylor & Norris 2007). Range reduction, on the
3 other hand, often results in habitat fragmentation, and consequently in population divergence
4 through reduced gene flow (Hanski 1998; Hanski & Ovaskainen 2000; Keller & Largiadèr
5 2003), but also in total population decrease (see *The King's Synnove*, p. 163) and associated loss of
6 genetic diversity (Arenas *et al.* 2012). Thus, there is an intricate connexion between census popu-
7 lation size change, population fragmentation, gene flow, and effective population size change,
8 and external (*e.g.* climatic) forcings are likely to influence all these parameters jointly. A typical
9 example of this difficulty is the European refugial structure that is often proposed for the Last
10 Glacial Maximum (Schmitt 2007). The Pleistocene continental ice sheet is thought to have
11 forced most species into fully isolated coastal refugia: concurrently, several studies have inferred
12 drastic population reductions during that period (*e.g.* in the Wild boar, see Groenen *et al.* 2012
13 or in the Brown bear, see Miller *et al.* 2012) - however, disentangling the respective roles of cen-
14 sus size reduction, and population fragmentation in refugial areas, in the observed effective popu-
15 lation size reduction, remains quite a challenge.

16 Although several studies have proposed, as a workaround, to start by establishing the genetic
17 structure of the extant populations, and then reconstruct the demographic history of each popu-
18 lation separately (see *e.g.* the remarkable work of Wallberg *et al.* 2014 on the honey bee *Apis*
19 *mellifera*), this is only possible when population structure has a clear-cut phylogeographic compo-
20 nent, that is to say, when populations have diverged in a tree-like manner, and remained isolated
21 ever since, and that divergence happened much before the demographic events of interest, so that
22 isolation and demographic processes do not interfere in shaping genetic diversity. A more realistic
23 case is that the degree of isolation between populations is fluctuating together with population
24 size and species density during the history of the population, and that the present-day population
25 structure, if ever it exists, does not fully reflect the past population structure - as groups that have
26 come in secondary contact may have totally re-shuffled in the meantime. Thus, suspicion of past

Discussion - §169

1 population structure may be an important hindrance to successful inference of past demography
2 solely from contemporary genetic data uninformed by, for example, paleontological or ancient
3 DNA data.

4 The *Aptenodytes* breeding system and genetic organisation is particularly favorable to past popula-
5 tion size inference. The present-day *synnome* structure appears to be rooted in behavioural charac-
6 teristics of the species (such as high dispersal ability) rather than being solely the result of recent
7 population history (see §167 p. 299): so that there is no compelling grounds to suspect that pop-
8 ulation fragmentation was ever an important trait in the past. However, we should be cautious
9 about the interpretation of strong population reduction events, such as the one observed during
10 the Last Glacial Maximum in the King penguin (see *The King synnome*, p. 163). What we really
11 reconstruct, in this study, is a reduction in *effective* population size, likely brought about by range
12 contraction and refugial fragmentation of the species. This explicitly involves a form of popula-
13 tion subdivision, and likely an associated reduction - even minor - in gene flow between refugial
14 areas. Thus, translating the observed reduction in *effective population size*, as a reduction in *census*
15 *population size* is not as straightforward as it would seem, even considering that the life history of
16 the species remained fully unchanged during the period. On the contrary, the observed restruc-
17 turation of genetic diversity during the Pleistocene is the joint consequence of all the forces we
18 mentioned earlier. Yet, our observed baseline (the present-day state of the species) allows us to go
19 further. With an effective population size of ~20,000 to 25,000 breeders, density is currently
20 such that gene flow precludes any population structure. Situation is likely to have been similar
21 prior to the LGM bottleneck, when inferred effective population size was strong. Thus, the state
22 change during the LGM gives us at least one crucial qualitative information: disruption of the
23 King penguin's habitat was strong enough to provoke a combination of population fragmenta-
24 tion and decline.

1 In contrast with the *Aptenodytes* synnomes, the very strong population structure that we observe
2 in a strongly philopatric seabird such as the Peruvian diving-petrel (see *Unexpected philopatry in*
3 *an insular seabird, the Peruvian diving-petrel*, p. 275) makes any attempt at inferring past demog-
4 raphy rather risky. Indeed, present-day isolation patterns are a convincing testimony to the be-
5 havioural capacity of the species to evolve strongly isolated structures: but it does not imply that
6 these structures have remained the same during the whole history of the species. Thus, even if we
7 restricted ourselves to one single present-day colony, that did not bear any signs of internal frag-
8 mentation, we would not be able to ascertain whether that population was, at any time in the
9 past, split into several units and later recast into its present shape. It is possible that further
10 methodological developments will alleviate that important limitation: it has been argued that the
11 full ancestral recombination graph (see §42 p. 107) should contain enough information to recon-
12 struct not only past population size, but also past population structure and admixture patterns.
13 An attempt has been made in that direction by Schiffels & Durbin 2014, although still within
14 the Sequential Markovian Coalescent framework - that is to say, still without taking the topology
15 of recombination events into account, which effectively limits reconstructions to tree-like diver-
16 gence patterns. We suggest that integrating the local recombination topology may be more infor-
17 mative as to past admixture events, since the probability of two given haplotypes merging
18 through recombination at any point of time is conditional on their probability of co-occurring in
19 a given deme. However, this would imply stepping out of the SMC framework, and would prob-
20 ably require considerable computation effort (Song & Hein 2005; Rasmussen *et al.* 2014). Al-
21 ternatively, within the multiple-SMC framework, reconstructing ancestral states at each node of
22 each inferred independent local tree, and performing admixture analysis (*e.g.* through K-means
23 clustering, see Jombart *et al.* 2010) for each reconstructed state at ancestral points of time may
24 give us an insight into the evolution of admixture complexity in along the past history of the
25 species. However, this is a broad endeavour, an far beyond the scope of our study. In the mean-
26 time, the wisest course would appear to be (*i*) to safely apply demographic reconstruction meth-

Discussion - §170

ods only to essentially panmictic species such as the *Aptenodytes* penguins, and (ii) when dealing with structured populations such as can be found in *Pelecanoides* species, to either avoid such approaches altogether, or to restrict the analysis to a single population and to stay aware of the considerable uncertainty that propensity to genetic isolation imposes on the results.

§-170 Pleistocene history of the Aptenodytes penguins. The various methods we applied in this work to reconstruct the demographic history of the King and Emperor penguin (in *The King synnome*, p. 163 and *The Emperor synnome*, p. 211) showed a satisfactory degree of consistency, (although all did not perform as accurately for the more complex history of the King penguin, see §98 p. 192): we can thus obtain a general picture of the recent past of the two species with relative confidence.

According to our reconstruction, the Emperor penguin population did not vary significantly during the past 100,000 years - a trait found consistently by three independent methods, and based on data either from the whole Antarctic continent, or from six individual colonies. The more accurate EBSP and Stairway Plot approaches (see §44 p. 112 and §46 p. 121) also suggest a population growth before 100,000 years, and back to 1,000,000 years according to the EBSP - however, these are deep times by the yard of penguin evolution, and at these scales demographic and speciation processes may start to interfere. This confidently stable population during the later Pleistocene contrasts with the findings of Younger *et al.* 2015: however, as we suggested in §137 p. 253, the steep population growth inferred from mtDNA data by Younger and colleagues rather seems to stem from inadequate analysis. Indeed, the use of unexplained diploid sites in haploid mitochondrial DNA (likely sequencing errors caused by secondary structures in the Emperor penguin control region) is expected to artificially increase recent population size by overestimating the importance of low frequency variants. Other improper choices, such as treating mitochondrial cytochrome-b gene and mitochondrial HVR as unlinked, independent markers in an EBSP framework (instead of concatenating them and remaining within the frame of a simple sin-

Discussion - §170

1 gle-locus Bayesian Skyline Plot), are likely to have biased this reconstruction: thus, they do not
2 contradict our findings convincingly. This stable effective population size also implies that
3 metapopulation structure was never prevalent in the species (see §169 p. 305, and let aside a very
4 unlikely hypothesis of balanced compensation between census size and gene flow) - not even dur-
5 ing the Last Glacial Maximum, when extensive winter sea ice cover and Antarctic ice sheet ex-
6 pansion could have been thought to have disrupted the Emperor's breeding cycle. Interestingly,
7 the persistence of the Emperor penguin population throughout the period seems to imply that
8 the conditions on the major part of the Antarctic shoreline were not significantly out of the
9 species' optimum - *i.e.* somewhere between the harsh conditions now found at the high latitudes
10 of the Ross Sea, and the milder ones that prevail at the lower latitudes of Wilkes Land in Eastern
11 Antarctica. Importantly, the stability of the Emperor penguin population means that coastal
12 polynya activity was comparable to what we observe nowadays (see §18 p. 65) despite very differ-
13 ent global sea ice conditions (see §9 p. 42) - a hypothesis confirmed by our preliminary analysis
14 of ice conditions in the coastal Antarctic area (see §116 p. 223 to §122 p. 235). It also confirms
15 the hypothesis of Wolff *et al.* 2006 who, based on iron flux reconstructed from the EPICA ice
16 core (Augustin *et al.* 2004), suggested that marine productivity did not change significantly dur-
17 ing the last Pleistocene glacial cycles: indeed, the robustness of the Emperor penguin population
18 implies that food resource was never a limiting factor for the species, even during the coldest
19 times of the last era.

20 The King penguin offers a very different picture (detailed in *The King penguin*, p. 163). Al-
21 lowance made for the possibly complex interplay of population decline and range fragmentation
22 (see above, §169 p. 305), the King penguin differs from the Emperor by its drastic demographic
23 response to the Last Glacial Maximum, and likely also to the previous Llanquihue glacial maxi-
24 mum - beyond which coalescence signal is lost in our sample, and possibly in the whole extant
25 gene pool of the species. A previous study by Trucchi *et al.* 2014 had proposed a strong post-
26 glacial expansion of the species based on RAD data for 8 individuals, with strong support based

Discussion - §170

1 both on EBSP, and a diffusion-approximation based spectrum approach (see §45 p. 117): but
2 that first study could not retrieve any signal past the LGM low-population-size period, an unsur-
3 prising consequence of the low sample size. The novel and more accurate method established by
4 Liu & Fu 2015, that we deploy in *The King synnome*, p. 163, fully confirms that finding - but sig-
5 nificantly c hanges i ts i nterpretation b y m aking t he p ost-glacial p opulation e xpansion e vent a
6 recovery rather than a *de novo* colonisation. The oscillating population size that we reconstruct in
7 the King penguin offers a striking contrast with the Emperor penguin: as we demonstrate, the pe-
8 riods of strong population reduction are caused by a drastic reduction in available habitat under
9 LGM conditions. The difference between the essentially fluid habitat of the Emperor penguin,
10 that may follow the sea ice, and the semi-rigid habitat of the King penguin, that can only breeds
11 on the available islands, but, just as the Emperor, is also sensitive to changes in the oceanic and
12 atmospheric spheres, appears as an important evolutionary divergence - which gave a clear advan-
13 tage to the most extreme adaptation, against the more generalist one, during the whole
14 Pleistocene.

15 Interestingly, all our demographic reconstructions suggest that the Emperor penguin was the
16 most abundant *Aptenodytes* penguin species during the late Pleistocene, but was replaced by the
17 King penguin during the early Holocene. The early Holocene climatic optimum is one of the pe-
18 riodic interglacials that punctuate the Pleistocene (Wolff *et al.* 2006 - or at least it resembles them
19 in every respect, but anthropogenic climate change might decide otherwise), and our demograph-
20 ic reconstructions suggest that the King penguin's breeding system and physiology evolved to
21 provide highest fitness under these conditions, at the cost of much diminished populations dur-
22 ing the regular glacial maximums. The specialisation in a relatively unstable niche (at the milenial
23 scale) thus forces the King penguin into cycles of growth and decline, while the paradoxically
24 more stable (because less firmly anchored in the midst of its unstable surroundings) niche of the
25 Emperor allowed it to go through the Pleistocene unhindered. However, current projections seem
26 to indicate that the Holocene-optimum conditions indeed provide the fitness optimum for the

1 King penguin, and that further increase in temperature beyond these conditions is equally likely
2 to bring about population decline. Indeed, the replacement of the Emperor penguin by the King
3 penguin as the dominant *Aptenodytes* biomass during the early Holocene implies that the latter is
4 already the «warm adapted» species of the clade - and its glory days may well be behind it.

5 *Seabird response to Anthropocene environmental change.*

6 §-171 *Seabird response to climate change.* Evidence is accumulating for a negative response of pen-
7 guins in general, and *Aptenodytes* penguins in particular, to ongoing climate change (Ainley *et al.*
8 2010a). Recent studies on the Emperor penguin (Barbraud & Weimerskirch 2001a; Jenouvrier *et*
9 *al.* 2009; Jenouvrier *et al.* 2012; Jenouvrier *et al.* 2014), the King penguin (Le Bohec *et al.* 2008;
10 Péron *et al.* 2012), or the Adélie penguin (Jenouvrier *et al.* 2006) have emphasised the vulnerabil-
11 ity of these species to climate warming. Our results on the King penguin (see *The King synnome*,
12 p. 163) clearly support this view. However, the two penguin species examined in this work are
13 predicted to respond to climate change for very different reasons, and based on very different
14 mechanisms - and both are subject to considerable uncertainties. The gaps in our knowledge of
15 the Emperor penguin are maybe the most clearly visible. All of the data used in recent projections
16 come from the single Pointe Géologie colony (see *The Emperor synnome*, p. 211), that has under-
17 gone tremendous disturbance in the recent past (see §24 p. 79 for an overview), so that the part
18 of climatic, and direct influences in its demography are very blurred to say the least. Adding to
19 these uncertainties are the basic assumptions of climate envelope modelling approach (Pearson &
20 Dawson 2003), which are correlative, and may or may not reflect true causal processes (see §49
21 p. 127) - and do not integrate the possibility of (lagged) plastic adaptation. Finally, in the study
22 of Jenouvrier *et al.* 2014, sea ice emerged as the main forcing on the Emperor's demography:
23 however, the representation of sea ice in the CMIP3 and CMIP5 ensemble models is still inaccu-
24 rate, and show trends that are contrary to the observations over the historical periods (Turner *et*
25 *al.* 2013; Shu *et al.* 2015). Combined, these three layers of uncertainty (in the reliability of flip-

1 per-banded, disturbed birds, in the robustness of the correlative envelope approach, and in the
2 accuracy of the CMIP sea ice projections) amount to a considerable lack of knowledge, that
3 leaves room for about any hypothesis.

4 The King penguin's response seems, at first glance, to be more robust. The negative correlation
5 between sea-surface temperature and adult survival and breeding success is well established (Le
6 Bohec *et al.* 2008), and its causal mechanism is equally well understood: the warmer the sea sur-
7 face temperature, the further South the Antarctic Polar Front and associated foraging areas, and
8 the longer the foraging trips for the Northern King penguin colonies (see Péron *et al.* 2012, Bost
9 *et al.* 2015, and *The King synnome*, p. 163) - a mechanism observed directly using GPS loggers
10 placed on penguins. However, a possibly important flaw in this model is the fact that until now,
11 the main driver of sea surface temperature variation is not directly anthropogenic climate
12 change¹, but the natural oscillations of the Southern Hemisphere Ocean-Atmosphere circulation,
13 such as the Southern Annular Mode, the El Niño Southern Oscillation or the Southern Indian
14 Ocean Dipole (see §8 p. 40 for a summary of these phenomena). The changes in temperature as-
15 sociated with these oscillations can reasonably be taken as proxies for global change (although
16 this is, in itself, subject to some uncertainties, see §11 p. 50). But approximating the response of
17 seabirds to global long-term change by their response short-term oscillations is another matter.
18 Indeed, oscillating phenomena such as the ENSO force the alternance of « warm » and « cold » pe-
19 riods over 3 to 6 years cycles: while breeding success is very low during the warm years, it may be
20 partly compensated during the cold years. Considering that the most exposed populations (on
21 Crozet archipelago, see *The King synnome*, p. 163) are still stable demographically (Delord *et al.*
22 2004; Saraux *et al.* 2011a), it is possible that the cyclic character of the changes makes even plas-
23 tic adaptation un-necessary for the time being: put simply, « waiting for better times » can still be

1. While anthropogenic climate change clearly has an influence on Southern Ocean sea surface temperature, as we showed for example in **The King synnome**, p. 163, it mostly affect the *mean* temperature, and its underlying trend. But interannual variability is still central in local processes, and is of the same, or of a larger order than man-made trends.

Discussion - §172

1 a beneficial strategy for a bird like the King penguin. Whether this strategy will still hold under
2 long-term irreversible changes, is another question. Several alternative strategies will still be possi-
3 ble: altogether not trying to breed during the warmer phases of ENSO (as has been reported in
4 the Galapagos islands, see Duffy & Siegfried 1987) and building up reserves for increased breed-
5 ing success during the more favorable years, developing nearshore foraging strategies, etc.
6 Whether King penguins will react differently to linear changes than to cyclic changes is of course
7 speculative, but it does add to the uncertainty of predicted demographic trajectories. If the actual
8 timing and extent of climate change effects is far from certain, the reality of these effects is, at any
9 rate, beyond doubt. Indeed, Pleistocene changes in King penguin population are qualitatively
10 consistent with the expectations set by oscillative climate change *The King synnome*, p. 163: in our
11 case, long term data validates hypotheses set by very short-term data, both converging to empha-
12 sise the importance of climate change in shaping the future of penguin.

13 *§-172 Seabird response to environmental change.* Despite the overwhelming evidence for a deep
14 impact of climate change on biological communities (and that our work clearly supports), reduc-
15 ing the current threats to the sole effects of climate change is misleading. In a recent study, Trivel-
16 piece *et al.* 2011 demonstrated how direct anthropogenic effects, such as the hunt of cetaceans to
17 near-extinction, or the present-day overexploitation of krill stocks (*e.g.* for farmed-fish food), had
18 a considerable effect on Chinstrap penguin *Pygoscelis antarctica* breeding success and recruitment,
19 that outweighed the impacts of climate change. Generally, overfishing has been identified as the
20 major cause of decline in seabird populations, far before the change in climate (Cury *et al.* 2011).
21 This observation may, in a sense, be an optimistic one, as recovery from overfishing may happen
22 faster and more fully than recovery from greenhouse-gas induced global warming (which may not
23 happen at all on a human time scale).

24 In that respect too, the Peruvian diving-petrel is an extreme case. For that species, several well
25 identified direct factors have contributed to reducing large populations to vestigial colonies in a

Discussion - §173

1 matter of decades: in the current state of knowledge, direct habitat destruction, predator and
2 competitor introduction, overfishing, and hunting (see §149 p. 279) are responsible for the dras-
3 tic population decline more than any effect of climate change, *e.g.* the intensification of the El
4 Niño dipole oscillation. This again is a rather hopeful situation: indeed, since the threats are
5 mostly local, local conservation measures such as land and marine habitat sanctuarisation and
6 reintroduction may still save the species and help it recover. Comparing the King penguin and
7 the Peruvian diving petrel gives a clear picture of the spectrum of threats weighing on seabirds in
8 the anthropocene: one century ago, the main threat on both species was direct destruction by hu-
9 mans (see §22 p. 75 and §149 p. 279). Full protection granted to the King penguin allowed it to
10 recover almost entirely, while insufficient or failing protection maintains the Peruvian diving-pe-
11 trel populations at a critical level. Yet currently, because of the species' deep connexion to global
12 circulation mechanisms (§18 p. 65), climate change has become the main threat to the King pen-
13 guin - while climate effects are dwarfed, in the diving-petrel, by the comparatively overwhelming
14 direct human impact.

15 *Further work on the penguin synnomes.*

16 *§-173 A final examination of penguin population structures.* Our studies explored most of the geo-
17 graphical range of both the Emperor and the King penguin, and concluded in the remarkable
18 level of mixing in these two *synnomes* (§167 p. 299). However, the findings of Younger *et al.*
19 2015, if they lack robustness for methodological reasons (see §137 p. 253), invite us to explore
20 further possible (yet unlikely) reduced-gene-flow areas in the Emperor penguin, *e.g.* in the Ross
21 Sea, Marie Byrd and Ellsworth lands, or the Antarctic peninsula. Similarly, two areas were not
22 fully covered by our study of the King penguin (Macquarie Island only through mtDNA, and
23 Tierra del Fuego not at all). Interestingly, Tierra del Fuego is a newly founded colony (Kusch &
24 Marín 2012), and could give us fascinating insights into the genetic aspects of colony foundation

Discussion - §174

1 (in particular the respective roles of intrinsic growth and immigration)¹. Finally, the precise ques-
2 tion of the subspecific taxonomy in the King penguin (see §25 p. 80), although we consider it as
3 good as settled here, may require explicit testing, including morphometric measurements to
4 match the observations of Barrat 1976.

5 In the Peruvian diving-petrel, on the other hand, additional sampling² will likely improve greatly
6 our understanding of ongoing population processes. Our Chilean sampling shows that isolation
7 is a major population process in this species, but an additional «outgroup» sample, distant from
8 out sampling by ~1,500 km, should allow us to set bounds on the migratory flow, and to under-
9 stand wether it is mainly driven by geographical distance (in which case we should observe a clear
10 phylogeographic structure, driven by isolation-by-distance gradients), or by more complex
11 mechanisms.

12 *§-174 Extending palaeodemographic reconstructions.* Our final reconstructions for population size
13 changes in the King and the Emperor penguin use a currently state-of-the-art method (Liu & Fu
14 2015), and are a large improvement over previous studies (Trucchi *et al.* 2014). However, coales-
15 cent-based reconstructions are inherently limited by the very structure of genetic diversity: once
16 all examined sequences have found a single common ancestor, no further information can be re-
17 trieved. This is expected to happen, in a stable population, around $4.N_e$ generations back in the
18 past (which, in *Aptenodytes* penguins, would be as considerable as 1 million years) - but much
19 sooner if demographic events, such as bottlenecks, temporarily increase the probability of coales-
20 cence at given points of time. This phenomenon is easily visible when contrasting the King and
21 the Emperor penguins (Fig. 19): although population sizes and generation times are comparable,

1. Sampling has been done in Tierra del Fuego, in Chile, with the collaboration of Dr Juliana Vianna, but will not be processed and integrated in the present work.

2. Sampling is currently underway in Isla San Gallán, in Peru, with the collaboration of Dr Carlos Zavalaga, but will not be finished in time to be processed and integrated in the present work.

Discussion - §174

1 the Emperor penguin population size can be reconstructed further back in time due to the ab-
2 sence of intermediate bottlenecks.

3 In order to retrieve evidence about more ancient times, two lines of action are possible. First, we
4 may try to enlarge the sample size, in the hope that the increased diversity may retain informa-
5 tion as to more ancient ancestors. However, past a modest sample size (as low as ~16 individuals),
6 the full coalescent should already be reasonably well described, and increasing sample size is not
7 expected to help much (Felsenstein 2006; Hoban *et al.* 2013a; Hoban *et al.* 2013b). Second, we
8 may directly try to shift the time limit by integrating ancient samples in our reconstructions. In a
9 «pure» coalescent framework, like the EBSP (see §44 p. 112), this may take the form of serial
10 sampling (*i.e.* individuals from different points of time can be integrated in a single time-tree) in
11 order to use the full evidence jointly. In a derived framework like the Stairway plot approach (see
12 §46 p. 121), however, serial sampling is not directly possible, since the model is fitted to a time-
13 flat object, the allele frequency spectrum (see §45 p. 117). Extending this method to serial sam-
14 pling would require (*i*) that several ancient, but mutually contemporaneous samples are available,
15 in order to establish a «palæospectrum». This is a rather drastic restriction imposed on the exper-
16 imental design, and would only be possible in rather specific cases (*e.g.* the ice-core sampling of
17 penguin moult feathers and remains, as performed by Emslie *et al.* 2007 for isotopic niche stud-
18 ies), and (*ii*) to have some knowledge about the relationship between the sampled palæo-popula-
19 tion, and the modern population (see *e.g.* Heupink *et al.* 2012). If these two conditions can be
20 met, further implementation could involve (*i*) a simple model in which independent reconstruc-
21 tions would be performed for modern and ancient samples, using the current algorithms, (*ii*)
22 preferably, an integrated reconstruction, in which population size at each point of time would be
23 chosen in order to maximise the likelihood of SNP classes in both spectra simultaneously (thus
24 assuming that the modern population is the direct descendant of the ancestral one), or (*iii*) more

Discussion - §175

1 accurately still, by estimating the joint allele-frequency spectrum of the two *chronodemes*¹, and
2 modelling allele inheritance as well as other demographic processes as unidirectional diffusion pa-
3 rameters (from the more ancient chronodeme to most recent one). In a Stairway plot framework,
4 this would considerably widen the parameter space (since the number of SNP classes become the
5 product, and not the sum, of the dimensions of the two spectra), but there is no obvious impossi-
6 bility there.

7 Technically, this would also probably require shifting to a whole-genome resequencing, rather
8 than RAD-sequencing, approach, as enzymatic digestion is notoriously unpredictable in ancient
9 DNA. But King penguin DNA, for example, has been successfully extracted for samples as an-
10 cient as 8,000 years old conserved in peat (Heupink *et al.* 2012), so that there is hope to retrieve
11 information prior to the most ancient of our reconstructed bottlenecks, and most importantly to
12 verify whether the oscillating population size pattern may be extended to the whole Pleistocene
13 period.

14 *§-175 From neutral to adaptive genomics.* RAD sequencing, since its original formulation in 2007
15 (Miller *et al.* 2007; Baird *et al.* 2008; Davey *et al.* 2011), has been successfully used in a wide
16 range of genomic applications. These have focused, as we did in this work, on the informative
17 content of neutrally evolving regions, in particular in the fields of phylogenetics (Cariou *et al.*
18 2013; Cruaud *et al.* 2014), phylogeography (Emerson *et al.* 2010), palæodemography (Trucchi *et*
19 *al.* 2014), and generally population genetics (Davey & Blaxter 2010) - but also on speciation
20 analysis (Gagnaire *et al.* 2013b), and on more resolutely functional aspects of genome evolution,
21 such as quantitative trait locus analysis and association mapping (Hohenlohe *et al.* 2010; Gag-
22 naire *et al.* 2013a), genetic maps (Amores *et al.* 2011; Amores *et al.* 2014), quantitative genomics
23 (McGuigan *et al.* 2010). In a preliminary study, Trucchi *et al.* 2014 proposed an extension of the

1. We propose this neologism to describe the extension in time of the *deme* idea, on the model of the *chronospecies* concept.

Discussion - §175

1 RADseq analysis paradigm to take advantage of its short-haplotype structure in a coalescent
2 framework (§29 p. 85 and §31 p. 87), using a subset of highly variable loci for EBSP reconstruc-
3 tion of past population size changes. In this work, we developed this idea further, and applied it
4 to population size and migration reconstruction, concurrently with SNP-based analyses. Based
5 on our results (see *e.g.* the demographic reconstructions in *The King synnome*, p. 163), RAD-se-
6 quencing offers clear advantages over whole-genome resequencing: data can be gathered for a
7 much larger number of individuals for a similar sequencing effort, and the accuracy of demo-
8 graphic reconstructions is increased using this wide dataset.

9 However, we did encounter a number of methodological limitations. Besides the random biases
10 that do occasionally appear during library preparation (see §35 p. 93), the main *caveat* of RAD-
11 seq is its discontinuous character (§31 p. 87): the unsequenced gaps between markers make it
12 difficult to identify continuous structural processes in the genome. Thus, while focusing on neu-
13 tral processes, we left aside functional phenomena - in particular the signature of local selection
14 (that may be detected through clues like selective sweeps - see Nielsen *et al.* 2005, or long runs of
15 homozygosity - see Kirin *et al.* 2010), or properly genomic processes such as the evolution of re-
16 combination hotspots (Smagulova *et al.* 2011) or replication origins (Lucas *et al.* 2007). Howev-
17 er, now that a strong neutral framework has been established for both *Aptenodytes* species, further
18 understanding may be gained from the functional and adaptive aspects of their genome. In par-
19 ticular, understanding whether speciation was the result of allopatric isolation and neutral drift
20 (see §168 p. 304), or rather of a strong and sudden local selection pressure, may give us insights
21 into the forces that shaped the Southern Ocean biodiversity during the Pliocene and the Pleis-
22 tocene. For these processes, whole-genome resequencing is likely to offer a framework for the
23 more precise interpretation of RAD data, *e.g.* by allowing us to locate RAD markers in relation to
24 genome landmarks. Further investigation methods, such as bisulfite-conversion (Deng *e t al.*
25 2009), possibly in a RADseq framework (Trucchi *et al. in prep.*) can offer insights into epigenetic
26 phenomena (*e.g.* methylated cytosine residues). RNAseq (*e.g.* Mortazavi *et al.* 2008) may allow us

Discussion - §176

1 to go further, by identifying very short-term local plastic responses through modulation of gene
2 expression levels and patterns, for example during development.

3 Yet the plethora of new techniques in the field of genome analysis should not divert us from the
4 real originality of the *Aptenodytes* model. The wealth of life-history data made accessible by the
5 advanced monitoring program on Possession Island (see §26 p. 81 and Gendner *et al.* 2005; Le
6 Bohec 2008) offers a unique framework for studying the interconnection of ecological and ge-
7 nomic processes in the wild, and further research will highly benefit by exploring this interface on
8 a large number of individuals. In particular, the study of QTL associations between behavioural
9 traits (such as breeding output, foraging efficiency, or partner and site fidelity) and genomic or
10 epigenomic features may shed new lights on the evolution of life histories. In this respect too, our
11 work lays a strong foundation for further investigation, since the identification of neutral popula-
12 tion structure is a pre-requisite to the study of adaptation (Price *et al.* 2006) - in our particular
13 case, the synneme structure also has implications on the possible realisation of processes such as
14 positive selection (see §176 p. 320).

15 *§-176 Synnemes and gene swamping.* Delving into functional and adaptive processes may also al-
16 low us to shed light on the relationship that short- and long-term processes entertain in the
17 species' dynamics and evolution. In §95 p. 186, we reconstructed the King penguin synneme
18 growth rate along its Pleistocene and Holocene history: despite the very large amplitude of the
19 fluctuations in population size, the growth parameter λ^1 remained very stable, ranging from a
20 minimum $\lambda_{\min} = 0.999$ in the decline periods to a maximum of $\lambda_{\max} = 1.002$ in the growth peri-
21 ods. In comparison, when growth rate is extracted over a short period on a yearly basis
22 (1999-2010, Le Bohec *et al in prep.*), variation is larger by one order of magnitude ($\lambda_{\min} = 0.99$
23 and $\lambda_{\max} = 1.08$). This is an expected observation, since growth rate is a highly stochastic parame-

1. The growth rate λ is the ratio of the population size between two successive years (in our case - otherwise any other time interval may be used).

1 ter, and decade-scaled variability is dominated by yearly environmental effects, that are evened
 2 out at longer scales. Yet, this raises the question of the methodological link that can be estab-
 3 lished between local demographic monitoring surveys and evolutionary processes. As we men-
 4 tioned in our introduction (see in particular §16 p. 60), most studies focusing on response to cli-
 5 mate change, in particular in vertebrates, are necessarily restricted to the observation of a limited
 6 number of individuals and generations: this individual-based monitoring is by nature focused on
 7 individual responses to changes. But as we see, making a transition from individual to collective
 8 responses, even in such a basic parameter as growth rate, is not trivial. Or, in Tolstoy's words: «*a*
 9 *human lives consciously for himself, but is an unconscious instrument in the accomplishment of the his-*
 10 *toric, universal, aims of humanity. A deed done is irrevocable: and its result, coinciding in time with*
 11 *the actions of millions of other men, assumes a historical significance [...]. As long as we will only write*
 12 *the history of singled-out characters - be they Cesar, Alexander, Luther or Voltaire - and not the history*
 13 *of every man without exception who took part in the event, we will not be able to explain the move-*
 14 *ments of mankind without conceiving a force constraining men to strive towards a common end» (Tol-*
 15 *stoy 1869). Put simply, individual responses matter less than the synchronised, convergent re-*
 16 *sponse of great numbers of individuals: population phenomena can be represented as global,*
 17 *species-scaled trends, yet the essence of these phenomena lies in the joint individual trajectories*
 18 *that underly the common response.*

19 Thus, for penguins, which are not organised in «nations» or populations, only the consensus re-
 20 action of the whole species should really be considered meaningful: local trajectories may lead
 21 (from a local perspective) to «local extinctions», or (from a global perspective) to «range shifts»,
 22 but the species' eventual evolutionary (and adaptive) response can only be determined globally.
 23 Indeed, free movement of alleles between colonies implies that *migration load*¹ certainly reaches

1. *Migration load* can be defined as the «decrease in mean fitness of a population because of immigration. This occurs because the phenotypic mean of the population is different from the local optimum value» (Lenormand 2002). Thus, in the case of positive selection leading to local adaptation, migration load is the counteracting force bringing random alleles into the local gene pool through migration events, and making local adaptation less efficient.

1 the point of *gene swamping*¹ (see a complete review of these concepts in Lenormand 2002): for a
 2 gene to be positively selected, it needs to offer consistent selective advantage throughout the
 3 whole species' range - and, to undergo purifying selection², it needs to be consistently detrimental
 4 everywhere. In short, the *migration-selection equilibrium* is brought to an extreme of *averaged se-*
 5 *lection*, where local habitat conditions are integrated over the whole distribution of the species.
 6 Whereas individual fitness selection occurs relative to the match of individual traits and local
 7 habitat characteristics, the final selective pressure on the allele will, on the contrary, be the result
 8 of its average selective advantage under the species' total habitat. This has two important conse-
 9 quences: (i) even though there may be drastic local selection for certain alleles, local adaptation
 10 will be prevented by the migration load, and (ii) as long as «*gene refugia*» (i.e. local conditions
 11 where a gene retains a positive or neutral selective coefficient) are maintained in the species'
 12 range, fixed adaptation is unlikely.

13 Shifting from a neutral evolution paradigm, such as the one we explored in the present work, to a
 14 functional and adaptive one, should enable us to identify such processes. Rapid local micro-adap-
 15 tation, for example, should be detectable through particularly shallow coalescence times at locally
 16 selected loci, and spatial homogeneity in the patterns of selection may be an important clue on
 17 ongoing gene swamping. Comparison with highly philopatric species such as the Peruvian
 18 diving-petrel might be especially informative: in that case, the full genetic sorting between even
 19 closely neighbouring colonies makes local adaptation a likely possibility. And the fitness benefit
 20 can be easily understood in the stable Atacama coast environment, where both the insular breed-
 21 ing grounds and the climate exhibits relatively high stability (e.g. in the ENSO cycles, see Moy *et*
 22 *al.* 2002): investing on a particular local evolutionary strategy can pay off, since the is good
 23 chance that the match between adaptation and environment will be maintained for some time.

1. *Gene swamping* can be used to describe « the situations where there is no significant response to selection because gene flow is too high » (Lenormand 2002) - in other words, when the migration load fully counteracts the effects of local selection.
 2. Or «negative selection» - i.e. active removal of detrimental alleles from the local gene pool.

1 Thus, the contrast between high philopatry and reduced gene flow in the Diving petrel's stable
2 environment, and the large penguins' panmictic system amidst unstable oceanic structures is
3 double-sided: from a «*neutral*» point of view, unstable environments are likely to promote higher
4 dispersal, in particular in response to local habitat destruction, as we showed in *Empirical evi-*
5 *dence of heterogeneous dispersal*, p. 259, and therefore to bias population systems towards pan-
6 mixia, whereas stable environments allow for perpetuation of colonial units on the long term,
7 and consequently allow for drift and genetic isolation. But from an «*adaptive*» standpoint, stable
8 environments allow species to benefit from local adaptation, and are expected to select for lower
9 dispersal in order to reduce migration load on the efficiency of selection (Billiard & Lenormand
10 2005), while unstable environments will rather enhance the benefit of conserving a more general-
11 ist gene pool, that is globally suited to the range of possible conditions, although it may not be
12 optimal in the current state of affairs - that is, ultimately, full *gene swamping*. Of course, distin-
13 guishing one line of explanation from the other is a lost cause, since these are probably but two
14 ways of describing the same, unified phenomenon¹.

15 *Conclusion*

16 Bringing together the tools of population genomics, numerical climate models and field ecology
17 allowed us to sketch a general picture of the effects of climate change on three remote seabird
18 species. And whereas the general perception of the Southern Ocean is often that of a pristine and
19 extreme environment, one of our most remarkable conclusions may be that it is, in fact, neither.
20 As we exposed throughout this work, even species as rarely met by humans as the King and the

1. Or, to pursue on Tolstoy's thought, «*when an apple has ripened and falls, why does it fall? Because of its attraction to the earth, because its stalk withers, because it is dried by the sun, because it grows heavier, because the wind shakes it, or because the boy standing below wants to eat it? Nothing is the cause. All this is only the coincidence of conditions in which all vital organic and elemental events occur. And the botanist who finds that the apple falls because the cellular tissue decays and so forth is equally right with the child who stands under the tree and says the apple fell because he wanted to eat it and prayed for it. Equally right or wrong is he who says that Napoleon went to Moscow because he wanted to, and perished because Alexander desired his destruction*» (Tolstoy 1869).

Discussion - §176

1 Emperor penguin are nonetheless at the forefront of anthropogenic change. The first and most
2 direct effect of these changes is, of course, the increase in temperature, both at the sea surface
3 (which results in polar front migration and loss of foraging grounds for the King penguin) and in
4 the atmosphere (which results in modified sea ice formation patterns, and potentially loss of for-
5 aging grounds too for the Emperor penguin). But, as we presented earlier (see §84 p. 174), cli-
6 mate change itself has other, more indirect effects: by affecting the physiology of plankton, it
7 changes the global primary productivity of the Southern Ocean; by modifying sea ice patterns, it
8 affects the blooms of Antarctic Krill (Brierley *et al.* 2002 - with cascading effects on seabirds, see
9 Nicol *et al.* 2008; Chapman *et al.* 2010; Trivelpiece *et al.* 2011); by changing the heat exchange
10 between sea surface and atmosphere, it is likely to change the circulation patterns of water mass-
11 es, with immediate effects on ecosystem structure (Nicol *et al.* 2000).

12 Nor does the remoteness of the Southern Ocean mean that it is isolated: directly connected to
13 the three major ocean basins, it is at the heart of the global thermohaline circulation (see §7 p.
14 35), and a major input into the lower latitudes (in particular through the eastern boundary cur-
15 rents such as the Humboldt Current, see §147 p. 276) - and changes in its circulation patterns
16 have strong consequences on the world's ocean circulation and productivity patterns. Thus, the
17 effects of climate change that we document here in three seabird species are by no means limited
18 to their particular ecosystems, but must rather be interpreted as a marker of large-scale changes in
19 the global state of the Oceans.

20 Indeed, the extreme adaptations of polar species often motivate the misleading impression that
21 they reveal equally extreme, and perhaps *marginal* effect of climate change. Yet although we can,
22 from a phylogenetic standpoint, infer that *Aptenodytes* penguins' adaptations are extreme com-
23 pared to their avian ancestors, there is nothing extreme in the species themselves, as seen from
24 within their own habitats: they are, in that sense, strongly representative of the world's ecosystems
25 at large. If the effects of climate change are more visible in polar regions than anywhere else, it is

Discussion - §176

1 also, as noted by Smetacek & Nicol 2005, because direct human impacts occurred later in the
2 high latitudes than in the tropical and temperate belts, and that less climate change is to some ex-
3 tent visible there in a more «raw» form. Yet it seems that we are now rapidly catching up on
4 these delayed impacts - to mention only the steep rise in marine overexploitation in the Southern
5 Ocean (Myers & Worm 2003; Ainley & Blight 2009), and the increasing concentration of conta-
6 minants at high latitudes (Nash 2011).

7 Thus, regarding the fate of penguins and other seabirds in the face of man-made climate change,
8 our study may have brought us as many questions as answers. Yet there may be some optimism in
9 cases such as the one of the Peruvian diving petrel, in which resolute conservation measures ap-
10 pear to be having actual effect. Yet in all cases, the disproportionate effect of direct human impact
11 (through population decimation in the King penguin, habitat destruction in the Diving-petrel,
12 and human disturbance on the Emperor penguin) competes with indirect impact through an-
13 thropogenic climate change, so that it may be vain to separate both questions. Indeed, in the
14 present state of things, climate change, ecosystem overexploitation and habitat destruction are
15 but different faces of one general stance, that doubtlessly needs to be renewed.

16 This is however a fully different topic, that we will hopefully have the opportunity to explore fur-
17 ther on. In the meantime, I wish to finish this work by thanking the reader for his patience and
18 perseverance, and by expressing once again all my gratitude to Dr Céline Le Bohec and Dr Emil-
19 iano Trucchi for their guidance, and to Dr Beaugrand and Dr Davey for accepting the heavy task
20 of reviewing it.

Discussion - §176

1 Annexes

2 King penguin demography since the last glaciation inferred from 3 genome-wide data

4 The following work was performed during the earliest stages on this PhD, with a limited partici-
5 pation on my side - but it served as an important methodological basis throughout this thesis.
6 We therefore include it here, mostly for reference:

7 Trucchi E, Gratton P, Whittington JD, Cristofari R, Le Maho Y, Stenseth NC, Le Bohec, C
8 (2014) **King penguin demography since the last glaciation inferred from genome-wide data.**
9 *Proceedings of the Royal Society B: Biological Sciences*, 281, 20140528.

10 *Abstract*

11 How natural climate cycles, such as past glacial/interglacial patterns, have shaped species distribu-
12 tions at the high-latitude regions of the Southern Hemisphere is still largely unclear. Here, we
13 show how the post-glacial warming following the Last Glacial Maximum (ca 18,000 years ago),
14 allowed the (re)colonisation of the fragmented sub-Antarctic habitat by an upper- level marine
15 predator, the king penguin *Aptenodytes patagonicus*. Using restriction site-associated DNA se-

1 quencing and standard mitochondrial data, we tested the behaviour of subsets of anonymous
2 nuclear loci in infer- ring past demography through coalescent-based and allele frequency spec-
3 trum analyses. Our results show that the king penguin population breeding on Crozet archipel-
4 ago steeply increased in size, closely following the Holocene warming recorded in the Epica
5 Dome C ice core. The following population growth can be explained by a threshold model in
6 which the eco- logical requirements of this species (year-round ice-free habitat for breeding and
7 access to a major source of food such as the Antarctic Polar Front) were met on Crozet soon after
8 the Pleistocene/Holocene climatic transition.

9 *Introduction*

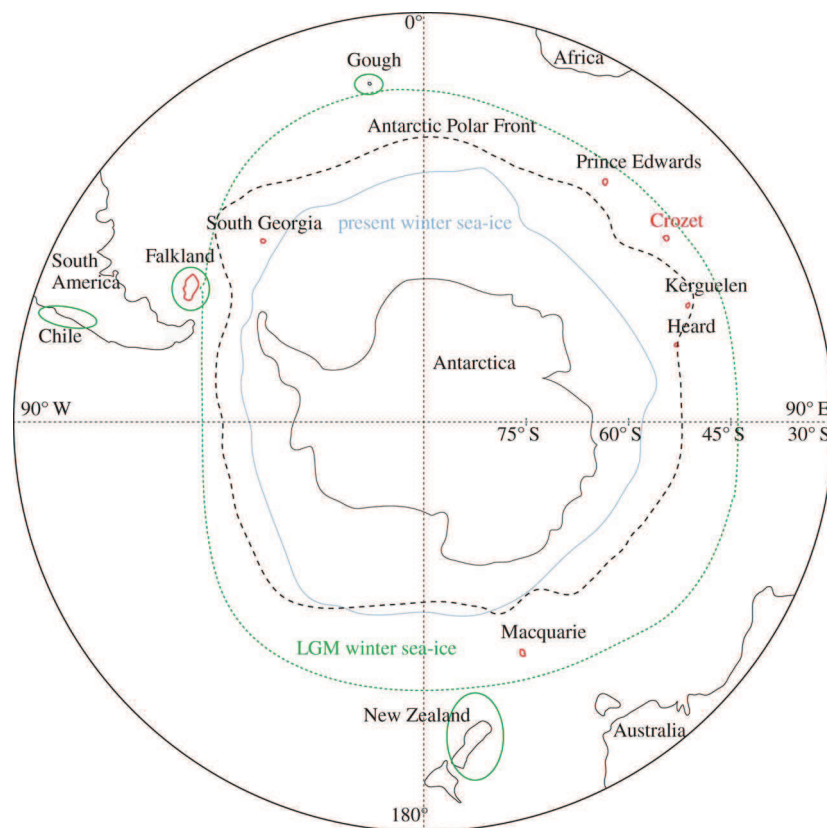
10 Environmental variation through time is one of the drivers of the evolutionary process and a key
11 mechanism in shaping biodiversity. Both short- and long- term climate shifts dramatically affect
12 the geographical distribution of species/populations according to their own dispersal abilities and
13 ecological requirements (Hewitt 2000; Parmesan & Yohe 2003). Understanding how past fluctu-
14 ations impacted the demography of key species in different ecosystems is essential for predicting
15 the response of communities to ongoing and future change, including anthropogenic-induced
16 climate forcing (Hoffmann & Sgrò 2011; Bellard *et al.* 2012). Using genomic data, we investi-
17 gate the impact of the last glaciation on the past demography of a king penguin (*Aptenodytes*
18 *patagonicus*) colony from the most important breeding areas for this species (the Crozet
19 archipelago).

20 While the biological impacts of recent Quaternary glaciation events on the largely terrestrial
21 Northern Hemisphere have been relatively well documented (Bellard *et al.* 2012), the oceanic
22 Southern Hemisphere, particularly at higher latitudes, remains poorly described (Fraser *et al.*
23 2012) owing to its relative inaccessibility. Recent hypotheses suggested that winter sea ice during
24 the Last Glacial Maximum (LGM) could have extended as far north as South Georgia, Crozet,

1 Kerguelen and Macquarie Islands, while Falkland, Gough and New Zealand Islands may have
2 acted as ice-free refugia (Fig. 45 p. 330, and Fraser *et al.* 2009). As southern ocean conditions
3 fluctuated between glacial and interglacial periods, local species had to contend with changes in
4 both habitat and resource availability. In particular, the distribution of most seabirds is primarily
5 constrained by the spatial location of suitable breeding sites on land and foraging areas in the
6 ocean (Hunt 1991). For a population to thrive, productive foraging areas must exist within reach
7 of a suitable breeding site. Climate oscillations affect both these key ecological constraints. On
8 one side, these fluctuations cause glaciers to expand and retreat, thus covering or exposing potential
9 breeding sites, on the other side, they affect the location and intensity of oceanic currents and
10 fronts that in turn determine marine primary production. According to (Thatje *et al.* 2008) the
11 large-scale Southern Ocean productivity may have been nearly shut down during the LGM, at
12 least at the latitude where it is currently present. Highly dispersive species negatively affected by
13 glacial conditions could have moved northwards to more suitable ice-free breeding habitats during
14 the LGM, while low-dispersive ones may have been strongly reduced in population size or
15 may even have gone extinct (Baroni & Orombelli 1994). Therefore, post-glacial warming and the
16 subsequent retreat of glaciers and sea ice, probably favoured the re-colonization of Antarctic and
17 sub-Antarctic territories by dispersive species dependent on ice-free breeding sites (Fraser *et al.*
18 2009; Van der Putten *et al.* 2010; Fraser *et al.* 2012).

19 Although some marine mammals or seabirds from Antarctic and sub-Antarctic areas have shown
20 dispersal responses to long-term climatic variations (e.g. southern elephant seal *Mirounga leonina*,
21 see de Bruyn *et al.* 2009, or Adélie penguin *Pygoscelis Adeliae*, see Roeder *et al.* 2001), the future
22 rate of environmental changes induced by a warming climate may outpace the ability of most upper-
23 trophic-level predators to adjust. In this context, projections simulated according to future
24 global warming scenarios given by the latest Intergovernmental Panel on Climate Change Fourth
25 Assessment Report (Solomon *et al.* 2007) reveal that the genus *Aptenodytes* may be under serious
26 threat of extinction before the end of the twenty-first century (Le Bohec *et al.* 2008; Jenouvrier *et*

1 *al.* 2012). Indeed, an abrupt increase in sea surface temperature (SST) has been demonstrated to
 2 be detrimental to king penguin survival and reproductive rates in populations breeding in the
 3 Crozet archipelago in the southern Indian Ocean (Le Bohec *et al.* 2008; Saraux *et al.* 2011a).
 4 This is caused by the longer distance the individuals have to swim to get to their main foraging
 5 ground (the Antarctic Polar Front during the summer season) whose latitudinal location is influ-
 6 enced by SST (the higher the SST, the higher the latitude, see Péron *et al.* 2012). However,
 7 when SST was lower, and food resources probably closer, the Crozet Islands were probably not
 8 suitable for the king penguin to complete their breeding cycle as this species needs year-round
 9 ice-free grounds for reproduction.



10

11 **Figure 45 | Winter sea ice extent** today (*pale blue solid line*) and during the Last Glacial Maximum (LGM;
 12 *green dashed line*), current position of the Antarctic Polar Front (*black dashed line*), putative refugia during the LGM
 13 (*solid green line*), and current location of king penguin populations (*red solid line*) and of the colony of 'La Baie du
 14 Marin' on Possession Island, Crozet archipelago, monitored in this study (*red text*); adapted from Fraser *et al.* 2012.

1 Although ca half of the global king penguin population is resident in the Crozet basin, and this
2 top predator represents one of the largest components of the sub-Antarctic marine ecosystem as
3 measured by biomass and energy flux (Charrassin & Bost 2001), the long-term response of this
4 species to the warming period following the LGM remains entirely unknown. Here, we use a ge-
5 nomic-based demographic reconstruction to test whether the king penguin population from the
6 Crozet archipelago ('La Baie du Marin' colony, Possession Island) was strongly affected by the
7 colder conditions characterising the LGM. We tested the applicability of genome-wide data
8 produced by restriction site-associated DNA (RAD) sequencing (Baird *et al.* 2008; Davey *et al.*
9 2011) within a multi-locus coalescent-based framework (Extended Bayesian Skyline Plot, EBSF,
10 see Heled & Drummond 2008) that aims to average the stochasticity of mutation and drift
11 across tens of genomic loci and to overcome the idiosyncrasy of the results obtained from one or
12 a few molecular markers.

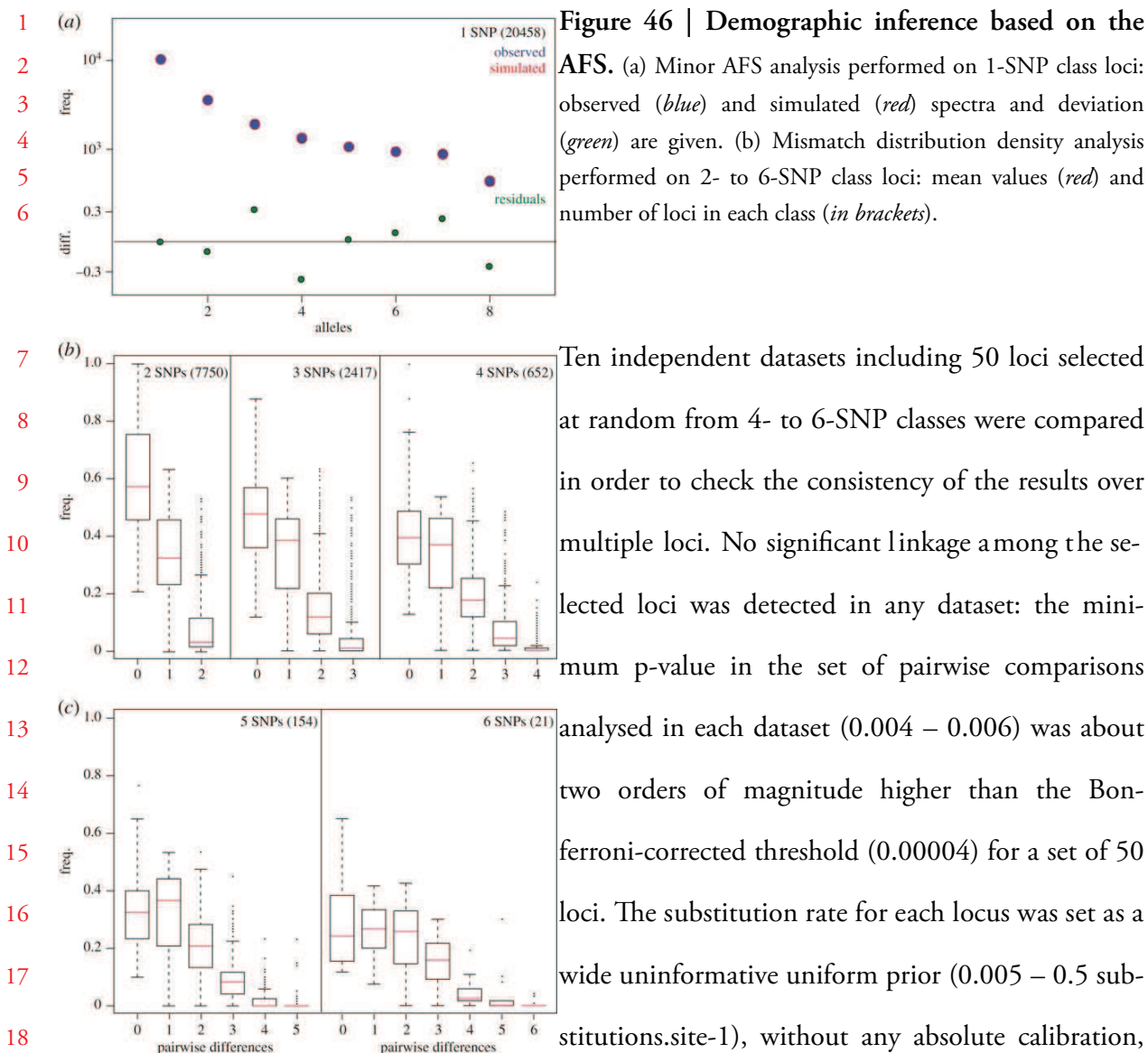
13 We first demonstrated that genomic regions with different levels of polymorphism consistently
14 show a similar pattern of diversity, and that a consistent demographic signature is recognizable
15 across the entire genome, thus providing evidence that a reduced subset of anonymous genomic
16 loci can be safely employed to estimate past population trends. Historic demography was then in-
17 ferred by employing a random selection of RAD loci sequenced in multiple individuals and time
18 calibrated according to the mitochondrial substitution rate as estimated in the Adélie penguin
19 Control Region (CR) (in substitutions.site⁻¹.Myr⁻¹: median = 0.55, 95% interval = 0.29–0.88, see
20 Millar *et al.* 2008). Our data show that the king penguin population from Crozet was strongly
21 reduced on this potentially ice-covered sub-Antarctic archipelago before the end of the LGM, or
22 king penguins may have been totally absent, but they (re)colonized the island as soon as the eco-
23 logical conditions required by this species were met during the following Holocene warming.

1 *Results*

2 ILLUMINA sequencing of a paired-end RAD library from eight king penguins yielded ca 65
3 million, 100 bp reads. Front reads only were used in our analyses and are available on GenBank
4 at the Sequence Read Archive (Run Num.: SRR942341). After quality filtering, trimming of the
5 last 5 bp, and barcode sorting, 101,115 anonymous loci (each as a 95 bp nucleotide sequence)
6 with 50 average coverage were aligned in an unreferenced catalogue. GC content of this dataset
7 was estimated as 50.45%. According to our quality criteria, the catalogue was further filtered to
8 66,172 loci (of which 31 452 were polymorphic) matching all eight individuals without missing
9 data. This dataset was then used in downstream statistical analyses.

10 First, we tested whether genomic regions with different degrees of polymorphism carried consis-
11 tent information about the demographic history of the king penguin population from Crozet.
12 Polymorphic RAD loci were sorted into six different classes according to the number of single
13 nucleotide polymorphisms (SNPs) observed (from 1 to 6 SNPs), and separate analyses were per-
14 formed for each category: 20,458 loci with 1 SNP; 7,750 loci with 2 SNPs, 2,417 loci with 3
15 SNPs; 652 loci with 4 SNPs; 154 loci with 5 SNPs; and 21 loci with 6 SNPs. None of the loci
16 with more than 6 SNPs passed our quality criteria. The signature of the past demography in the
17 1-SNP class was investigated using the allele frequency spectrum (AFS) (Fig. 46 p. 334) in $\delta a \delta i$
18 (Gutenkunst *et al.* 2009). A likelihood ratio test strongly supported a 2-epoch sudden demo-
19 graphic expansion model over a constant-size model ($\chi^2 = 8063$; $p < 0.0001$). When scaled by
20 our estimated mutation rate for the 1-SNP class (see below), estimates of the parameters from the
21 best model indicate a demographic expansion from ca 1,400 to ca 7,000 breeders around 18500
22 years ago. The mismatch distribution density was then checked for the loci in the 2- to 6-SNP
23 classes (Fig. 46 p. 334). A unimodal distribution of mismatches, characteristic of a recent popula-
24 tion expansion, was typical across loci in all SNP classes, though more evident in those classes
25 with higher number of SNPs. Notably, the number of outlier loci (in this case, loci showing high

1 frequency of mismatch at the highest number of differences) was negligible in all classes, high-
2 lighting the consistent mismatch distributions across loci, which is expected in a population that
3 has recently undergone a demographic expansion. The same 2- to 6-SNP classes of loci were then
4 used to reconstruct the demographic function through time by a multi-locus coalescent-based
5 Bayesian approach (EBSP, see Heled & Drummond 2008). Highly consistent results were ob-
6 tained using separate datasets with 2, 3 or 4 – 6-SNP loci (c supplementary material, Fig. 47 p.
7 335). However, loci from 4- to 6-SNP classes were chosen for further analyses because of their
8 higher information content.



19 and the EBSP reconstructions were thus unscaled. The same pattern of population growth is

20 clearly evident in all runs (supplementary material, Fig. 47 p. 335). Test analyses including sub-

21 sets of 10, 25, 50 and 100 loci selected at random from the same set of 100 loci were also run.

22 Smaller subsets resulted in broader credibility regions, although the median value of the popula-

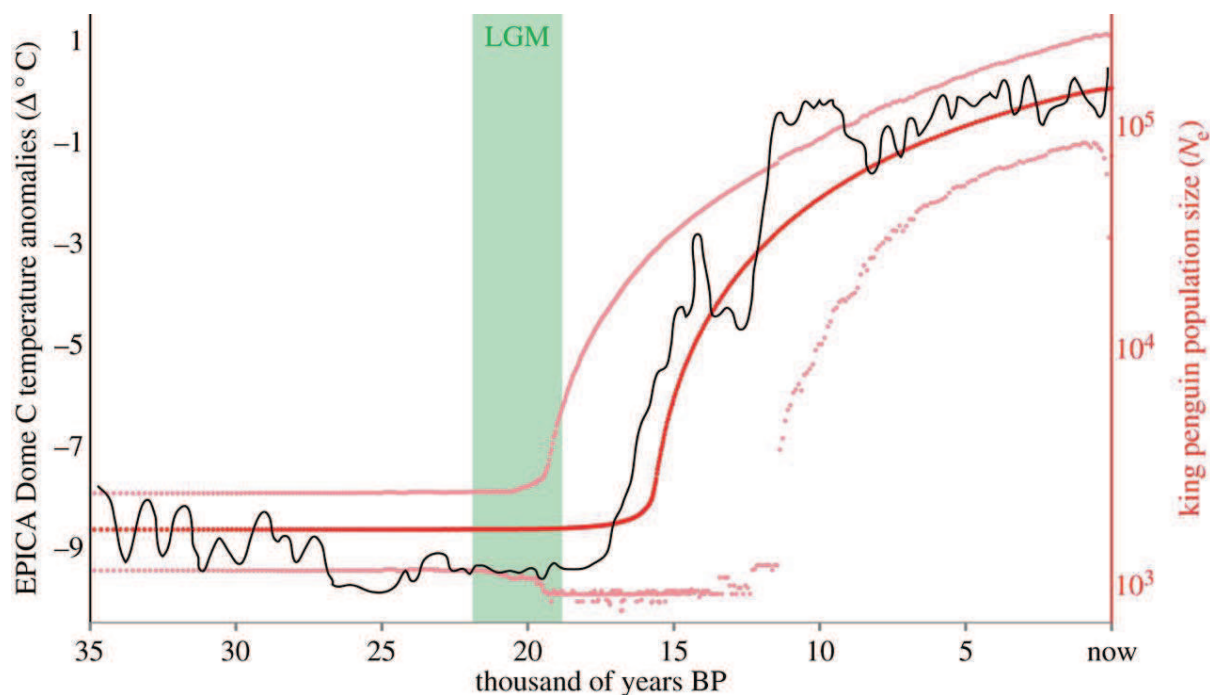
23 tion size through time was not substantially different (supplementary material, Fig. 49 p. 345).

24 All runs showed highly consistent EBSP trends, but the effective sample size of the posterior in-

25 creased very slowly in chain increments from 50 to 200 million generations, showing a general

26 poor mixing. However, likelihood, substitution model parameters and clock model parameters

1 were adequately sampled in all runs. When analysing 100 loci in the same run, the length of the
 2 analysis and the poor mixing of the Markov chain Monte Carlo (MCMC) simulations (meaning
 3 that it should run much longer) were the limiting factors making the demographic analysis im-
 4 practical on our computational resources (see Material and methods).



5
 6 **Figure 47 | Demographic reconstructions of the Crozet king penguin colony** employing the Ex-
 7 tended Bayesian Skyline Plot analysis. Consistency in the pattern inferred is compared among different data selec-
 8 tions including 50 loci chosen at random from different classes of variation: a) 2 SNPs, 4 independent datasets; b) 3
 9 SNPs, 4 independent datasets; c) 4-6 SNPs, 9 independent datasets. In order to facilitate comparison of uncalibrated
 10 EBSP reconstruction and solely for visualization purpose, all runs were scaled to have the same value of the demo-
 11 graphic function at t_0 . To do so, we divided all demographic estimates by the ratio $K_i = N_i(t_0)/\max(N_i(t_0))$, where
 12 $N_i(t_0)$ is the median value from the posterior distribution of the demographic function at t_0 for each run. Corre-
 13 spondingly, the time intervals of each run were multiplied by the same ratio K_i , reflecting the assumption that the
 14 actual population size is the same across all runs. Median (black), 95% HPD lower (red) and upper (green) values
 15 are reported.

16 In order to calibrate our reconstruction, the mitochondrial CR dataset was included in the EBSP
 17 analysis together with 50 randomly selected nuclear loci (4- to 6-SNP classes), using the substitu-

1 tion rate ($0.55 \text{ substitutions.site}^{-1}.\text{Myr}^{-1}$) published for the Adélie penguin (Millar *et al.* 2008).
2 When analysed on its own, the mitochondrial locus supported the demographic expansion of the
3 Crozet king penguin population. Sequences of 354 bp fragments of the mitochondrial DNA
4 (mtDNA) CR were produced in 140 samples (number of haplotypes = 112; haplotypes diversity
5 = 0.997; nucleotide diversity = 0.025) and all new haplotypes were uploaded to GenBank (acces-
6 sion no. KF530582 – KF530720). Demographic statistics (Tajima's $D = 21.75$; Ramos-Onsins &
7 Rozas (R_2) = 0.039*; $F_u F_s = 2168.4^*$; * $p < 0.05$), unimodal mismatch distribution of pairwise
8 differences (not shown), and the Bayesian Skyride analysis (supplementary material, Fig. 50 p.
9 346), all strongly support the hypothesis of a population expansion, thus providing an indepen-
10 dent support to the results obtained from random sets of RAD nuclear loci. The 0.55 substitu-
11 $\text{tions.site}^{-1}.\text{Myr}^{-1}$ substitution rate (considering its upper and lower 95% CR boundaries, Millar *et*
12 *al.* 2008) was then used to scale the Bayesian Skyride plot to an absolute timescale (supplemen-
13 tary material, Fig. 50 p. 346).

14 Our EBSP reconstruction based on a random set of 50 RAD loci and one mitochondrial frag-
15 ment indicated that the king penguin population of 'La Baie du Marin' had a N_e of ca 2,000 in-
16 dividuals during the LGM, and it steeply increased at the onset of the Holocene, around 15,000
17 years ago. This demographic growth closely matched the trend of temperature anomalies record-
18 ed in the Epica Dome C ice core (Jouzel *et al.* 2007), which shows a rapid increase between
19 18,000 and 10,000 years ago (Fig. 48 p. 339). Then, the colony reached an estimated present
20 population size of ca 170,000 individuals. Using the published mtDNA rate as a reference, the
21 average substitution rate of loci in our dataset was estimated at $2.2.10^{-8} \text{ substitutions.site}^{-1}.\text{Myr}^{-1}$
22 ($2.6.10^{-7} \text{ substitutions.site}^{-1}.\text{Myr}^{-1}$). To obtain this estimate, we used the substitution rates esti-
23 mated in BEAST for loci with different numbers of SNPs (mean values of posterior probabilities
24 in $\text{substitutions.site}^{-1}.\text{Myr}^{-1}$: 0 SNP = 0.007; 1 SNP = 0.028; 2 SNPs = 0.050; 3 SNPs = 0.071; 4
25 SNPs = 0.092; 5 SNPs = 0.114; 6 SNPs = 0.135). Note that the number of SNPs at each locus is,
26 in the first place, a result of the stochasticity of the mutational and coalescent process. Hence,

1 mutation rates estimated in BEAST for each SNP class should not be regarded as estimates of the
2 actual substitution rates across different genomic fragments, but rather as posterior probabilities
3 conditional to the number of SNPs at each locus. We then estimated the average substitution rate
4 for our RAD dataset as a weighted average accounting for the frequency of each SNP class across
5 the 66,172 loci.

6 *Discussion*

7 *§-177 Using restriction site-associated DNA sequencing data in a coalescent-based framework.* Our
8 analyses empirically supported that RAD sequencing data can be effectively used for inferring
9 past demography in a coalescent-based framework and that a small subset of these data (i.e. 50
10 loci only) is sufficient to describe the post-glacial history of the king penguin colony breeding on
11 Crozet. This result is particularly important because, though thousands of markers can now be
12 easily and cost-effectively sequenced in any biological system (Davey *et al.* 2011), many existing
13 analytical tools for demographic inference are not yet optimized for genome-level data. Recently
14 published methods, like the pairwise sequentially Markovian coalescence (Li & Durbin 2011) or
15 DiCal (Sheehan *et al.* 2013) require phased diploid genomic data over a long fragment of the
16 chromosome that are not readily available for most non-model species. The development of new
17 algorithms to exploit SNP data in coalescent-based frameworks (i.e. ‘SNPs and amplified frag-
18 ment length polymorphisms phylogenies’, see Bryant *et al.* 2012) are paving the way for a full use
19 of genome-scan data in phylogeography and population genetics, although linkage needs to be
20 taken into account for extensive SNP datasets.

21 We found that EBSP analysis is not really optimized for handling high numbers of loci (very high
22 parametrization) and this may result in slow convergence and poor mixing of the MCMC chains.
23 We estimated that 3–5 billion iterations would be needed to get effective sample size above 200
24 for all model parameters when analysing our dataset of 50 loci and eight individuals. Neverthe-

1 less, our EBSP demographic reconstructions were highly consistent with each other, with those
2 produced by the diffusion approximation of the AFS (Fig. 46 p. 334) and with standard single-
3 locus mtDNA analyses (supplementary material, Fig. 50 p. 346). The AFS analysis, based on ap-
4 proximation to the neutral Wright–Fisher diffusion of an allele (Gutenkunst *et al.* 2009), has
5 been strongly criticized as theoretically unfit to distinguish between competing population histo-
6 ries even if strongly simplified assumptions (panmixia, infinite site mutation model and neutrali-
7 ty) are met in the studied system (Myers *et al.* 2008). In particular, it seems that compensating
8 events in population dynamics are impossible to detect, thus making the analysis of complex
9 histories unreliable. In our test case, the different analytical approaches converged on a quite sim-
10 ple history of sudden population growth making it a likely suitable case for AFS analysis. Both
11 EBSP and AFS analyses, based on independent datasets (1-SNP class versus 4- to 6-SNP classes
12 loci), produced largely similar inferences of the time and trend of the past demography of our
13 king penguin colony. A major difference between results from the two methods was in the esti-
14 mate of the current effective population size ($N_e = 7,000$ from AFS analysis versus $N_e = 170,000$
15 from EBSP). Both estimates are inconsistent with the direct count of breeding birds on ‘La Baie
16 du Marin’ colony, Possession Island (*i.e.* 32,000 breeding birds, see Delord *et al.* 2004). However,
17 it is likely that a certain degree of gene flow exists within the Possession Island, and that ‘La Baie
18 du Marin’ colony is part of a metapopulation at the island level. According to the most recent es-
19 timate by direct count, the breeding population of Possession Island is ca 150,000 individuals
20 (Delord *et al.* 2004). Therefore, the population size we estimated from genomic data may reflect
21 this larger meta-population, rather than the single colony. A study investigating the level of
22 connectivity and gene flow among the colonies of Possession Island is currently ongoing.

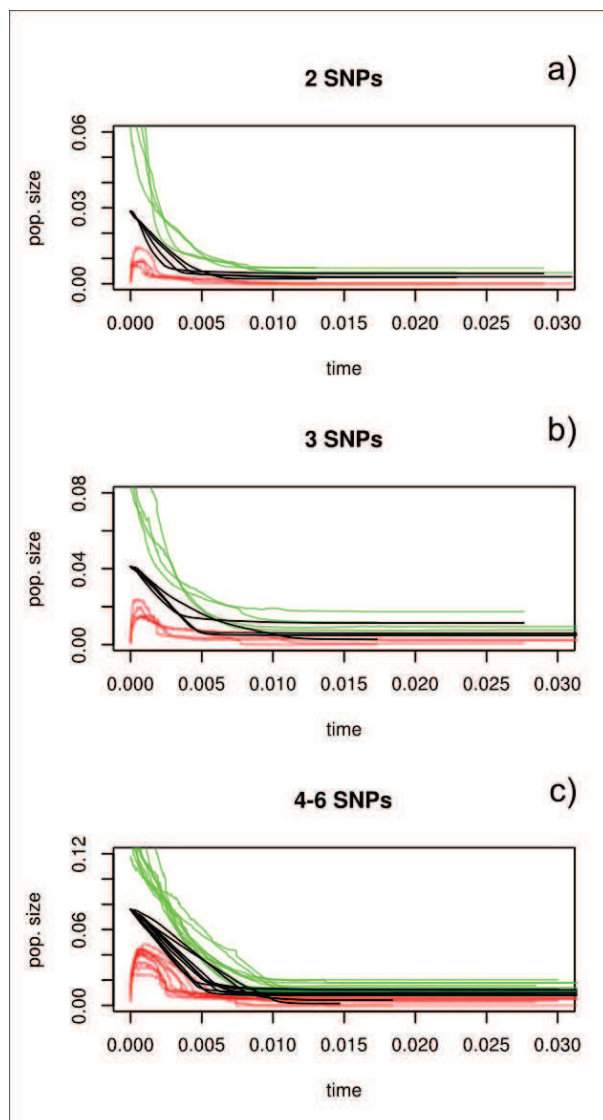


Figure 48 | Past demographic trend of the king penguin colony of 'La Baie du Marin' on Possession Island, Crozet archipelago: median value (red) and 95% confidence interval (pale red). Trend of temperature anomalies recorded in the EPICA Dome C ice core (black Jouzel *et al.* 2007). The LGM period is given in green.

In our test case, the demographic trend inferred from a single mitochondrial locus was consistent with a similar analysis carried out on 50 nuclear loci (Fig. 48 p. 339 and supplementary material, Fig. 50 p. 346). Nevertheless, analysing a fairly large number of (presumably) unlinked loci should be considered as best practice as it allows to us average the stochasticity of mutation and drift over many independent genealogies, thus avoiding the risk of a biased demographic inference owing to the idiosyncratic history of a single marker. In addition, when a large number of

loci are genotyped, many fewer samples are needed to estimate population-level statistics and obtain demographic inferences (Felsenstein 2006; Willing *et al.* 2012).

Thanks to the inclusion of the mitochondrial marker in our genomic analyses, we were also able to propose an average substitution rate for the loci included in our RAD dataset. Our estimated rate is about one order of magnitude faster than previously reckoned in other bird species (Ellegren 2013). However, recent evidence showed that genomic substitution rates can be faster than formerly estimated from pedigree studies (O'Roak *et al.* 2012; Gibb & Hills 2013). Moreover, the RAD sequencing protocol may introduce biases in the genome representation: when using a

1 digestion enzyme like SbfI, whose restriction site is rich in GC, a selection for GC-rich regions
2 occurs. Indeed, the GC content in our dataset was 50.45% that corresponds to the highest esti-
3 mates for chicken microchromosomes (Hillier *et al.* 2004). Hence, our dataset is probably cen-
4 tred on GC-rich regions (e.g. microchromosomes), which are characterized by increased levels of
5 gene density, recombination rate, number of CpG sites, methylation and mutation rate (Axelsson
6 *et al.* 2005). Excess of hypermutable CpG nucleotides in GC-rich sequences has been identified
7 as one possible explanation of increased mutation rates (Zhao & Boerwinkle 2002) so that GC-
8 rich microchromosomes can show a 1.2–1.3 faster substitution rate than the average of the
9 genome (Webster *et al.* 2006). This effect probably contributes, at least in part, at explaining the
10 high mutation rate estimated for our dataset. Even if further investigation is clearly necessary, our
11 interpretation sounds like a cautionary tale about the bias in the selection of genomic regions
12 when using sequence-based restriction enzymes.

13 *§-178 King penguin population history during the Last Glacial Maximum.* Our genomic data from
14 the Crozet king penguin colony contain a consistent signal of demographic expansion at all levels
15 of variability, with very few outliers in the mismatch analysis. Our demographic reconstruction
16 strongly supports a sudden population expansion following the LGM and starting ca 15,000
17 years ago. Therefore, our results show that even a cold-adapted species, such as the king penguin,
18 was limited by colder glacial conditions, and benefited from post-glacial warming, which offered
19 suitable breeding sites and foraging conditions in the Crozet region.

20 Like several other top-level predators in the Antarctic and Sub-Antarctic, king penguins depend
21 on marine ecosystems for food and on terrestrial habitats for reproduction. This species needs a
22 combination of two ecological factors that are both influenced by temperature: *(i)* year-round
23 ice-free breeding grounds, and *(ii)* access to major feeding grounds (e.g. the Antarctic Polar
24 Front) within swimming distance and compatible with birds' breeding duties (exchanging the egg
25 with their fasting partner, feeding their chick). The king penguin has the longest breeding cycle

1 of all penguin species, spanning 14 – 16 months (Barrat 1976). This implies that the reproduc-
2 tive success depends on both summer and winter conditions. Yet, LGM climate strongly influ-
3 enced both marine productivity and the availability of suitable breeding grounds from which
4 feeding areas could be exploited. Lower global temperature favoured a northward expansion of
5 both winter and summer sea ice (Gersonde *et al.* 2005; Fraser *et al.* 2009; Fraser *et al.* 2012). In
6 turn, ice-sheet and sea ice extensions profoundly affected the location (shifted northward) and
7 strength of oceanic fronts and circumpolar currents encircling Antarctica (Charles *et al.* 1991;
8 Gersonde *et al.* 2005; Fraser *et al.* 2009), which determine primary productivity in the Southern
9 Ocean (Thatje *et al.* 2008). Some polar species tracked favourable ecological niches northwards
10 (Burckle 1984; Janko *et al.* 2007; Fraser *et al.* 2012), and suitable refugia for land-breeding
11 species have been proposed on the Falklands, Gough Island and Southern New Zealand. Genetic
12 evidence showed recent (re)colonizations of the sub-Antarctic islands by several taxa that drifted
13 with the strong eastward flow of the Antarctic Circumpolar Current (Waters 2008; Fraser *et al.*
14 2009). However, current reconstructions of LGM conditions lack the spatial resolution to un-
15 equivocally detect the number and geographical distribution of all glacial refugia (Convey *et al.*
16 2009). In particular, it is not clear whether ice-free patches existed in the Crozet archipelago dur-
17 ing the LGM (Hunt 1991).

18 EPICA Dome C ice core (Jouzel *et al.* 2007) data show an increase of ca 9°C between 18,000
19 and 10,000 years ago. In our demographic reconstruction, this interval coincides with a sustained
20 phase of growth for the Crozet king penguin population (Fig. 48 p. 339). This growth trajectory
21 is consistent with a threshold model, where a new colony is established as soon as all of the criti-
22 cal ecological requirements are met. We propose that the availability of breeding grounds within
23 feeding-trip distance from foraging areas played the most critical role in the population expan-
24 sion of Crozet king penguins after the LGM. Even if a major food resource (linked to the pres-
25 ence of the AFP) may have already been available within a king penguin home-range distance
26 from the Crozet islands during the LGM, winter sea ice and possibly land glaciers, did not allow

1 the settlement of a viable king penguin population on the archipelago before the post-LGM
2 warming. As soon as breeding grounds became available, the king penguin could extensively set-
3 tle on Crozet and exploit the food resource. However, with growing SST temperature owing to
4 the global warming, the AFP can soon be too far from Crozet, thus making the area unsuitable
5 again. Extinction, beginning locally and potentially mitigated by migration events (depending on
6 their ability to follow spatial and temporal changes in food/breeding resource availability, as seen
7 in the Gentoo penguin *Pygoscelis papua*, see Ducklow *et al.* 2007), is a probable outcome of the
8 ongoing climate warming.

9 *Material and methods*

10 *§-179 Sampling, DNA extraction and quality assessment.* Sampling was conducted during the 2010
11 breeding season in the king penguin colony of ‘La Baie du Marin’ on Possession Island (46°250
12 S, 51°8450 E), Crozet Archipelago. Blood (approx. 100 ml) was collected from the brachial vein
13 of chicks hatched in the long-term monitored area ‘ANTAVIA’, transferred to a filter paper
14 (Whatman 113), dried and later frozen at -20°C. Individuals were randomly selected along a 120
15 m axis at the periphery of the colony, in order to maximize separation distance. A total of 140 in-
16 dividuals were chosen for mtDNA CR analysis, and eight of these were randomly selected for
17 RAD sequencing analysis. Total DNA was extracted from the filter papers using standard meth-
18 ods and controlled for quantity and quality.

19 *§-180 Restriction site-associated DNA sequencing.* Genomic DNA from eight king penguins was
20 individually barcoded, pooled and genotyped by RAD sequencing (Baird *et al.* 2008) in one li-
21 brary sequenced on an ILLUMINA HiSeq2000 at the Norwegian Sequencing Centre, University
22 of Oslo. Raw reads were trimmed, demultiplexed and aligned in an unreferenced catalogue using
23 the STACKS software pipeline (Catchen *et al.* 2011) running on the server facility on the ABEL
24 cluster, University of Oslo. Further quality filtering using custom bash and python scripts was ap-

1 plied to produce the final dataset used in downstream statistical analyses (detailed protocol in the
2 supplementary material).

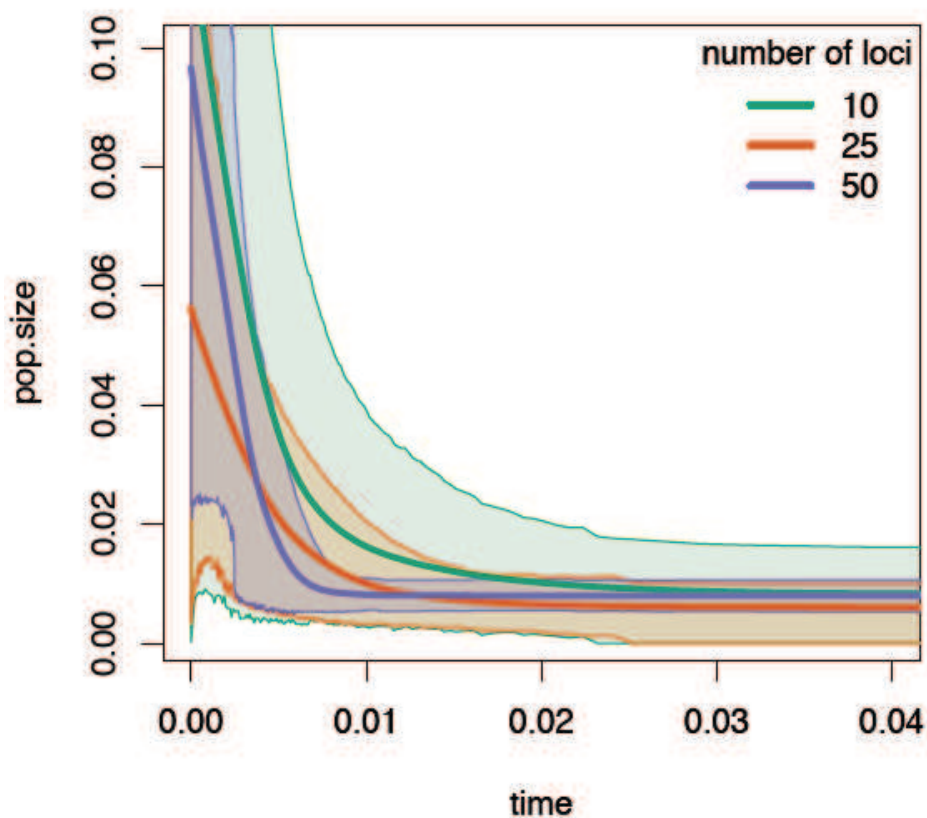
3 *§-181 Genome-wide demographic analysis and calibration.* Genomic RAD fragments (95 bp) were
4 sorted into six different classes according to the number of SNPs observed (from 1 to 6 SNPs)
5 and separate analyses were performed for each category. Minor allele frequency spectra were cal-
6 culated by functions available in the R package ‘adegenet’ (Jombart 2008) using loci included in
7 the 1-SNP class. The signature of the past demography in this class of polymorphism was investi-
8 gated using the AFS analysis in $\delta a \delta i$ (Gutenkunst *et al.* 2009). Using a diffusion approximation
9 of the AFS, this analysis allows demographic inferences to be made from genetic data for testing
10 alternative demographic scenarios in a maximum-likelihood framework. A sudden growth in
11 population size was tested against the null hypothesis of constant population size using the ‘2-
12 epoch’ and the ‘snc’ functions, respectively. The mismatch distribution density (average mismatch
13 distribution of pairwise differences) was then analysed to check for the same demographic pattern
14 in the 2- to 6-SNP classes. Functions included in the R package ‘ape’ (Paradis *et al.* 2004) and
15 the R standard boxplot function were used to estimate and plot the mismatch distribution densi-
16 ty in each SNP class. Random selections of 50–100 loci in the 2- to 6-SNP classes were used to
17 infer the past demography of the king penguin population using the coalescent-based multi-locus
18 analysis implemented in BEAST v. 1.7.4 (Drummond & Rambaut 2007), setting the EBSP
19 (Heled & Drummond 2008) as the tree prior model (see the supplementary material for details).
20 The robustness of the approach was tested with respect to: (i) the number of SNPs per locus, (ii)
21 the different random selection of loci, and (iii) the number of loci included in the random selec-
22 tion. All analyses were run on the Bioportal facility (now LifePortal) of the ABEL cluster, Univer-
23 sity of Oslo. An estimate proposed for the substitution rate of the mitochondrial CR in the Ade-
24 lie penguins ($0.55 \text{ substitutions.site}^{-1}.\text{Myr}^{-1}$, see Millar *et al.* 2008) was used to calibrate our de-
25 mographic reconstruction. We then plotted the population trend of the last 35,000 years togeth-
26 er with the trend of temperature anomalies (in D°C) as inferred by the analysis of the EPICA

1 Dome C ice core (Jouzel *et al.* 2007). Concerning the calibration of the mean genome-wide sub-
2 stitution rate: first, the mean of the MCMC posterior median values for each SNP class included
3 in the selected EBSP analysis (4- to 6-SNP classes) was calculated; then, a linear regression was
4 used to infer the substitution rate of those SNP classes excluded from the final EBSP analysis (0-
5 to 3-SNP classes); finally, we calculated the mean genomic substitution rate weighting each SNP
6 class according to the frequency (number of loci) of each class.

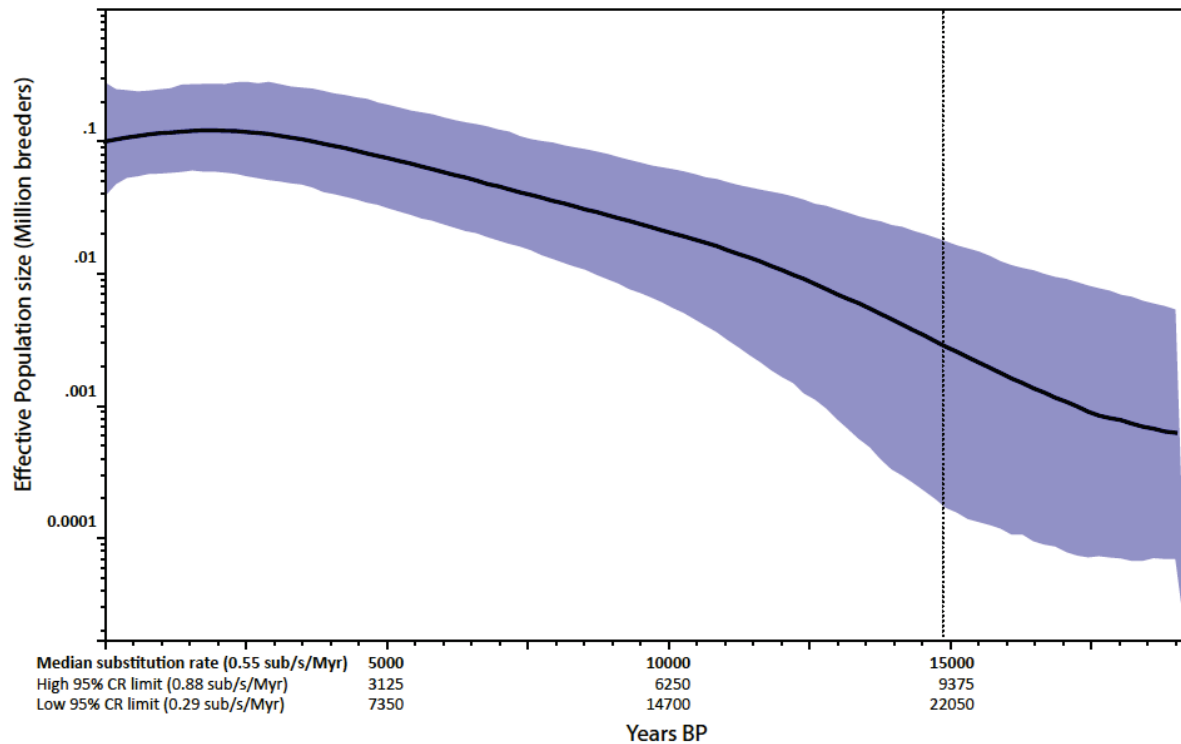
7 *§-182 Mitochondrial DNA control region analysis.* Partial sequences of the CR (354bp) were am-
8 plified and sequenced in 140 samples according to the protocol published in Heupink *et al.*
9 2012. Summary molecular statistics (Haplotype diversity: Hd; nucleotide diversity: pi), demo-
10 graphic parameters as Tajima's D and Fu's Fs, and the mismatch distribution of pairwise differ-
11 ences were calculated. This dataset was used to infer the king penguin past demography employ-
12 ing the Bayesian Skyride plot (Minin *et al.* 2008) reconstruction, which produces estimates of
13 population size through time and its associated credibility intervals, combining both phylogenet-
14 ic and coalescent uncertainties, as implemented in BEAST v. 1.7.4. The estimate of the CR sub-
15 stitution rate proposed for Adélie penguins ($0.55 \text{ substitutions.site}^{-1}.\text{Myr}^{-1}$, see Millar *et al.* 2008)
16 was used to calibrate our analyses.

17 All animals in this study were handled only once in order to mark them with a plastic tag (Floy-
18 tagw), to blood-sample them, and to conduct morphological measurements. All procedures em-
19 ployed during this fieldwork were approved by the Ethical Committee of the French Polar Insti-
20 tute (Institut Paul Emile Victor 2 IPEV) and conducted in accordance with its guidelines, also
21 complying with French laws including those relating to conservation and welfare. Authorisations
22 to enter the breeding site (permits n82009-57 issued on the 26th of August 2009) and handle
23 birds (permits n82009-59 issued on the 29th of August 2009) were delivered first by the French
24 "Ministère de l'Aménagement du Territoire et de l'Environnement" and then by the Terres Aus-
25 trales et Antarctiques Françaises (TAAF).

1 Handled animals were removed from the colony in order to minimise the disturbance to neigh-
 2 bouring birds and taken to a few meters away for manipulation. They were hooded to reduce
 3 their stress and manipulations lasted between 5 and 10 minutes. All blood-sampling (quantities
 4 adjusted according to the age of the chick) and tagging material was sterilized (either sealed, or
 5 through chemical sterilisation). Moreover, Vétédine soap and alcoholic antiseptic solutions were
 6 used to disinfect the skin before bleeding and tagging. Flesh wounds did not seem infected there-
 7 after (personal observations on a subset of recaptured birds).



8 **Figure 49 | Demographic reconstructions of the Crozet king penguin colony** employing the Ex-
 9 tended Bayesian Skyline Plot analysis. Consistency of the inference is compared across three nested datasets (10, 25
 10 and 50 loci) randomly selected from the 4-6-SNP class. Population size and time are unscaled. Median (solid line)
 11 and 95% upper and lower (filled areas) values are reported.



1 **Figure 50 | Bayesian Skyride Plot inferred from mitochondrial Control Region data** describing
 2 population trend through time. Time is scaled according to median (solid black line), 95% upper and lower credibil-
 3 ity region (filled blue area), as estimated in Millar *et al.* 2008.

4 *§-183 Acknowledgements.* We are very grateful to H el ene Gachot-Neveu and Magali Beaugey for
 5 their help in the laboratory. We also thank M. Le Vaillant, O. Prudhomme and M. Ripoche for
 6 the sample collection. We are very grateful to Sissel Jentoft and core members of the Centre for
 7 Ecological and Evolutionary Synthesis, University of Oslo, for the support in implementing the
 8 RAD sequencing platform and to Sanne Boessenkool for helpful comments on early versions of
 9 the manuscript. We also thank the editors and two anonymous referees for their very helpful
 10 comments. **Data accessibility.** RADseq data: Sequence Read Archive (Run Num.: SRR942341).
 11 MtDNA data: GenBank accession nos. KF530582– KF530720. **Funding statement.** This study
 12 was supported by the Institut Polaire Fran ais Paul-Emile Victor (Programme 137—ECOPHY),
 13 the Centre National de la Recherche Scientifique (Programme Zone Atelier de Recherches sur

1 l'Environnement Antarctique et Subantarctique), the Fondation de France/Fondation Ars Cuttoli
2 (to J.D.W.) and Marie Curie Intra European Fellowships (FP7-PEOPLE-IEF-2008, European
3 Commission; project no. 235962 to C.L.B. and FP7-PEOPLE-IEF-2010, European Commis-
4 sion; project no. 252252 to E.T.).

5 *Supporting information*

6 *§-184 Sampling, DNA extraction, quality assessment.* Sampling was conducted during the
7 2010/2011 summer season in the king penguin colony of 'La Baie du Marin' on Possession Is-
8 land (46°25'S, 51°45'E) in the Crozet Archipelago. Blood (~100 µL) was collected from the
9 brachial vein of chicks hatched in the long-term monitored area 'ANTAVIA', transferred to a fil-
10 ter paper (Whatman 113 °), dried, and later frozen at -20°C. Individuals were randomly selected
11 along a 120m-axis at the periphery of the colony, in order to maximise distance separation. A
12 total of 140 individuals were chosen for mitochondrial DNA Control Region analysis, and 8 of
13 these were randomly selected for restriction site-associated DNA (RAD) sequencing analysis.
14 Total DNA were extracted from the filter papers using a Phenol-Chloroform protocol or the Qia-
15 gen DNAase blood & tissue kit according to manufacturer's instructions. After extraction, DNA
16 quantity and quality were tested in each sample by fluorimetric-based measurement (Qubit, In-
17 vitrogen) and gel electrophoresis.

18 *§-185 MtDNA marker analysis.* Partial sequences of the Control Region (354 bp) were amplified
19 and sequenced in 140 samples according to the protocol published in Heupink *et al.* 2012. PCR
20 products were Sanger-sequenced in the ABI-LAB at the University of Oslo. Sequences were then
21 manually edited and aligned in Bioedit (Hall 1999). All new haplotype sequences have been up-
22 loaded to GenBank (Accession number: KF530582-KF530720). Summary molecular statistics,
23 demographic parameters and the mismatch distribution of pairwise differences were calculated in
24 DNAsp v5 (Rozas & Rozas 1999). This dataset was used to infer the king penguin past demogra-

1 phy employing the Bayesian Skyride plot (Minin *et al.* 2008), where inferred population history
2 is bounded by credibility intervals that combine phylogenetic and coalescent uncertainties, as im-
3 plemented in the BEAST 1.7.4 package (Drummond & Rambaut 2007). Analyses were per-
4 formed on the Biportal facility (now LifePortal) running on the ABEL cluster, University of
5 Oslo. A GTR+G+I substitution model was set for the mitochondrial sequence. A relaxed uncor-
6 related log-normal clock prior was set for the substitution rate to take into account fluctuations
7 of the molecular clock along different branches of the phylogeny; a log-normal priors with mean
8 in the real space of 0.55substitution/site/Myr respectively were set (Millar *et al.* 2008). The
9 Bayesian Skyline plot was set as coalescent tree prior model. Convergence among three runs, with
10 a MCMC length of 30 million generations for each parameter setting was checked. Effective
11 sample size was checked in Tracer 1.5 (Rambaut & Drummond 2005) and plots of population
12 size change through time were drawn.

13 *§-186 RAD sequencing and genome-wide demographic inference.* Eight king penguin individuals
14 were pooled and genotyped by RAD sequencing [8] in one library sequenced on an ILLUMINA
15 HiSeq2000, yielding ca. 65 million 100-bp reads. All raw sequence reads are available on Gen-
16 Bank at the Sequence Read Archive (Run Num.:SRR942341).

17 After quality assessment, samples showing high molecular weight and highly concentrated DNA
18 were employed in next-generation sequencing (NGS) of RADtags (Baird *et al.* 2008). The follow-
19 ing RADseq protocol was adopted: (i) approximately 100ng of genomic DNA pers ample were
20 digested with the restriction enzyme SbfI (NEB); (ii) each sample was then ligated to a unique
21 barcoded P1 adapter prior to pooling in a single library. The library was then sheared by sonica-
22 tion, and gel electrophoresis of small library aliquots were run after the first 5 cycles (30" ON –
23 30" OFF) and then every 2-3 cycles of sonication; (iii) the target size range fraction (300-500
24 bp) was achieved after 8 cycles of sonication and was then selected by gel electrophoresis and
25 manual excision; (iv) before size selection on the gel,sonicated libraries were concentrated to 25

1 μl by DNA capture on magnetic beads (beads solution:DNA = 0.8:1), thus further reducing the
2 carry-over of non-ligated P1 adapters; (v) capture on magnetic beads using the same beads:DNA
3 ratio (0.8:1) was then employed in all following purification steps (after blunt-end repairing,
4 poly-A tailing, P2 adapter ligation and library enrichment by PCR); (vi) PCR amplification was
5 performed in 8 x 12.5 μl aliquots pooled after the amplification in order to reduce amplification
6 bias on few loci due to random drift; (vii) the library was then quantified by a fluorimetric-based
7 method (Qubit, Invitrogen) and molarity was checked on an Agilent Bioanalyzer chip (Invitro-
8 gen). A final volume of 20 μl with a DNA concentration of 45 ng/ μl was submitted for sequenc-
9 ing on an ILLUMINA HiSeq2000 sequencer at the Norwegian Sequencing Centre, University of
10 Oslo.

11 Raw reads were then processed using the scripts included in the Stacks package (Catchen *et al.*
12 2013) running on our server facility on the ABEL cluster, University of Oslo. Raw reads were
13 quality filtered and grouped according to individual barcodes. Then individual loci were retrieved
14 and SNPs were called by a maximum-likelihood function that excluded likely sequencing errors.
15 Several runs with different settings of read trimming parameter, quality thresholds, mismatches al-
16 lowed when building the individual and the population catalogs, were performed to check for
17 consistency of the results. The parameters setting used to build the final catalog included: -t 95
18 and the default values for the quality checking when using “process_radtags.pl”; -m 10, -n 7, -M
19 3 when running “denovo_map.pl”. 101,115 loci with 50X average coverage were aligned in an
20 unreferenced catalog. A table including all loci matching the eight sequenced individuals was
21 built using “export_sql.pl” Stacks script. This table was further filtered by python scripts (avail-
22 able upon request) excluding loci with missing data, with more than 2 alleles per individual, and
23 deleveraged by Stacks algorithm.

24 Loci were then grouped according to the number of SNPs allowing from 0 to a maximum of 6
25 substitutions per locus (0 to 6-SNP classes). Loci with 4-6 SNPs were then directly checked

1 through the catalog web-based interface provided by Stacks. Loci with more than 2 SNPs in the
2 last 5 base pairs or with observed heterozygosity higher than 0.6 were blacklisted and removed
3 from the table as likely sequencing errors or paralogous loci. Only those loci hosting 1 single bi-
4 allelic SNP were employed in AFS analysis in order to minimise linkage among the data. Not
5 having a reference genome, we could not exclude loci produced by adjacent genomic regions or
6 by the two sides of each restriction site. Custom python scripts (available upon request) were em-
7 ployed to edit this 1-SNP dataset as a suitable input file for downstream statistical analysis encod-
8 ing SNPs as 0-2 when homozygote for the two alleles respectively or 1 when heterozygote. On
9 the other hand, loci in 2 to 6-SNP classes were treated as short sequences and locus-by-locus
10 edited using python script as NEXUS format files each containing 16 sequences 95 bp long (two
11 sequences per individual).

12 Minor allele frequency spectrum was calculated by functions available in the R package “ade-
13 genet” (Jombart 2008) using loci included in the 1-SNP class. This information was then passed
14 to the python-based software ∂adi (Gutenkunst *et al.* 2009) that, using a diffusion approximation
15 to the allele frequency spectrum, allows demographic inference from genetic data testing alterna-
16 tive demographic scenarios in a maximum-likelihood framework. A sudden growth in
17 population size was tested against the null hypothesis of constant population size using
18 the “two_epoch” and the “snc” functions, respectively. Several runs of likelihood optimisation
19 were performed changing the extent of the search by the “fold” parameter in the “dadi.Misc.per-
20 turb_params” function. Optimised log-likelihood and Theta values were recorded. In order to
21 calculate effective population size from Theta values produced by ∂adi , a total sequence length of
22 1,943,510 bp (95 bp X 20,458 loci used in this analysis) was used.

23 Functions included in the R package “ape” (Paradis *et al.* 2004) and the R standard boxplot func-
24 tion[13] were used to calculate the joint mismatch distribution in the pairwise differences (from

1 here onwards referred to as mismatch distribution density). Calculations were performed and
2 plotted in each 2 to 6-SNP classes separately.

3 Different random combinations of 50-100 loci in 2 to 6-SNP classes were compared when infer-
4 ring the past demography of the king penguin population using the coalescent-theory based mul-
5 ti-locus analysis implemented in BEAST 1.7.4. Linkage disequilibrium was tested in all subsets
6 using Genepop (Raymond & Rousset 1995) with the default setting in the web tool and the
7 Bonferroni correction for multiple tests. The robustness of the approach was tested with respect
8 to i) the number of SNPs per locus, ii) the different random selection of loci and iii) the number
9 of loci included in the random selection: 50 loci in 2-SNP class (5 runs), 50 loci in 3-SNP class (5
10 runs), 50 loci in 4-6-SNP class (10 runs), 10 loci in 4-6-SNP class (1 run), 25 loci in 4-6-SNP
11 class (1 run) and 100 loci in 4-6-SNP class (1 run). Three runs showing hints of multiple optima
12 for the demographic function were discarded. Different settings of the parameters and priors have
13 been explored in preliminary analyses, but the following was the definitive setting: (i) markers
14 were unlinked concerning site substitution model, clock model and tree prior model; (ii) site sub-
15 stitution model was set as a HKY with empirical base frequency; (iii) a strict molecular clock was
16 estimated for each marker with a uniform prior distribution bounded within 0.5 and 0.005 sub/
17 s/Myr; (iv) the Extended Bayesian Skyline Plot (EBSP; Heled & Drummond 2008) was selected
18 as tree prior model and ploidy of the markers was set accordingly. Fine tuning of operators did
19 not improve our results as running the analyses with longer MCMC simulations; 200 million it-
20 erations were set as run length.

21 Mitochondrial Control Region data were included in the analysis in order to test the consistency
22 of the information provided by the two genetic dataset (genomic and mitochondrial) and to cali-
23 brate the genomic substitution rate using the mtDNA ControlRegion substitution rate as esti-
24 mated in the Adélie penguin. Site substitution and clock models were set as in the analyses of the
25 mitochondrial marker alone (see below). All analyses were run on the Biportal facility (now

1 LifePortal) of the Abel cluster, University of Oslo. Results were checked on Tracer 1.5 and plot of
2 the EBSP data were drawn in R. An extensive study on Adélie penguin ancient DNA suggested a
3 fast estimate (0.88 sub/s/Myr; Lambert *et al.* 2002) for the substitution rate of the mitochondrial
4 Control Region. Further analyses confirmed this high figure but it was downscaled to 0.55 sub/s/
5 Myr (Millar *et al.* 2008). We used the more conservative 0.55 sub/s/Myr for our calibrated de-
6 mography. A generation time of 11.49 years (Le Bohec *in prep*) was used to convert the popula-
7 tion size estimates on the EBSP(given by default in effective population size * generation time).
8 We then plotted the population trend for the last 35,000 years together with the trend of temper-
9 ature anomalies as inferred by the analysis of the EPICA Dome C ice core (Jouzel *et al.* 2007).
10 Concerning the calibration of the mean genome-wide substitution rate: first, the mean of the me-
11 dian values in each SNP-ratio class included in the EBSP analysis was calculated (4 to 6-SNP
12 classes); then, a linear regression was used to infer the substitution rate of those SNP classes ex-
13 cluded from the EBSP analysis (0 to 3-SNP classes); finally, we calculated the mean genomic sub-
14 stitution rate weighting each SNP class accordingly with the frequency (number of loci) of each
15 class.

1 RAD-seq bench protocol optimisation.

2 In order to maximise sequencing yield and efficiency, we conducted thorough protocol optimisa-
 3 tion. We used Lambda phage DNA as a control (50,000 genomic DNA, NEB): Lambda phage
 4 has 5 known restriction sites for Sbf-I, and known restriction fragment lengths, which makes it
 5 possible to assess preparation efficiency. In order to assess potential inhibitor effects, all tests were
 6 repeated with a) lambda DNA alone, b) lambda DNA + penguin DNA at equal concentrations,
 7 c) penguin DNA alone.

8 (i) *Digestion* was tested both with the recommended NEB protocol, and with Baird *et al.* (Baird
 9 *et al.* 2008) modified protocol (reduced enzyme molarity). No difference could be found, and
 10 high-molecular-weight lambda-phage had disappeared totally in both cases. Increasing digestion
 11 time from 45min to overnight did not increase efficiency.

12 (ii) *P1 ligation*. We tested four different conditions, removing in each test condition one of the
 13 three components: (a) $\lambda + P1 + T4$, (b) $\lambda + P1 + \emptyset$, (c) $\lambda + \emptyset + T4$, (d) $\emptyset + P1 + T4$.

14 - Overall efficiency: ligation was first performed under recommended conditions (NEB concen-
 15 trations, temperatures and times). Re-ligation of digested DNA (condition (c)) appeared sub-op-
 16 timal, as only a fraction of high-molecular-weight DNA was re-formed.

17 - Time and temperature: tests were then repeated with three different conditions: (a) 25°C for 45
 18 min (recommended), (b) at 16°C for 16 h (NEB test conditions) and (c) cycling between 10°C
 19 and 30°C (2°C increments every 30'') for 16 h (Lund *et al.* 1996). Under all test mixes, ligation

1 efficiency clearly increased from time conditions (a) to (c). We retained conditions (c) as optimal
2 for our updated protocol.

3 - P1 adapters dimers: in order to test the efficiency of P1 adapter ligation, we did a series of dilu-
4 tions (1:1, 1:10, 1:100) on all ligation conditions containing P1 adapters ((a) λ + P1 + T4, (b) λ
5 + P1 + \emptyset , and (d) a) \emptyset + P1 + T4). We performed 25 short cycles (25") of amplification using
6 only forward Solexa primers, that reverse-complement a section of P1 sequence. We thus expect
7 all P1-containing DNA fragments to be linearly amplified; and the band strength on a 1.5%
8 agarose gel should be proportional to initial concentration. Concentration of P1 dimers reflects
9 the efficiency of the ligation reaction. As expected, P1 dimer concentration increased from 25°C
10 incubation, to 16h cycle conditions.

11 - P1 adapter molarity: in order to test if current P1 molarity is simply limiting the reaction, we
12 performed ligation under «ideal» (cycled) conditions, with 0.1 pmol, 0.5 pmol (Baird *et al.*
13 2008), 1.0 pmol, 5 pmol, 10.0 pmol, 20 pmol, 50 pmol, and 100 pmol of P1 adapters per reac-
14 tion. In order to assess the efficiency of the P1 ligation process, we measured the molarity of
15 high-molecular-weight chimeric lambda phage re-constructs. Indeed, we consider the reaction
16 efficient when all available genomic DNA sticky-ends are ligated to a P1 adapter, which requires
17 these to be in large excess: under those conditions, genomic DNA religation should be minimal.
18 Fragment length and molarity were assessed on a TapeStation (Agilent (c)) genomic DNA chip.
19 With 0.1 pmol P1, chimeric religation is very important, but seems to decrease with increasing
20 P1 molarity. Saturation is reached at 5 pmol: chimeric DNA has disappeared, and P1 adapter
21 dimer molarity starts to increase. This new molarity is 10 times higher than recommended by
22 (Baird *et al.* 2008), and leads to enhanced library efficiency. However, when testing our stock P1
23 adapters on a BioAnalyzer DNA chip (Agilent (c)), it was found to be 10 times less concentrated
24 than expected from the manufacturer's specifications. Moreover, when using a fresh stock of
25 adapters, 0.5 pmol proved sufficient to obtain the same result. The most likely hypothesis here is

1 thus a manufacturer's failure - a possibility that we had not taken into account at first, but which
2 may be worth testing on all new adapter stocks.

3 (iii) *P2 ligation*. The same set of conditions tested for P1 ligation was assessed for P2 ligation, but
4 with no noticeable effect on library efficiency. P2 ligation was therefore considered already
5 optimal.

6 (iv) *Purification steps*. The original protocol includes a series of purification steps, realised with a
7 spin-column kit (Qiagen Minelute (c)). These cause massive DNA loss along the preparation. We
8 tested an alternative protocol, in which magnetic beads (Agencourt AMPure XP (c)) are added to
9 the library after size selection, and kept together with the library until the final purification step.
10 Solid-state reversible immobilisation (SPRI) beads are inert, paramagnetic carboxyl-coated parti-
11 cles. Reversible precipitation of DNA on the beads happens under high-salt, low-DNA-solubility
12 conditions, which are created by adding a high-salt polyethylene-glycol (PEG-8000) solution to
13 the library. Under low-tonicity conditions, DNA is dissolved in the surrounding buffer, and
14 beads do not interfere with enzymatic reactions. In order to implement this, we prepared a bind-
15 ing solution with 10 g PEG-8000, 7.3 g NaCl, and 50 mL H₂O. We calibrated its DNA precipi-
16 tation efficiency by purifying 50bp DNA ladder with a gradient of binding solution to ladder ra-
17 tio ranging from 0.3X to 1.4X. Purified product was assessed on a BioAnalyzer DNA 1000 chip
18 (Agilent (c)). For our solution, a ratio of 1 (equal volumes of binding solution and target DNA
19 solution) was found to provide the appropriate size selection for RADseqapplications, by binding
20 all DNA fragments longer than ~200bp.

21 Using this updated purification method, DNA was re-immobilised on the magnetic beads 3
22 times (after end repair, A-tailing, and P2-ligation), and beads are removed at the final step. DNA
23 loss was found to be greatly reduced, and library preparation efficiency (i.e. molarity of properly
24 prepared fragments compared to initial DNA input) was approximately doubled. An additional
25 benefit is the reduction in handling time, and in costs, given the high price of SPRI beads.

Annexes - §186

- 1 (v) *Size selection*. The original protocol (Baird *et al.* 2008) included a manual DNA excision from
- 2 an agarose gel. This method was found to lack precision and repeatability, especially due to the
- 3 inaccurate migration of several DNA ladders. It was replaced by an automated size selection sys-
- 4 tem (BluePippin (c)), with higher DNA recovery and more precise size range as a result.
- 5 (vi) *New protocol*. All these optimisations have been gathered in a revised bench protocol *Updated*
- 6 *RAD-sequencing protocol*, p. 357.

1 Updated RAD-sequencing protocol

2 This is a modified bench version of the original protocol by Paul D. Etter Baird *et al.* 2008. We
3 found it to improve the overall yield and efficiency of the process, as well as speeding it up. As it
4 is, the library preparation proper can be performed in two (long) days, not including library vali-
5 dation. A typical workflow would be (1) DNA extraction optimization – that may take a while.
6 (2) On the first run, some steps need parameter optimization. This may take some time too. (3)
7 Sample selection, library planning, DNA quantification – one day. (4) Library preparation Stage
8 1 – one day. (5) Library preparation Stages 2 and 3 may be combined in one day. The only limit
9 to this is that the whole beads-in Stage 3 should not be stopped overnight, but rather performed
10 in one stretch.

11 Required reagents and stock solutions

12 **Adapter alignment buffer:** Prepare a 10X stock that will be kept at +4°C.

13 - 500 mM NaCl

14 - 100 mM Tris-Cl

15 - adjust pH to 7.5-8.0 with H₂SO₄ (or NaOH).

16 **Binding solution:** Prepare a 1X stock that will be kept at +4°C. An aliquot will be brought to
17 room temperature for library preparation. Polyethylene-glycol with high NaCl allows the re-pre-
18 cipitation of DNA on magnetic beads. The final concentration of NaCl determines the lower se-
19 lected DNA size clip-off. Is it critical to calibrate each batch of this solution so that only DNA

Annexes - §186

1 fragments smaller than 200bp are left out. The ratio between the (theoretically salt-free) DNA so-
2 lution and the binding solution that allows for proper size selection varies from batch to batch,
3 we call it k along the protocol. k should be anywhere between 0.8 and 1.2.

4 - 10g polyethylene-glycol (PEG-8000)

5 - 7.3g NaCl

6 - 50mL H₂O

7 - mix until the solution is totally clear.

8 How to calibrate that : Dilute 50bp DNA ladder to 0.5X in EB buffer, make 12*50µL aliquots.

9 Add 75 µL AmpureXP beads to each, mix for 10min. Rinse twice in fresh 80% EtOH according

10 to the standard protocol. Dry to near-cracking point. Resuspend each bead pellet in 50 µL EB

11 buffer. Add binding solution in increasing volume, from 15 µL (0.3X) to 70 µL (1.4X). Mix for

12 10min. Repeat the rinsing and drying steps. Elute in 50 µL EB buffer. Run on a BioAnalyser

13 DNA1000 chip. Appropriate ratio should cut the ladder around 200bp without affecting the

14 molarity of the 250bp peak.

15 **DNA extraction and purification:** We will need high-quality, pure genomic DNA - Qiagen

16 DNeasy do fine for animal blood or tissue samples, but other methods perform just as well.

17 Check size range on a 1% agarose gel: there should be a clear, concentrated band above 10 kb. If

18 in doubt, perform additional bead cleaning with stringent size selection to keep only high mole-

19 cular weight DNA.

20 **Setting up the library:** Starting setup : small pools seem better for diversity. Say 12 Gb genome

21 per pool (*e.g.* 8 penguins of 16 tuna fishes). Take enough for 2 sonication runs if possible, *i.e.*

22 each pool should sum up to 6 µg DNA (8 samples with 750 ng each for ex).

23 **STAGE 1 : Barcoding steps**

1 **Dilution:** Prepare dilutions in PCR strips. For each sample, dilute the chosen amount of DNA
2 up to 40 μL . If DNA concentrations range widely (more than a 1:3 ratio between the extremes),
3 it is safer to do a pre-dilution of the more concentrated samples in EB buffer, in order to keep the
4 Tris concentration roughly equal in all samples.

5 **Restriction:** Performed in a total volume of 50 μL with 1X NEB Cutsmart buffer. Conditions
6 may need to be adjusted for different restriction enzymes. If the enzyme can't be heat-inactivated,
7 it is not necessary to purify it, but one should be careful to choose P1 adapters that will not recre-
8 ate the cutsite, in case there would be residual restriction activity at the lower ligation tempera-
9 ture. I.e. if using PstI-HF (cutting on GTGCA|C), one should avoid P1 adapters whose barcodes
10 end with a C base, as these will reconstruct the cutsite, and P1 adapter may be re-cut as soon as
11 ligated into the next step.

- 12 - 5 μL 10X Cutsmart buffer
- 13 - 4.5 μL H₂O (or qsp 10 μL)
- 14 - 0.5 μL SbfI-HF
- 15 - 37°C for 90min (restriction)
- 16 - 80°C for 20min (inactivation)
- 17 - cooldown at -1°C per minute to 8°C (re-hybridization).

18 **Ligation:** Determine appropriate amount of P1 adapters. This depends on the number of cutsites
19 and the amount of DNA. For SbfI on the penguin genome : 5 μL at 2 μM for 1 μg DNA.

20 How to calibrate that : in the initial setup, pick a representative sample with enough excess
21 DNA. Prepare 8 digestions, ideally with the planned final amount of DNA (eg 750 ng). After di-
22 gestion, pool them, mix, and re-aliquot into 8x50 μL reactions. Keep a sample as a digestion con-
23 trol. Add a gradient of P1 concentrations, starting at 0. Perform ligation on all except the digest
24 control. Run the product on a TapeStation gDNA chip. Concentration is good when no higher-
25 molecular-weight (religated) DNA is observed. Highest possible molecular weight should be visi-

Annexes - §186

1 ble in the ligation with no P1. Pay attention to the apparition of P1 dimers around 130bp : when
2 they start increasing noticeably, P1 has reached saturation. Just-saturated P1 is fine.

3 *NOTE: with the Vienna adapters, use 5µL of 0.1µM adapters per sample.*

4 Into each digested product, add the barcoded P1 adapter (it helps to prepare the P1 dilutions in
5 PCR strips and use a multichannel pipette to add them to the reactions).

6 - n µL P1.

7 - mix thoroughly.

8 Adapters must be added before the ligation premix, in order to avoid gDNA religation.

9 - 1 µL NEB Cutsmart buffer

10 - 0.6 µL rATP (100 nM)

11 - 0.5 µL T4 ligase (2,000,000 U/mL)

12 - H2O qsp (7.9-n) µL

13 Add (10-n) µL of this mix to each reaction. Mix thoroughly by pipetting (no vortexing). Liga-
14 tion is done overnight with a cycle protocol :

15 - [10°C +2°C/30" until 30°C, then -2°C/30" until 22°C]

16 (96 cycles of 10min, 16h total)

17 - 65°C for 20'

18 - -1°C per minute down to 8°C.

19 This takes about 17 hours to run and should be performed overnight. The increase in efficiency
20 compared to room temperature ligation is worth it.

21 **STAGE 2 : Shearing and selection**

1 **Pooling:** At this stage, every sample is hopefully uniquely barcoded. Aliquots from each sample
2 can be combined in a single tube, for a total of 3 µg DNA per pool. Volume will vary depending
3 on initial DNA concentration. The rest of the samples is better not pooled in case barcoding
4 needs to be re-done, and can be frozen until then.

5 **Sonication:** Performed on the Bioruptor at 2°C, high power. Conditions need to be calibrated for
6 each type of organism and gDNA quality. For penguins, 7 cycles of 30"ON/30"OFF gives opti-
7 mal results. Each TPX tube can sonicate a maximum of 300 µL : if the pool is larger than this,
8 split it across several tubes.

9 How to calibrate that : pick a representative sample with excess DNA, take a series of 1.5 µg
10 DNA aliquots in a volume of 200 µL, and try a range of cycle numbers (6-9-12-15 on the first
11 try). Purify the product into 30 µL EB buffer and run on a 2% agarose gel. The best range should
12 put the bulk of the DNA between 300 and 600bp. When the range is determined, repeat with
13 finer steps, purify and run on a BioAnalyzer DNA chip. Optimal conditions give the highest mo-
14 larity (not concentration !) between 300 and 600bp. Use the manual integration tool to obtain
15 the molarity in the 300-600bp range and in the 600-900bp range. Calculate the M600-900/
16 M300-600 ratio. The lowest value gives the best condition possible, as it indicates the conditions
17 that allow most fragments to be recovered in a 300-600bp gel cut. Purify and concentrate the
18 product in a QiaQuick column. Elute in 32 µL EB buffer.

19 **Size selection:** Prepare a BluePippin 2% DF gel cassette according to instructions. Set the ma-
20 chine on V1 (internal) marker, with a size selection in the 300bp to 650bp range.

21 - Add 10 µL Blue Pippin V1 marker solution to each sample (total 40 µL).

22 - Run the gel (about 90 minutes).

23 - Run the BluePippin according to instruction sheet.

24 - Take out the ~40 µL eluate into a LoBind tube.

25 - Replace by 40 µL EB buffer. Let stand for 5 min.

1 - Add these 40 μL to the eluate.

2 - Add EB buffer qsp. 100 μL .

3 STAGE 3 : Fragment preparation

4 **Initial bead purification:** The beads we add at this stage will remain in the reaction tube until the
5 end of the library preparation. Make sure they are at room temperature and perfectly
6 homogeneous.

7 - Add 80 μL AmpureXP beads.

8 - Mix for 10 min on a vortex or thermomixer at maximum speed.

9 - Place on the magnetic rack. Let stand for 5 min.

10 - Remove the 180 μL supernatant and discard it.

11 - Add 500 μL of fresh 80% ethanol and leave for 2 min.

12 - Remove and replace by another 500 μL 80% ethanol, leave for 2 min.

13 - Remove ca. 470 μL (leave a small volume at the bottom of the tube).

14 - Spin the beads down. Replace on the magnet.

15 - Remove the remaining ethanol, dry open until bead cracking point.

16 - Resuspend the beads in 20 μL EB buffer.

17 **End repair:** This step converts the overhangs created by sonication into phosphorylated blunt
18 ends. We follow exactly the NEB blunting kit protocol. The reaction is performed directly in the
19 previous tube, with the beads in, but off the magnet. To each library, add :

20 - 2.5 μL 10X NEB blunting buffer

21 - 2.5 μL dNTP mix (1 mM)

22 - 1 μL Blunt enzyme mix.

23 - Incubate at room temperature for 30min.

Annexes - §186

- 1 - Add k*26 μ L binding solution.
- 2 - Mix for 10 min.
- 3 - Clean (see above) and re-suspend in 43 μ L EB buffer.
- 4 **A-tailing:** This step adds a single 3' A-overhang on virtually every blunt end. This makes them
- 5 compatible with the T-overhang on the P2 adapter. The reaction is performed directly in the pre-
- 6 vious tube, with the beads in, but off the magnet. To each library, add :
 - 7 - 5 μ L 10X NEB2 buffer
 - 8 - 1 μ L dATP (10 mM)
 - 9 - 3 μ L Klenow exo-
 - 10 - 37°C for 30min with light mixing (300rpm)
 - 11 - Cool down to RT.
 - 12 - Add k*52 μ L binding solution.
 - 13 - Mix for 10 min.
 - 14 - Clean (see above) and re-suspend in 45 μ L EB buffer.
- 15 **P2 ligation:** This step adds the second adapter to every single A-tailed fragment available. For low
- 16 plexity runs, non-barcoded adapters (P2-Ø) are fine. For higher plexities with combinatorial bar-
- 17 coding, use different P2 adapters (one per library pool). The reaction is performed directly in the
- 18 previous tube, with the beads in, but off the magnet. To each library :
 - 19 - 5.2 μ L 10X NEB T4 ligase buffer
 - 20 - 1 μ L P2 adapters (10 μ M), mix thoroughly
 - 21 - 0.5 μ L T4 ligase (2,000,000 U/mL)
 - 22 - Incubate 45min at room temperature.
 - 23 - Add k*52 μ L binding solution.
 - 24 - Mix for 10 min.
 - 25 - Clean (see above)and elute in 52 μ L EB buffer.

1 - Transfer to a new LoBind tube.

2 Mark this tube as « Template ». This is the raw library, it will need double-strand adapter conver-
3 sion and enrichment before sending out for sequencing.

4 **Library validation**

5 **Amplification test:** A first non-quantitative PCR is done in order to check for correct library
6 preparation. A successful library should amplify without problem (although the strength will vary
7 from one library to another). For each library pool, prepare a reaction with :

8 - 12.5 µL NEB Phusion polymerase premix

9 - 7.5 µL H₂O

10 - 1 µL Solexa primer mix (10 µM)

11 - 4 µL library template.

12 - Solexa PCR conditions : 98°C 30" ; [98°C 10" | 65°C 30" | 72°C 30"]X18 ; 72°C 5'

13 - Add 5 µL 6X orange loading dye,

14 - Load the whole product on a gel with Low Range DNA ladder.

15 A successful library should yield a bright, well-defined band in the 400-700bp range. If amplifi-
16 cation is poor or absent, something probably went wrong during the prep. Good luck with that.

17 If amplification is really good, you may want to test fewer cycles in order to reduce duplication
18 biases to a minimum.

19 **Qubit test:** Use 2 µL of each library to measure accurately the dsDNA concentration on the
20 Qubit. A good library with standard DNA loss should be anywhere between 6 and 10 ng.µL⁻¹. A
21 lower concentration may indicate either a sub-optimal sonication (hence excessive DNA loss dur-
22 ing size selection) or a generally inefficient purification steps.

23 **Library enrichment:** If everything looks good, perform a large-volume PCR with :

Annexes - §186

- 1 - 62.5 μL NEB Phusion polymerase premix
 - 2 - 37.5 μL H_2O
 - 3 - 5 μL Solexa primer mix (10 μM)
 - 4 - 20 μL library template
 - 5 - Aliquot into 5*25 μL reactions
 - 6 - Amplify
 - 7 - Pool together the aliquots and purify into a final 30 μL EB buffer.
- 8 Now, the library should consist mostly double-stranded, properly ligated fragments ready for se-
- 9 quencing. Though this must be checked.
- 10 **BioAnalyzer check:** The final library should be run together with its template on a BioAnalyzer
- 11 DNA chip. In order to be in roughly comparable conditions, 2 μL of template should be diluted
- 12 in 11.5 μL EB buffer (as this is the dilution state in the PCR reaction prior to amplification). On
- 13 the BioAnalyzer, the library should be at roughly twice as concentrated as the template, and
- 14 exhibit a clean distribution with no obvious spikes along the amplified section. Spikes indicate
- 15 clonal populations of highly duplicated fragments. One should also check there is no adapter
- 16 carryover, especially P1 dimers (which should have all been left out since size selection).
- 17 **qPCR check:** qPCR assay (Kappa dedicated SybrGreen Illumina kit) allows for the most exact
- 18 measurement of ligated fragment molarity. This is the final check before loading on the cBot.

References

A

1. Abbott CL, Double MC, Trueman JWH, Robinson A, Cockburn A (2005) An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche* albatrosses. *Molecular ecology*, 14, 3605–3613.
2. Abram NJ, Gagan MK, Cole JE, Hantoro WS, Mudelsee M (2008) Recent intensification of tropical climate variability in the Indian Ocean. *Nature Geoscience*, 1, 849–853.
3. Abram NJ, Mulvaney R, Vimeux F, Phipps SJ, Turner J, England MH (2014) Evolution of the Southern Annular Mode during the past millennium. *Nature Climate Change*, 4, 564–569.
4. Ackerley D, Highwood EJ, Frame DJ (2009) Quantifying the effects of perturbing the physics of an interactive sulfur scheme using an ensemble of GCMs on the climateprediction.net platform. *J. Geophys. Res.*, 114,
5. Acosta CA (1999) Benthic dispersal of Caribbean spiny lobsters among insular habitats: implications for the conservation of exploited marine species. *Conservation Biology*, 13, 603–612.
6. Adams NJ, Klages NT (1987) Seasonal variation in the diet of the king penguin (*Aptenodytes patagonicus*) at sub Antarctic Marion Island. *Journal of Zoology*, 212, 303–324.
7. Aguiar D, Istrail S (2012) HapCompass: a fast cycle basis algorithm for accurate haplotype assembly of sequence data. *Journal of Computational Biology*, 19, 577–590.
8. Ahmed S, Hart T, Dawson DA, Horsburgh GJ, Trathan PN, Rogers AD (2009) Isolation and characterization of macaroni penguin (*Eudyptes chrysolophus*) microsatellite loci and their utility in other penguin species (Spheniscidae, AVES). *Mol Ecol Resour*, 9, 1530–1535.
9. Ainley D, Russell J, Jenouvrier S et al. (2010a) Antarctic penguin response to habitat change as Earth's troposphere reaches 2° C above preindustrial levels. *Ecological Monographs*, 80, 49–66.
10. Ainley D, Ballard G, Ackley S et al. (2007) Paradigm lost, or is top-down forcing no longer significant in the Antarctic marine ecosystem. *Antarctic Science*, 19, 283.
11. Ainley D, Ballard G, Blight LK et al. (2010b) Impacts of cetaceans on the structure of Southern Ocean food webs. *Marine Mammal Science*, 26, 482–498.
12. Ainley DG, Blight LK (2009) Ecological repercussions of historical fish extraction from the Southern Ocean. *Fish and Fisheries*,
13. Akst EP, Boersma PD, Fleischer RC (2002) A comparison of genetic diversity between the Galápagos penguin and the Magellanic penguin. *Conservation Genetics*, 3, 375–383.
14. Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11, 697–709.
15. Amores A, Catchen J, Nanda I et al. (2014) A RAD-tag genetic map for the platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish. *Genetics*, 197, 625–641.
16. Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an

- outgroup for the teleost genome duplication. *Genetics*, 188, 799–808.
17. Amos W, Harwood J (1998) Factors affecting levels of genetic diversity in natural populations. *Philos Trans R Soc Lond B Biol Sci*, 353, 177–186.
 18. Ancel A, Beaulieu M, Gilbert C (2013) The different breeding strategies of penguins: a review. *Comptes rendus biologies*, 336, 1–12.
 19. Ancel A, Beaulieu M, Le Maho Y, Gilbert C (2009) Emperor penguin mates: keeping together in the crowd. *Proc Biol Sci*, 276, 2163–2169.
 20. Ancel A, Cristofari R, Fretwell PT et al. (2014) Emperors in hiding: when ice-breakers and satellites complement each other in Antarctic exploration. *PLoS One*, 9, e100404.
 21. Ancel A, Rodary D, Wienecke B, Le Maho Y, Gilbert C (1999) Exploration biologique de la polynie de Mertz. *Rapport d'Activité de l'IFRTP*, 40–44.
 22. Ancel A, Kooyman GL, Ponganis PJ et al. (1992) Foraging behaviour of emperor penguins as a resource detector in winter and summer. *Nature*, 360, 336–339.
 23. Anderson JB, Shipp SS, Lowe AL, Wellner JS, Mosola AB (2002) The Antarctic Ice Sheet during the Last Glacial Maximum and its subsequent retreat history: a review. *Quaternary Science Reviews*, 21, 49–70.
 24. Anderson MG, Rhymer JM, Rohwer FC (1992) Philopatry, dispersal, and the genetic structure of waterfowl populations. *Ecology and management of breeding waterfowl*. University of Minnesota Press, Minneapolis, MN, 365–395.
 25. Anderson RF, Ali S, Bradtmiller LI et al. (2009) Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO₂. *science*, 323, 1443–1448.
 26. Arango CP, Soler-Membrives A, Miller KJ (2011) Genetic differentiation in the circum-Antarctic sea spider *Nymphon australe*. *Deep Sea Research Part II: Topical Studies in Oceanography*, 58, 212–219.
 27. Araya Modinger B, Duffy DC (1987) Animal introduction to Isla Chanaral, Chile; their history and effect on seabirds. *Cormorant*, 15, 3–6.
 28. Arenas M (2013) The importance and application of the ancestral recombination graph. *Front Genet*, 4, 206.
 29. Arenas M, Ray N, Currat M, Excoffier L (2012) Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution*, 29, 207–218.
 30. Armand LK, Leventer A (2010) Palaeo sea ice distribution and reconstruction derived from the geological records. In: *Sea ice*, pp. 469–530.
 31. Armitage KB (1981) Sociality as a life-history tactic of ground squirrels. *Oecologia*, 48, 36–49.
 32. Arnaud CM, Dobson FS, Murie JO (2012) Philopatry and within-colony movements in Columbian ground squirrels. *Molecular Ecology*, 21, 493–504.
 33. Arrigo K, Worthen DL, Dixon PL, Lizotte MP (1998) Primary productivity of near surface communities within Antarctic pack ice. *Antarctic Sea Ice: Biological Processes, Interactions and Variability*, 23–43.
 34. Arrigo KR, Worthen DL, Lizotte MP, Dixon P, Dieckmann G (1997) Primary production in Antarctic sea ice. *Science*, 276, 394–397.
 35. Ashok K, Guan Z, Saji NH, Yamagata T (2004) Individual and combined influences of ENSO and the Indian Ocean dipole on the Indian summer monsoon. *Journal of Climate*, 17, 3141–3155.

36. Augustin L, Barbante C, Barnes PR et al. (2004) Eight glacial cycles from an Antarctic ice core. *Nature*, 429, 623–628.
37. Auld JR, Agrawal AA, Relyea RA (2009) Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20091355.
38. Avise JC, Nelson WS, Bowen BW, Walker D (2000) Phylogeography of colonially nesting seabirds, with special reference to global matrilineal patterns in the sooty tern (*Sterna fuscata*). *Molecular Ecology*, 9, 1783–1792.
39. Axelsson E, Webster MT, Smith NGC, Burt DW, Ellegren H (2005) Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Research*, 15, 120–125.
- B**
40. Baird NA, Etter PD, Atwood TS et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, 3, e3376.
41. Baker AJ, Tavares ES, Elbourne RF (2009) Countering criticisms of single mitochondrial DNA gene barcoding in birds. *Molecular Ecology Resources*, 9, 257–268.
42. Baker AJ, Pereira SL, Haddrath OP, Edge K-A (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proceedings of the Royal Society B: Biological Sciences*, 273, 11–17.
43. Baker CS, Slade RW, Bannister JL et al. (2008) Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, world-wide. *Molecular Ecology*, 3, 313–327.
44. Bakun A (2006) Wasp-waist populations and marine ecosystem dynamics: navigating the predator pit topographies. *Progress in Oceanography*, 68, 271–288.
45. Ballard G, Toniolo V, Ainley DG, Parkinson CL, Arrigo KR, Trathan PN (2010) Responding to climate change: Adélie penguins confront astronomical and ocean boundaries. *Ecology*, 91, 2056–2069.
46. Ballou JD (1997) Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *Journal of Heredity*, 88, 169–178.
47. Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Mol Ecol*, 13, 3021–3031.
48. Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular ecology*, 11, 155–165.
49. Banks J, Van Buren A, Cherel Y, Whitfield JB (2006) Genetic evidence for three species of rockhopper penguins, *Eudyptes chrysocome*. *Polar Biology*, 30, 61–67.
50. Banks SC, Peakall ROD (2012) Genetic spatial autocorrelation can readily detect sex-biased dispersal. *Molecular ecology*, 21, 2092–2105.
51. Barbante C, Barnola J-M, Becagli S et al. (2006) One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature*, 444, 195–198.
52. Barber-Meyer SM, Kooyman GL, Ponganis PJ (2007) Estimating the relative abundance of emperor penguins at inaccessible colonies using satellite imagery. *Polar Biology*, 30, 1565–1570.
53. Barbosa A, Palacios MJ (2009) Health of Antarctic birds: a review of their parasites, pathogens and diseases. *Polar Biology*, 32, 1095–1115.

54. Barbraud C, Gavriilo M, Mizin Y, Weimerskirch H (2011) Comparison of emperor penguin declines between Pointe Géologie and Haswell Island over the past 50 years. *Ant Sci*, 23, 461–468.
55. Barbraud C, Jouventin P (1998) What causes body size variation in the Snow Petrel *Pagodroma nivea*. *Journal of Avian Biology*, 29, 161–171.
56. Barbraud C, Weimerskirch H (2001a) Emperor penguins and climate change. *Nature*, 411, 183–186.
57. Barbraud C, Weimerskirch H (2001b) Contrasting effects of the extent of sea ice on the breeding performance of an Antarctic top predator, the Snow Petrel *Pagodroma nivea*. *Journal of Avian Biology*, 32, 297–302.
58. Barker JSFS (2008) Adaptation and Fitness in Animal Populations: Evolutionary and Breeding Perspectives on Genetic Resource Management. *Adaptation and Fitness in Animal Populations*, 3–14.
59. Barlow EJ, Daunt F, Wanless S, Reid JM (2013) Estimating dispersal distributions at multiple scales: within-colony and among-colony dispersal rates, distances and directions in European Shags. *Ibis*, 155, 762–778.
60. Barnosky AD, Hadly EA, Bascompte J et al. (2012) Approaching a state shift in Earth's biosphere. *Nature*, 486, 52–58.
61. Baroni C, Orombelli G (1994) Abandoned penguin rookeries as Holocene paleoclimatic indicators in Antarctica. *Geology*, 22, 23–26.
62. Barrat A (1976) Quelques aspects de la biologie et de l'écologie du manchot royal *Aptenodytes patagonicus* des îles Crozet. *Com. Nat. Fr. Rech. Ant*, 40, 9–51.
63. Barron JA (1996) Diatom constraints on the position of the Antarctic Polar Front in the middle part of the Pliocene. *Marine micropaleontology*, 27, 195–213.
64. Bathmann UV, Scharek R, Klaas C, Dubischar CD, Smetacek V (1997) Spring development of phytoplankton biomass and composition in major water masses of the Atlantic sector of the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 44, 51–67.
65. Bauer S, Gienapp P, Madsen J (2008) The relevance of environmental conditions for departure decision changes en route in migrating geese. *Ecology*, 89, 1953–1960.
66. Beaufort L, de Garidel-Thoron T, Mix AC, Pisiac NG (2001) ENSO-like forcing on oceanic primary production during the late Pleistocene. *Science*, 293, 2440–2444.
67. Beaugrand G, Edwards M, Raybaud V, Goberville E, Kirby RR (2015) Future vulnerability of marine biodiversity compared with contemporary and past changes. *Nature Clim. Change*, 5, 695–701.
68. Beaugrand G, Reid PC, Ibanez F, Lindley JA, Edwards M (2002) Reorganization of North Atlantic marine copepod biodiversity and climate. *Science*, 296, 1692–1694.
69. Becker BH, Beissinger SR (2006) Centennial decline in the trophic level of an endangered seabird after fisheries decline. *Conservation Biology*, 20, 470–479.
70. Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, 185, 313–326.
71. Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, 13, 827–836.
72. Beerli P (2006) Comparison of Bayesian and maximum-likelihood infer-

- ence of population genetic parameters. *Bioinformatics*, 22, 341–345.
73. Behrenfeld MJ, O Malley RT, Siegel DA et al. (2006) Climate-driven trends in contemporary ocean productivity. *Nature*, 444, 752–755.
74. Behrenfeld MJ, Randerson JT, McClain CR et al. (2001) Biospheric primary production during an ENSO transition. *Science*, 291, 2594–2597.
75. Bellard C, Bertelsmeier C, Leadley P, Thuiller... W (2012) Impacts of climate change on the future of biodiversity. *Ecology ...*,
76. Berger A (2013) Milankovitch and climate: understanding the response to astronomical forcing. *Nato ASI Series*, 126,
77. Bergson H (1911) Life and consciousness.
78. Berry PM, Dawson TP, Harrison PA, Pearson RG (2002) Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global ecology and biogeography*, 11, 453–462.
79. Berthold P (1991) Genetic control of migratory behaviour in birds. *Trends in Ecology & Evolution*, 6, 254–257.
80. Bertrand R, Lenoir J, Piedallu C et al. (2011) Changes in plant community composition lag behind climate warming in lowland forests. *Nature*, 479, 517–520.
81. Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting FST: the impact of rare variants. *Genome Res*, 23, 1514–1521.
82. Bicknell AWJ, Knight ME, Bilton D, Reid JB, Burke T, Votier SC (2012) Population genetic structure and long-distance dispersal among seabird populations: Implications for colony persistence. *Molecular Ecology*, 21, 2863–2876.
83. Billiard S, Lenormand T (2005) Evolution of migration under kin selection and local adaptation. *Evolution*, 59, 13–23.
84. Billing TM, Guay PJ, Peucker AJ, Mulder RA (2007) Isolation and characterization of polymorphic microsatellite loci for the study of paternity and population structure in the little penguin *Eudyptula minor*. *Molecular Ecology Notes*, 7, 425–427.
85. Bintanja R, van Oldenborgh GJ, Katman CA (2015) The effect of increased fresh water from Antarctic ice shelves on future trends in Antarctic sea ice. *Annals of Glaciology*, 56, 69.
86. Bivand RS, Hauke, J., and Kossowski, T. (2013) Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, 45, 150–179.
87. Blais JM, Kimpe LE, McMahon D et al. (2005) Arctic seabirds transport marine-derived contaminants. *Science*, 309, 445–445.
88. Blois JL, Zarnetske PL, Fitzpatrick MC, Finnegan S (2013) Climate change and the past, present, and future of biotic interactions. *Science*, 341, 499–504.
89. Boessenkool S, Austin JJ, Worthy TH et al. (2009a) Relict or colonizer? Extinction and range expansion of penguins in southern New Zealand. *Proceedings of the Royal Society B: Biological Sciences*, 276, 815–821.
90. Boessenkool S, Star B, Waters JM, Seddon PJ (2009b) Multilocus assignment analyses reveal multiple units and rare migration events in the recently expanded yellow-eyed penguin (*Megadyptes antipodes*). *Molecular ecology*, 18, 2390–2400.
91. Bolnick DI, Amarasekare P, Araújo MS et al. (2011) Why intraspecific trait variation matters in community ecology. *Trends in ecology & evolution*, 26, 183–192.

92. Bonin CA, Goebel ME, Forcada J, Burton RS, Hoffman JI (2013) Unexpected genetic differentiation between recently recolonized populations of a long-lived and highly vagile marine mammal. *Ecol Evol*, 3, 3701–3712.
93. Borboroglu PG, Boersma PD (2013) Penguins: natural history and conservation.
94. Bost C-A, Delord K, Barbraud C et al. (2013) King Penguin. In: *Penguins, Natural history and conservation*, (eds Borboroglu PG, Boersma D), pp. 7–21. University of Washington Press, Seattle & London.
95. Bost CA, Cotté C, Terray P et al. (2015) Large-scale climatic anomalies affect marine predator foraging behaviour and demography. *Nat Commun*, 6, 8220.
96. Bost C-A, Cotté C, Bailleul F et al. (2009) The importance of oceanographic fronts to marine birds and mammals of the southern oceans. *Journal of Marine Systems*, 78, 363–376.
97. Both C, Bouwhuis S, Lessells CM, Visser ME (2006) Climate change and population declines in a long-distance migratory bird. *Nature*, 441, 81–83.
98. Bouckaert R, Heled J, Kuhnert D et al. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*, 10, e1003537.
99. Bouzat JL, Walker BG, Boersma PD (2009) Regional genetic structure in the Magellanic Penguin (*Spheniscus magellanicus*) suggests metapopulation dynamics. *The Auk*, 126, 326–334.
100. Bowen WD, Siniff DB (1999) Distribution, population biology, and feeding ecology of marine mammals. *Biology of marine mammals*, 423–484.
101. Bowler DE, Benton TG (2005) Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews*, 80, 205–225.
102. Box EEO (1981) Macroclimate and plant forms: an introduction to predictive modeling in phytogeography.
103. Boyd IL, Arnborn T (1991) Diving behaviour in relation to water temperature in the southern elephant seal: foraging implications. *Polar Biology*, 11, 259–266.
104. Braconnot P, Harrison SP, Kageyama M et al. (2012) Evaluation of climate models using palaeoclimatic data. *Nature Climate Change*, 2, 417–424.
105. Braconnot P, Harrison SP, Otto-Bliesner B, Abe-Ouchi A, Jungclaus J, Peterschmitt J-Y (2011) The paleoclimate modeling intercomparison project contribution to CMIP5. *CliVAR Exchanges*, 56, 2.
106. Bradshaw WE, Holzapfel CM (2006) Evolutionary response to rapid climate change. *Science(Washington)*, 312, 1477–1478.
107. Braudel F (1949) La Méditerranée et le monde méditerranéen à l 'époque de Philippe II.
108. Braudel F (1958) La Longue durée. *Histoire et Sciences Sociales*, 13, 725–753.
109. Braudel F (1969) Ecrits sur l 'histoire.
110. Bried J, Jiguet F, Jouventin P (1999) Why do Aptenodytes penguins have high divorce rates? *The Auk*, 504–512.
111. Bried J, Jouventin P (2001) The King Penguin *Aptenodytes patagonicus*, a non-nesting bird which selects its breeding habitat. *Ibis*, 143, 670–673.
112. Bried J, Nicolaus M, Jarne P, Dubois MP, Jouventin P (2007) Population biology of the wandering albatross (*Diomedea exulans*) in the Crozet and Kerguelen archipelagos, southern Indian Ocean, approached

- through genetic and demographic methods. *Journal of Zoology*, 272, 20–29.
113. Brierley AS, Fernandes PG, Brandon MA et al. (2002) Antarctic krill under sea ice: elevated abundance in a narrow band just south of ice edge. *Science*, 295, 1890–1892.
114. Brodin A, Olsson O, Clark CW (1998) Modeling the breeding cycle of long-lived birds: Why do king penguins try to breed late? *The Auk*, 115, 767–771.
115. Bromwich DH, Nicolas JP, Monaghan AJ et al. (2013) Central West Antarctica among the most rapidly warming regions on Earth. *Nature Geoscience*, 6, 139–145.
116. Browning SR, Browning BL (2011) Haplotype phasing: existing methods and new developments. *Nature Reviews Genetics*, 12, 703–714.
117. Bryant D, Bouckaert R, Felsenstein... J (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular biology and ...*,
118. Budd GM (1961) The biotopes of emperor penguin rookeries. *Emu*, 61, 171–189.
119. Budd GM (1962) Population studies in rookeries of the Emperor Penguin *Aptenodytes forsteri*. Proceedings of the Zoological Society of London 139, 365–388.
120. Budd GM (2000) Changes in Heard Island glaciers, king penguins and fur seals since 1947. Papers and Proceedings of the Royal Society of Tasmania 133, 47–60.
121. Burckle LH (1984) Diatom distribution and paleoceanographic reconstruction in the Southern Ocean Present and Last Glacial Maximum. *Marine Micropaleontology*, 9, 241–261.
122. Burg TM, Croxall JP (2004) Global population structure and taxonomy of the wandering albatross species complex. *Molecular Ecology*, 13, 2345–2355.
123. BurrIDGE CP, Peucker AJ, Valautham SK, Styan CA, Dann P (2015) Nonequilibrium Conditions Explain Spatial Variability in Genetic Structuring of Little Penguin (*Eudyptula minor*). *Journal of Heredity*, 106, 228–237.
124. Burrows MT, Schoeman DS, Buckley LB et al. (2011) The pace of shifting climate in marine and terrestrial ecosystems. *Science*, 334, 652–655.
125. Burrows MT, Schoeman DS, Richardson AJ et al. (2014) Geographical limits to species-range shifts are suggested by climate velocity. *Nature*, 507, 492–495.
- ## C
126. Cabré A, Marinov I, Bernardello R, Bianchi D (2015) Oxygen minimum zones in the tropical Pacific across CMIP5 models: mean state differences and climate change trends. *Biogeosciences Discussions*, 12,
127. Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, 87, 1217–1227.
128. Cairns DK (1988) Seabirds as indicators of marine food supplies. *Biological Oceanography*,
129. Calvo E, Pelejero C, De Deckker P, Logan GA (2007) Antarctic deglacial pattern in a 30 kyr record of sea surface temperature offshore South Australia. *Geophysical Research Letters*, 34,
130. Cariou M, Duret L, Charlat S (2013) Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecol Evol*, 3, 846–852.
131. Carr MH, Neigel JE, Estes JA, An- delman S, Warner RR, Largier JL (2003) Comparing marine and terrestrial ecosys-

- tems: implications for the design of coastal marine reserves. *Ecological Applications*, 13, 90–107.
132. Carr M-E, Friedrichs MAM, Schmeltz M et al. (2006) A comparison of global estimates of marine primary production from ocean color. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53, 741–770.
133. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1, 171–182.
134. Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular ecology*, 22, 3124–3140.
135. Celia B, Claire S, Viblanc Vincent A, Helene G-N, Magali B, Le Maho Yvon LBC (2014) Inter-Annual Variability of Fledgling Sex Ratio in King Penguins. *Plos One*, 9, 1–17.
136. Chapman EW, Hofmann EE, Patterson DL, Fraser WR (2010) The effects of variability in Antarctic krill (*Euphausia superba*) spawning behavior and sex/maturity stage distribution on Adélie penguin (*Pygoscelis adeliae*) chick growth: a modeling study. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57, 543–558.
137. Charles CD, Froelich PN, Zibello MA, Mortlock RA, Morley JJ (1991) Biogenic opal in Southern Ocean sediments over the last 450,000 years: Implications for surface water chemistry and circulation. *Paleoceanography*, 6, 697–728.
138. Charmantier A, McCleery RH, Cole LR, Perrins C, Kruuk LE, Sheldon BC (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, 320, 800–803.
139. Charrassin J-B, Bost C-A (2001) Utilisation of the oceanic habitat by king penguins over the annual cycle. *Marine Ecology Progress Series*, 221, 285–298.
140. Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. *Science*, 333, 1024–1026.
141. Cherel Y, Pütz K, Hobson KA (2002) Summer diet of king penguins (*Aptenodytes patagonicus*) at the Falkland Islands, southern Atlantic Ocean. *Polar Biology*, 25, 898–906.
142. Chereskin TK, Donohue KA, Watts DR, Tracey KL, Firing YL, Cutting AL (2012) cDrake: Dynamics and transport of the Antarctic Circumpolar Current in Drake Passage. *Oceanography*, 25, 134–135.
143. Cherry-Garrard A (1922) The worst journey in the world.
144. Chevin LM, Decorzent G, Lenormand T (2014) Niche dimensionality and the genetics of ecological speciation. *Evolution*, 68, 1244–1256.
145. Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, 186, 983–995.
146. Chown SL, Clarke A, Fraser CI, Cary SC, Moon KL, McGeoch MA (2015) The changing form of Antarctic biodiversity. *Nature*, 522, 431–438.
147. Ciasto LM, Thompson DWJ (2008) Observations of large-scale ocean-atmosphere interaction in the Southern Hemisphere. *Journal of Climate*, 21, 1244–1259.
148. Clarke AL, Sæther BE, Røskaft E (1997) Sex biases in avian dispersal: a reappraisal. *Oikos*, 429–438.

149. Clarke JA, Ksepka DT, Stucchi M et al. (2007) Paleogene equatorial penguins challenge the proposed relationship between biogeography, diversity, and Cenozoic climate change. *Proceedings of the National Academy of Sciences*, 104, 11545–11550.
150. CLIMAP (1981) Seasonal Reconstructions of the Earth's surface at the Last Glacial Maximum.
151. Collins MA, Xavier JC, Johnston NM et al. (2008) Patterns in the distribution of myctophid fish in the northern Scotia Sea ecosystem. *Polar Biology*, 31, 837–851.
152. Collins M, An S-I, Cai W et al. (2010) The impact of global warming on the tropical Pacific Ocean and El Niño. *Nature Geoscience*, 3, 391–397.
153. Comte L, Murienne J, Grenouillet G (2014) Species traits and phylogenetic conservatism of climate-induced range shifts in stream fishes. *Nat Commun*, 5, 5023.
154. Convey P, Barnes DKA, Griffiths HJ, Grant SM, Linse K, Thomas DN (2012) Biogeography and regional classifications of Antarctica. *Antarctic Ecosystems: An extreme environment in a changing world*, 469–491.
155. Convey P, Morton A, Poncet J (1999) Survey of marine birds and mammals of the South Sandwich Islands. *Polar Record*, 35, 107–124.
156. Convey P, Stevens MI, Hodgson DA et al. (2009) Exploring biological constraints on the glacial history of Antarctica. *Quaternary Science Reviews*, 28, 3035–3048.
157. Cooper CB, Daniels SJ, Walters JR (2008) Can we improve estimates of juvenile dispersal distance and survival. *Ecology*, 89, 3349–3361.
158. Coria NR, Montalti D (2000) A newly discovered breeding colony of emperor penguins *Aptenodytes forsteri*. *Marine Ornithology*, 28, 119–120.
159. Cornet C (2014) The adaptive capacities of seabirds to face environmental variability: the role of heterogeneity within populations. PhD.
160. Cortés Labra M (2007) Ajustes fisiológicos para la nidificación en cuevas: el caso de *Pelecanoides garnotii*, un enfoque evolutivo. Magister.
161. Coté SD (2000) Aggressiveness in king penguins in relation to reproductive status and territory location. *Anim Behav*, 59, 813–821.
162. Coulson JC (2002) Colonial breeding in seabirds. In 'Biology of Marine Birds.(Eds EA Schreiber and J. Burger.) pp. 87–113.
163. Cowie RH, Holland BS (2006) Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *Journal of Biogeography*, 33, 193–198.
164. Crawford TJ (1984) The estimation on neighbourhood parameters for plant populations. *Heredity*, 52, 273–283.
165. Cressey D (2015) World's whaling slaughter tallied. *Nature*, 519, 140–141.
166. Croxall JP, Butchart SHM, Lascelles B et al. (2012) Seabird conservation status, threats and priority actions: a global assessment. *Bird Conservation International*, 22, 1–34.
167. Croxall JP, Silk JRD, Phillips RA, Afanasyev V, Briggs DR (2005) Global circumnavigations: tracking year-round ranges of nonbreeding albatrosses. *Science*, 307, 249–250.
168. Croxall JP, Trathan PN, Murphy EJ (2002) Environmental change and Antarctic seabird populations. *Science*, 297, 1510.
169. Cruaud A, Gautier M, Galan M et al. (2014) Empirical assessment of RAD sequencing for interspecific phylogeny. *Mol Biol Evol*, 31, 1272–1274.

170. Cury PM, Boyd IL, Bonhommeau S et al. (2011) Global seabird response to forage fish depletion—one-third for the birds. *Science*, 334, 1703–1706.

D

171. Damerou M, Matschiner M, Salzburger W, Hanel R (2012) Comparative population genetics of seven notothenioid fish species reveals high levels of gene flow along ocean currents in the southern Scotia Arc, Antarctica. *Polar biology*, 35, 1073–1086.

172. Danchin E, Wagner RH (1997) The evolution of coloniality: the emergence of new perspectives. *Trends in Ecology & Evolution*, 12, 342–347.

173. Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

174. Darwin C (1859) On the origin of species.

175. Davey JW, Blaxter ML (2010) RAD-Seq: next-generation population genetics. *Briefings in Functional Genomics*, 9, 416.

176. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510.

177. Davis AJ, Jenkinson LS, Lawton JH, Shorrocks B, Wood S (1998) Making mistakes when predicting shifts in species range in response to global warming. *Nature*, 391, 783–786.

178. Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, 292, 673–679.

179. De Boer AM, Toggweiler JR, Sigman DM (2008) Atlantic dominance of the meridional overturning circulation. *Journal of Physical Oceanography*, 38, 435–450.

180. de Bruyn M, Hall BL, Chauke LF, Baroni C, Koch PL, Hoelzel AR (2009) Rapid response of a marine mammal species to holocene climate and habitat change. *PLoS Genet*, 5, e1000554.

181. De Dinechin M, Dobson S, Zehntindjiev P et al. (2012) The biogeography of Gentoo Penguins. *Can. J. Zool*, 90, 353–361.

182. De Dinechin M, Ottvall R, Quillfeldt P, Jouventin P (2008) Speciation chronology of rockhopper penguins inferred from molecular, geological and palaeoceanographic data. *Journal of Biogeography*, 36, 693–702.

183. de Lavergne C, Palter JB, Galbraith ED, Bernardello R, Marinov I (2014) Cessation of deep convection in the open Southern Ocean under anthropogenic climate change. *Nature Climate Change*, 4, 278–282.

184. Dearborn DC, Anders AD, Schreiber EA, Adams RMM, Mueller UG (2003) Inter-island movements and population differentiation in a pelagic seabird. *Molecular Ecology*, 12, 2835–2843.

185. Delord K, Barbraud C, Weimerskirch H (2004) Long-term trends in the population size of king penguins at Crozet archipelago: environmental variability and density dependence? *Polar Biology*, 27, 793–800.

186. Delord K, Roudaut G, Guinet C, Barbraud C, Bertrand S, Weimerskirch H (2015) Kite aerial photography: a low cost method for monitoring seabird colonies. *Journal of Field Ornithology*, 86, 173–179.

187. Deng J, Shoemaker R, Xie B et al. (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature biotechnology*, 27, 353–360.

188. Denton GH, Hughes TJ (2002) Reconstructing the Antarctic ice sheet at the Last Glacial Maximum. *Quaternary Science Reviews*, 21, 193–202.

189. DePristo MA, Banks E, Poplin R et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, 43, 491–498.
190. Descamps S, Gauthier-Clerc M, Gendner JP, Le Maho Y (2002) The annual breeding cycle of unbanded king penguins *Aptenodytes patagonicus* on Possession Island (Crozet). *Avian Science*, 2, 87–98.
191. Descamps S, Le Bohec C, Le Maho Y, Gendner JP, Gauthier-Clerc M (2009) Relating demographic performance to breeding-site location in the King Penguin. *The Condor*, 111, 81–87.
192. Díaz S, Fargione J, Chapin FS, Tilman D (2006) Biodiversity loss threatens human well-being. *PLoS biology*, 4, 1300–1305.
193. Dickinson MG, Orme CD, Suttle KB, Mace GM (2014) Separating sensitivity from exposure in assessing extinction risk from climate change. *Sci Rep*, 4, 6898.
194. Diffenbaugh NS, Field CB (2013) Changes in ecologically critical terrestrial climate conditions. *Science*, 341, 486–492.
195. Dingle H, Drake VA (2007) What is migration. *Bioscience*, 57, 113–121.
196. Doak DF, Morris WF (2010) Demographic compensation and tipping points in climate-induced range shifts. *Nature*, 467, 959–962.
197. Dong S, Sprintall J, Gille ST (2006) Location of the Antarctic polar front from AMSR-E satellite sea surface temperature measurements. *Journal of Physical Oceanography*, 36, 2075–2089.
198. Double MC, Peakall R, Beck NR, Cockburn A (2005) Dispersal, philopatry, and infidelity: dissecting local genetic structure in superb fairy-wrens (*Malurus cyaneus*). *Evolution*, 59, 625–635.
199. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22, 1185.
200. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7, 214.
201. Du Plessis CJ, Van Heezik YM, Seddon PJ (1994) Timing of king penguin breeding at Marion Island. *Emu*, 94, 216–219.
202. Ducklow HW, Baker K, Martinson DG et al. (2007) Marine pelagic ecosystems: the west Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 67–94.
203. Duffy DC, Siegfried WR (1987) Historical variations in food consumption by breeding seabirds of the Humboldt and Benguela upwelling regions. *Seabirds: feeding ecology and role in marine ecosystems*, 327–346.
204. Dugger KM, Ainley DG, Lyver POB, Barton K, Ballard G (2010) Survival differences and the effect of environmental instability on breeding dispersal in an Adélie penguin meta-population. *Proceedings of the National Academy of Sciences*, 107, 12375–12380.
205. Dugger KM, Ballard G, Ainley DG, Barton KJ (2006) Effects of flipper bands on foraging behavior and survival of adélie penguins (*Pygoscelis adeliae*). *The Auk*, 123, 858–869.
206. Dunford C (1977) Kin selection for ground squirrel alarm calls. *The American Naturalist*, 111, 782–785.
207. Dunn CW (2005) Complex colony level organization of the deep sea siphonophore *Bargmannia elongata* (Cnidaria, Hydrozoa) is directionally asymmetric and arises by the subdivision of pro

- buds. *Developmental dynamics*, 234, 835–845.
208. Dunn PO, Winkler DW, Møller AP, Fiedler W, Berthold P (2010) Effects of climate change on timing of breeding and reproductive success in birds. *Effects of climate change on birds*, 113–128.
209. Dunn PO, Winkler DW, Whittingham LA, Hannon SJ, Robertson RJ (2011) A test of the mismatch hypothesis: How is timing of reproduction related to food abundance in an aerial insectivore. *Ecology*, 92, 450–461.
210. Durner GM, Douglas DC, Nielson RM et al. (2009) Predicting 21st-century polar bear habitat distribution from global climate models. *Ecological Monographs*, 79, 25–58.
- ## E
211. Earl DA (2012) Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation genetics resources*, 4, 359–361.
212. Edwards AWF (2008) GH Hardy (1908) and Hardy Weinberg Equilibrium. *Genetics*, 179, 1143–1150.
213. Ehrlén J, Morris WF (2015) Predicting changes in the distribution and abundance of species under environmental change. *Ecol Lett*, 18, 303–314.
214. Eichhorn G, Drent RH, Stahl J, Leito A, Alerstam T (2009) Skipping the Baltic: the emergence of a dichotomy of alternative spring migration strategies in Russian barnacle geese. *Journal of Animal Ecology*, 78, 63–72.
215. Ellegren H (2013) The evolutionary genomics of birds. *Annual Review of Ecology, Evolution, and Systematics*, 44, 239–259.
216. Ellis JC (2005) Marine birds on land: a review of plant biomass, species richness, and community composition in seabird colonies. *Plant Ecology*, 181, 227–241.
217. Emerson KJ, Merz CR, Catchen JM et al. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, 107, 16196–16200.
218. Emslie SD, Coats L, Licht K (2007) A 45,000 yr record of Adélie penguins and climate change in the Ross Sea, Antarctica. *Geol*, 35, 61.
219. Epperson BK (2005) Estimating dispersal from short distance spatial autocorrelation. *Heredity*, 95, 7–15.
220. Esler D, Iverson SA, Rizzolo DJ (2006) Genetic and demographic criteria for defining population units for conservation: the value of clear messages. *The Condor*, 108, 480–483.
221. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Mol Ecol*, 14, 2611–2620.
222. Ewens WJ (1979) *Mathematical Population Genetics*.
223. Ewins PJ (1986) The ecology of black guillemots (*Cephus grylle*) in Shetland.
224. Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet*, 9, e1003905.
225. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*, 1, 47–50.
226. Excoffier L, Smouse PE, Quattro JM (1992a) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochon-

- drial DNA restriction data. *Genetics*, 131, 479–491.
227. Excoffier L, Foll M (2011) Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27, 1332–1334.
228. Excoffier L, Smouse PE, Quattro JM (1992b) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491.
- F
229. Fahrbach E, Rohardt G, Schröder M, Strass V (1994) Transport and structure of the Weddell Gyre. *Annales Geophysicae* 12, 840–855.
230. Falla RA (1937) Birds of the British, Australian, and New Zealand Antarctic Research Expedition (1929-1931) and of the Australasian Antarctic Expedition (1911-1914).
231. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
232. Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci. *Molecular biology and evolution*, 23, 691–700.
233. Fernández Zamora C (2011) Midiendo la densidad reproductiva del petrel buceador (*Pelecanoides garnotii*) antes de la erradicación de conejos europeos sobre las islas de Coquimbo, Chile. Postgrado.
234. Fernández-Chacón A, Genovart M, Pradel R et al. (2013) When to stay, when to disperse and where to go: survival and dispersal patterns in a spatially structured seabird population. *Ecography*, 26, 1–10.
235. Fernandez C, Miranda Urbina D, Portflitt Toro M, Luna Jorquera G (2014) Estimación de la densidad reproductiva del yunco *Pelecanoides garnotii* en Isla Choro. *XI Congreso Chileno de Ornitología*, 24.
236. Figueroa J, Stucchi M (2008) Peruvian diving petrels *Pelecanoides garnotii* on Lobos de afuera islands, Peru. *Marine Ornithology*, 36, 189–190.
237. Figueroa J, Suazo E, Santillan L (2011) Registros actuales e históricos del Potoyunco peruano *Pelecanoides Garnotii* en el Perú. *The Biologist*,
238. Fischer H, Wahlen M, Smith J, Mastroianni D, Deck B (1999) Ice core records of atmospheric CO₂ around the last three glacial terminations. *Science*, 283, 1712–1714.
239. Fogt RL, Bromwich DH (2006) Decadal variability of the ENSO teleconnection to the high-latitude South Pacific governed by coupling with the Southern Annular Mode. *Journal of Climate*, 19, 979–997.
240. Fogt RL, Bromwich DH, Hines KM (2011) Understanding the SAM influence on the South Pacific ENSO teleconnection. *Climate dynamics*, 36, 1555–1576.
241. Forcada J, Trathan PN (2009) Penguin responses to climate change in the Southern Ocean. *Global Change Biology*, 15, 1618–1630.
242. Fordyce RE, Jones CM, Davis LS, Darby JT (1990) Penguin history and new fossil material from New Zealand. *Penguin biology*, 419, 446.
243. Frankham R (1995) Inbreeding and extinction: a threshold effect. *Conservation Biology*, 9, 792–799.
244. Frankham R (2005) Stress and adaptation in conservation genetics. *Journal of Evolutionary Biology*, 18, 750–755.

245. Fraser CI, Nikula R, Ruzzante DE, Waters JM (2012) Poleward bound: biological impacts of Southern Hemisphere glaciation. *Trends in ecology & evolution*, 27, 462–471.
246. Fraser CI, Nikula R, Spencer HG, Waters JM (2009) Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum. *Proceedings of the National Academy of Sciences*, 106, 3249–3253.
247. Fraser CI, Nikula R, Waters JM (2011) Oceanic rafting by a coastal community. *Proceedings of the Royal Society B: Biological Sciences*, 278, 649–655.
248. Frederiksen M, Petersen A (1999) Philopatry and dispersal within a Black guillemot colony. *Waterbirds*, 22, 274–281.
249. Frederiksen M, Edwards M, Richardson AJ, Halliday NC, Wanless S (2006) From plankton to top predators: bottom up control of a marine food web across four trophic levels. *Journal of Animal Ecology*, 75, 1259–1268.
250. Frederiksen M, Harris MP, Daunt F, Rothery P, Wanless S (2004) Scale dependent climate signals drive breeding phenology of three seabird species. *Global Change Biology*, 10, 1214–1221.
251. Freer JJ, Mable BK, Clucas G et al. (2015) Limited genetic differentiation among chinstrap penguin (*Pygoscelis antarctica*) colonies in the Scotia Arc and Western Antarctic Peninsula. *Polar Biology*, 1–10.
252. Fretwell PT, Trathan PN, Wienecke B, Kooyman GL (2014) Emperor penguins breeding on iceshelves. *PLoS one*, 9, e85285.
253. Fretwell PT, Larue MA, Morin P et al. (2012) An emperor penguin population estimate: the first global, synoptic survey of a species from space. *PLoS One*, 7, e33751.
254. Fretwell PT, Trathan PN (2009) Penguins from space: faecal stains reveal the location of emperor penguin colonies. *Global Ecology and Biogeography*, 18, 543–552.
255. Friesen VL (2015) Speciation in seabirds: why are there so many species and why aren't there more. *Journal of Ornithology*, 1–13.
256. Friesen VL, Burg TM, McCoy KD (2007) Mechanisms of population differentiation in seabirds. *Molecular Ecology*, 16, 1765–1785.
257. Froget G, Gautier-Clerc M, Le Maho Y, Handrich Y (1998) Is penguin banding harmless? *Polar Biology*, 20, 409–413.
258. Frolking S, Roulet NT (2007) Holocene radiative forcing impact of northern peatland carbon accumulation and methane emissions. *Global Change Biology*, 13, 1079–1088.
259. Fumagalli M, Vieira FG, Korneliusen TS et al. (2013) Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*, 195, 979–992.
260. Fumagalli M, Vieira FG, Linderoth T, Nielsen R (2014) ngsTools: methods for population genetics analyses from Next-Generation Sequencing data. *Bioinformatics*, 1–2.
261. Furness RW (1978) Energy requirements of seabird communities: a bioenergetics model. *The Journal of Animal Ecology*, 39–53.

G

262. Gagnaire PA, Normandeau E, Pavey SA, Bernatchez L (2013a) Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, 22, 3036–3048.
263. Gagnaire P, Pavey SA, Normandeau E, Bernatchez L (2013b) The genetic archi-

- ture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by rad sequencing. *Evolution*, 67, 2483–2497.
264. Gales R, Pemberton D (1988) Recovery of the king penguin, *Aptenodytes patagonicus*, population on Heard Island. *Wildlife Research*, 15, 579–585.
265. García-Godos I, Goya E (2006) Diet of the Peruvian diving petrel *Pelecanoides garnotii* at La Vieja island, Peru, 1997–2000: potential fishery interactions and conservation implications. *Marine Ornithology*, 34, 33–41.
266. Gaspar P, Benson SR, Dutton PH et al. (2012) Oceanic dispersal of juvenile leatherback turtles: going beyond passive drift modeling. *Marine Ecology Progress Series*, 457, 265.
267. Gauthier-Clerc M, Mangin S, Le Bohec C, Gendner JP, Le Maho Y (2003) Comparison of behaviour, body mass, haematocrit level, site fidelity and survival between infested and non-infested king penguin *Aptenodytes patagonicus* by ticks *Ixodes uriae*. *Polar Biology*, 26, 379–382.
268. Gauthier-Clerc M, Gendner JP, Ribic CA et al. (2004) Long-term effects of flipper bands on penguins. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271, S423.
269. Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ (2015) Bayesian total evidence dating reveals the recent crown radiation of penguins. *preprint (arXiv)*, 1506.04797.
270. Gendner JP, Gauthier-Clerc M, Le Bohec C, Descamps S, Le Maho Y (2005) A new application for transponders in studying penguins. *Journal of Field Ornithology*, 76, 138–142.
271. Gent PR, Danabasoglu G, Donner LJ et al. (2011) The community climate system model version 4. *Journal of Climate*, 24, 4973–4991.
272. Gersonde R, Crosta X, Abelmann A, Armand L (2005) Sea-surface temperature and sea ice distribution of the Southern Ocean at the EPILOG Last Glacial Maximum a circum-Antarctic view based on siliceous microfossil records. *Quaternary Science Reviews*, 24, 869–896.
273. Gibb GC, Hills SFK (2013) Intergenerational mutation rate does not equal long-term evolutionary substitution rate. *Proceedings of the National Academy of Sciences*, 110, E611–E611.
274. Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular ecology*, 17, 167–178.
275. Gilbert C, Robertson G, Le Maho Y, Ancel A (2008) How do weather conditions affect the huddling behaviour of emperor penguins. *Polar Biology*, 31, 163–169.
276. Giraud T, Pedersen JS, Keller L (2002) Evolution of supercolonies: the Argentine ants of southern Europe. *Proceedings of the National Academy of Sciences*, 99, 6075–6079.
277. Goberville E, Beaugrand G, Hautekèete NC, Piquot Y, Luczak C (2015) Uncertainties in the projection of species distributions related to general circulation models. *Ecology and evolution*, 5, 1100–1116.
278. Gómez-Díaz E, González-Solís J, Peinado MA (2009) Population structure in a highly pelagic seabird, the Cory's shearwater *Calonectris diomedea*: an examination of genetics, morphology and ecology. *Marine Ecology Progress Series*, 382, 197–209.
279. González-Wevar CA, Nakano T, Cañete JI, Poulin E (2010) Molecular phylogeny and historical biogeography of *Na-*

- cella* in the Southern Ocean. *Molecular Phylogenetics and Evolution*, 56, 115–124.
280. Gorospe KD, Karl SA (2013) Genetic relatedness does not retain spatial pattern across multiple spatial scales: dispersal and colonization in the coral, *Pocillopora damicornis*. *Molecular Ecology*, 22, 3721–3736.
281. Goudet J (1995) Fstat: a computer program to calculate F-Statistics. *The Journal of Heredity*, 86, 485–486.
282. Gould SJ, Eldredge N (1972) Punctuated equilibria: an alternative to phyletic gradualism.
283. Gouretski V (1999) The large-scale thermohaline structure of the Ross Gyre. In: *Oceanography of the Ross Sea Antarctica*, pp. 77–100. Springer,
284. Grassé PP (1950) *Traité de Zoologie. Oiseaux*.
285. Gray GR (1844) Aptenodytes. *The Annals and Magazine of Natural History*, 13, 315.
286. Gregg WW, Conkright ME, Ginoux P, O'Reilly JE, Casey NW (2003) Ocean primary production and climate: Global decadal changes. *Geophys. Res. Lett.*, 30, 1–4.
287. Grémillet D, Charmantier A (2010) Shifts in phenotypic plasticity constrain the value of seabirds as ecological indicators of marine ecosystems. *Ecological Applications*, 20, 1498–1503.
288. Greminger MP, Stölting KN, Nater A et al. (2014) Generation of SNP datasets for orangutan population genomics using improved reduced-representation sequencing and direct comparisons of SNP calling algorithms. *BMC Genomics*, 15, 16.
289. Groenen MAM, Archibald AL, Uenishi H et al. (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, 491, 393–398.
290. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59, 307–321.
291. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*, 5, e1000695.
- ## H
292. Hackett SJ, Kimball RT, Reddy S et al. (2008) A phylogenomic study of birds reveals their evolutionary history. *Science*, 320, 1763–1768.
293. Haddad NM, Brudvig LA, Clobert J et al. (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, 1, e1500052.
294. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series* 41, 95–98.
295. Halpern BS, Walbridge S, Selkoe KA et al. (2008) A global map of human impact on marine ecosystems. *Science*, 319, 948–952.
296. Hamilton M (2011) Population genetics.
297. Handrich Y, Bevan RM, Charrassin J-B et al. (1997) Hypothermia in foraging king penguins. *Nature*, 388, 64–67.
298. Hanna E, Navarro FJ, Pattyn F et al. (2013) Ice-sheet mass balance and climate change. *Nature*, 498, 51–59.
299. Hanski I (1998) Metapopulation dynamics. *Nature*, 396, 41–49.

300. Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. *Nature*, 404, 755–758.
301. Harrison SP, Bartlein PJ, Izumi K et al. (2015) Evaluation of CMIP5 palaeo-simulations to improve climate projections. *Nature Clim. Change*, 5, 735–743.
302. Hart T, Clucas G, Levy H et al. (2011) Characterization of microsatellite loci for utility in a range of penguin species (*unpublished*).
303. Hartl DL, Clark AG, Clark AG (1988) Principles of population genetics.
304. Harvey BP, Gwynn Jones D, Moore PJ (2013) Meta analysis reveals complex marine biological responses to the interactive effects of ocean acidification and warming. *Ecology and Evolution*, 3, 1016–1030.
305. Hawkins BA, Diniz Filho JAF, Jaramillo CA, Soeller SA (2006) Post Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *Journal of Biogeography*, 33, 770–780.
306. Hays C (1989) The Peruvian diving petrel in Peru. *Oryx*, 23, 102–105.
307. Hayward BW, Scott GH, Crundwell MP et al. (2008) The effect of submerged plateaux on Pleistocene gyral circulation and sea-surface temperatures in the Southwest Pacific. *Global and Planetary Change*, 63, 309–316.
308. Haywood AM, Smellie JL, Ashworth AC et al. (2008) Middle Miocene to Pliocene history of Antarctica and the Southern Ocean. *Developments in Earth and Environmental Sciences*, 8, 401–463.
309. He F, Hubbell SP (2011) Species-area relationships always overestimate extinction rates from habitat loss. *Nature*, 473, 368–371.
310. Heg D, Rothenberger S, Schürch R (2011) Habitat saturation, benefits of philopatry, relatedness, and the extent of cooperative breeding in a cichlid. *Behavioral Ecology*, 22, 82–92.
311. Heithaus MR, Frid A, Wirsing AJ, Worm B (2008) Predicting ecological consequences of marine top predator declines. *Trends in Ecology & Evolution*, 23, 202–210.
312. Heled J, Drummond A (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, 8, 289.
313. Hellenthal G, Stephens M (2007) msHOT: modifying Hudson’s ms simulator to incorporate crossover and gene conversion hotspots. *Bioinformatics*, 23, 520–521.
314. Heupink TH, van den Hoff J, Lambert DM (2012) King penguin population on Macquarie Island recovers ancient DNA diversity after heavy exploitation in historic times. *Biol Lett*, 8, 586–589.
315. Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–913.
316. Hillier LW, Miller W, Birney E et al. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432, 695–716.
317. Ho SY, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol Ecol Resour*, 11, 423–434.
318. Hoban S, Gaggiotti O, Bertorelle G (2013a) Sample Planning Optimization Tool for conservation and population Genetics (SPOTG): a software for choosing the appropriate number of markers and samples. *Methods in Ecology and Evolution*, 4, 299–303.
319. Hoban SM, Gaggiotti OE, Bertorelle G (2013b) The number of markers and sam-

- ples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulation based study. *Molecular ecology*, 22, 3444–3450.
320. Hodell DA, Kanfoush SL, Shemesh A, Crosta X, Charles CD, Guilderson TP (2001) Abrupt cooling of Antarctic surface waters and sea ice expansion in the South Atlantic sector of the Southern Ocean at 5000 cal yr BP. *Quaternary Research*, 56, 191–198.
321. Hodgson DA, Graham AGC, Roberts SJ et al. (2014) Terrestrial and submarine evidence for the extent and timing of the Last Glacial Maximum and the onset of deglaciation on the maritime-Antarctic and sub-Antarctic islands. *Quaternary Science Reviews*, 100, 137–158.
322. Hoegh-Guldberg O, Bruno JF (2010) The impact of climate change on the world's marine ecosystems. *Science*, 328, 1523–1528.
323. Hoelzel AR (1998) Genetic structure of cetacean populations in sympatry, parapatry, and mixed assemblages: implications for conservation policy. *Journal of Heredity*, 89, 451–458.
324. Hoffman JI, Simpson F, David P et al. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proc Natl Acad Sci U S A*, 111, 3775–3780.
325. Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, 470, 479–485.
326. Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, 9, 421–432.
327. Hofmann M, Maqueda M (2011) The response of Southern Ocean eddies to increased midlatitude westerlies: A non eddy resolving model study. *Geophysical Research Letters*, 38,
328. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, 6, e1000862.
329. Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics*, 10, 639–650.
330. Hope AG, Waltari E, Payer DC, Cook JA, Talbot SL (2013) Future distribution of tundra refugia in northern Alaska. *Nature Clim. Change*, 3, 931–938.
331. Howard WE (1960) Innate and environmental dispersal of individual vertebrates. *American Midland Naturalist*, 152–161.
332. Huffaker C, Gutierrez AP (1984) Ecological entomology.
333. Hunt GL (1991) Occurrence of polar seabirds at sea in relation to prey concentrations and oceanographic factors. *Polar Research*, 10, 553–560.
334. Hunt GL, McKinnell S (2006) Interplay between top-down, bottom-up, and wasp-waist control in marine ecosystems. *Progress in Oceanography*, 68, 115–124.
335. Hunter CM, Caswell H, Runge MC, Regehr EV, Amstrup SC, Stirling I (2010) Climate change threatens polar bear populations: a stochastic demographic analysis. *Ecology*, 91, 2883–2897.
336. Hunter S (1991) The impact of avian predator-scavengers on King Penguin *Aptenodytes patagonicus* chicks at Marion Island. *Ibis*, 133, 343–350.
337. Huntington CE (1951) *Ortstreue* and Subspecies Formation in the Pied Flycatcher.
338. Husby A, Nussey DH, Visser ME, Wilson AJ, Sheldon BC, Kruuk LEB (2010) Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (*Parus*

- major) populations. *Evolution*, 64, 2221–2237.
339. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 23, 254–267.
340. Hutchinson GE (1957) Cold spring harbor symposium on quantitative biology. *Concluding remarks*, 22, 415–427.
341. Hyde WT, Crowley TJ, Baum SK, Peltier WR (2000) Neoproterozoic snowball Earth simulations with a coupled climate/ice-sheet model. *Nature*, 405, 425–429.
- ## I
342. Isenmann P (1971) Contribution à l'éthologie et à l'écologie du manchot empereur (*Aptenodytes forsteri* Gray) à la colonie de Pointe Géologie (Terre Adélie). PhD.
343. Isenmann P, Jouventin EP (1970) Ecoéthologie du Manchot empereur (*Aptenodytes forsteri*) et comparaison avec le Manchot Adélie (*Pygoscelis adeliae*) et le Manchot royal (*Aptenodytes paragonica*). *Expéditions Polaires Françaises*, 322, 136–159.
- ## J
344. Jackson JBC (2008) Ecological extinction and evolution in the brave new ocean. *Proceedings of the National Academy of Sciences*, 105, 11458–11465.
345. Jahncke J, Goya E (1998) The status of the Peruvian diving-petrel population at its main breeding areas along the coast of Peru. *Colonial Waterbirds*, 94–97.
346. Jahncke J, García-Godos A, Goya E (1999) The Diet of the Peruvian Diving-Petrel at la Vieja and San Gallan, Perú (La Dieta de Pelecanoides garnotii a la Vieja y San Gallan, Perú). *Journal of Field Ornithology*, 71–79.
347. Janko K, Lecointre G, DeVries A, Couloux A, Cruaud C, Marshall C (2007) Did glacial advances during the Pleistocene influence differently the demographic histories of benthic and pelagic Antarctic shelf fishes? Inferences from intraspecific mitochondrial and nuclear DNA sequence diversity. *BMC Evolutionary Biology*, 7, 220.
348. Jarvis ED, Mirarab S, Aberer AJ et al. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346, 1320–1331.
349. Jenouvrier S (2004) Influence de la variabilité environnementale sur les stratégies démographiques des populations de prédateurs supérieurs : La communauté d'oiseaux marins en Antarctique. PhD.
350. Jenouvrier S, Caswell H, Barbraud C, Holland M, Stroeve J, Weimerskirch H (2009) Demographic models and IPCC climate projections predict the decline of an emperor penguin population. *Proc Natl Acad Sci U S A*, 106, 1844–1847.
351. Jenouvrier S, Weimerskirch H, Barbraud C, Park YH, Cazelles B (2005) Evidence of a shift in the cyclicity of Antarctic seabird dynamics linked to climate. *Proceedings of the Royal Society B: Biological Sciences*, 272, 887.
352. Jenouvrier S, Barbraud C, Weimerskirch H (2006) Sea ice affects the population dynamics of Adélie penguins in Terre Adélie. *Polar Biol*, 29, 413–423.
353. Jenouvrier S, Holland M, Stroeve J et al. (2012) Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. *Global Change Biology*, 18, 2756–2770.
354. Jenouvrier S, Holland M, Stroeve J et al. (2014) Projected continent-wide declines of the emperor penguin under climate change. *Nature Climate change*, 4, 715–718.

355. Jeyasingham WS, Taylor SA, Zavalaga CB, Simeone A, Friesen VL (2013) Specialization to cold water upwellings may facilitate gene flow in seabirds: new evidence from the Peruvian pelican *Pelecanus thagus* (Pelecaniformes: Pelecanidae). *Journal of avian biology*, 44, 297–304.
356. Johnson K (1981) Social organization in a colony of rock squirrels (*Spermophilus variegatus*). *The Southwestern Naturalist*, 237–242.
357. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.
358. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, 11, 94.
359. Jones OR, Wang J (2012) A comparison of four methods for detecting weak genetic structure from marker data. *Ecology and Evolution*, 2, 1048–1055.
360. Jouventin P (1971a) Comportement et structure sociale chez le manchot empereur.
361. Jouventin P (1971b) Incubation et élevage itinérants chez les manchots empereurs de Pointe Géologie (Terre Adélie).
362. Jouventin P (1972a) Note sur l'existence et la signification d'une rythmicité des parades mutuelles.
363. Jouventin P (1972b) Un nouveau système de reconnaissance acoustique chez les oiseaux. *Behaviour*, 43, 176–185.
364. Jouventin P (1975) Mortality parameters in emperor penguins *Aptenodytes forsteri*. *The biology of penguins*. London: Macmillan, 435–446.
365. Jouventin P, Cuthbert RJ, Ottvall R (2006) Genetic isolation and divergence in sexual traits: evidence for the northern rock-hopper penguin *Eudyptes moseleyi* being a sibling species. *Molecular Ecology*, 15, 3413–3423.
366. Jouzel J, Masson-Delmotte V, Cattani O et al. (2007) Orbital and millennial Antarctic climate variability over the past 800,000 years. *science*, 317, 793–796.
367. Jump AS, Mátyás C, Peñuelas J (2009) The altitude-for-latitude disparity in the range retractions of woody species. *Trends in ecology & evolution*, 24, 694–701.
- ## K
368. Kanno Y, Vokoun JC, Letcher BH (2011) Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. *Molecular Ecology*, 20, 3711–3729.
369. Kaplan MR, Schaefer JM, Denton GH et al. (2010) Glacier retreat in New Zealand during the Younger Dryas stadial. *Nature*, 467, 194–197.
370. Karpouzi VS, Watson R, Pauly D (2007) Modelling and mapping resource overlap between seabirds and fisheries on a global scale: a preliminary assessment. *Marine Ecology Progress Series*, 343, 87–99.
371. Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American statistical association*, 90, 773–795.
372. Kauffman MJ, Frick WF, Linthicum J (2003) Estimation of habitat-specific demography and population growth for peregrine falcons in California. *Ecological Applications*, 13, 1802–1816.
373. Keller I, Largiader CR (2003) Recent habitat fragmentation caused by major roads leads to reduction of gene flow and loss of genetic variability in ground beetles. *Pro-*

- ceedings of the Royal Society of London B: Biological Sciences*, 270, 417–423.
374. Keller LF, Waller DM (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17, 230–241.
375. Kemp AES, Grigorov I, Pearce RB, Garabato ACN (2010) Migration of the Antarctic Polar Front through the mid-Pleistocene transition: evidence and climatic implications. *Quaternary Science Reviews*, 29, 1993–2009.
376. Kendall WL, Nichols JD (2004) On the estimation of dispersal and movement of birds. *The Condor*, 106, 720–731.
377. Keppel G, Van Niel KP, Wardell Johnson GW et al. (2012) Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, 21, 393–404.
378. Kierkegaard S (1843) *Gjentagelsen*. Et Forsøg i den experimenterende Psychologi af Constantin Constantius.
379. Kim SY, Lohmueller KE, Albrechtsen A et al. (2011) Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC bioinformatics*, 12, 231.
380. Kimura M (1955) Solution of a process of random genetic drift with a continuous model. *Proc Natl Acad Sci*, 41, 144.
381. Kimura M (1983) The neutral theory of molecular evolution.
382. Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics*, 49, 725.
383. Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability*, 27–43.
384. Kingman JFC (2000) Origins of the coalescent: 1974–1982. *Genetics*, 156, 1461–1463.
385. Kinsky FC, Falla RA (1976) A sub-specific revision of the Australasian blue penguin (*Eudyptula minor*) in the New Zealand area.
386. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF (2010) Genomic runs of homozygosity record population history and consanguinity. *PloS one*, 5, e13996.
387. Knorr G, Lohmann G (2014) Climate warming during Antarctic ice sheet expansion at the Middle Miocene transition. *Nature Geoscience*, 7, 376–381.
388. Kohfeld KE, Graham RM, de Boer AM et al. (2013) Southern Hemisphere westerly wind changes during the Last Glacial Maximum: paleo-data synthesis. *Quaternary Science Reviews*, 68, 76–95.
389. Kooyman GL (1993) Breeding habitats of emperor penguins in the western Ross Sea. *Antarctic Science*, 5, 143–148.
390. Kooyman GL, Ponganis PJ (1998) The physiological basis of diving to depth: birds and mammals. *Annual Review of Physiology*, 60, 19–32.
391. Kooyman GL, Ponganis PJ (2014) Chick production at the largest emperor penguin colony decreases by 50% from 2008–10. *Antarctic Science*, 26, 33–37.
392. Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, 15, 356.
393. Koudil M, Charrassin J-B, Le Maho Y, Bost C-A (2000) Seabirds as monitors of upper-ocean thermal structure. King penguins at the Antarctic polar front, east of Kerguelen sector. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, 323, 377–384.
394. Krebs CJ (2013) *Ecology: The Experimental Analysis of Distribution and Abundance*.

395. Krehenwinkel H, Rödder D, Tautz D (2015) Eco-genomic analysis of the poleward range expansion of the wasp spider *Argiope bruennichi* shows rapid adaptation and genomic admixture. *Glob Chang Biol*, 21, 4320–4332.
396. Kucera M, Rosell-Melé A, Schneider R, Waelbroeck C, Weinelt M (2005) Multi-proxy approach for the reconstruction of the glacial ocean surface (MARGO). *Quaternary Science Reviews*, 24, 813–819.
397. Kuhlbrodt T, Rahmstorf S, Zickfeld K et al. (2009) An integrated assessment of changes in the thermohaline circulation. *Climatic Change*, 96, 489–537.
398. Kusahara K, Sato T, Oka A et al. (2015) Modelling the Antarctic marine cryosphere at the Last Glacial Maximum. *Annals of Glaciology*, 56, 425.
399. Kusch A, Marín M (2012) Sobre la distribución del Pingüino Rey *Aptenodytes Patagonicus* (Aves: Spheniscidae) en Chile. *Anales del Instituto de la Patagonia* 40, 157–163.
- L
400. Lack DL (1968) Ecological adaptations for breeding in birds.
401. Lam F, Tarpine R, Istrail S (2010) The imperfect ancestral recombination graph reconstruction problem: upper bounds for recombination and homoplasy. *Journal of Computational Biology*, 17, 767–781.
402. Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, Baroni C (2002) Rates of evolution in ancient DNA from Adélie penguins. *Science*, 295, 2270.
403. Langham GM, Schuetz JG, Distler T, Soykan CU, Wilsey C (2015) Conservation Status of North American Birds in the Face of Future Climate Change. *PLoS One*, 10, e0135350.
404. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9, 357–359.
405. LaRue MA, Kooyman G, Lynch HJ, Fretwell P (2015) Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography*, 38, 114–120.
406. Laubscher RK, Perissinotto R, McQuaid CD (1993) Phytoplankton production and biomass at frontal zones in the Atlantic sector of the Southern Ocean. *Polar biology*, 13, 471–481.
407. Le Boeuf BJ, Crocker DE, Costa DP, Blackwell SB, Webb PM, Houser DS (2000) Foraging ecology of northern elephant seals. *Ecological monographs*, 70, 353–382.
408. Le Bohec C (2008) Stratégies d’histoire de vie d’un oiseau longévif: le manchot royal (*Aptenodytes patagonicus*). PhD.
409. Le Bohec C, Durant JM, Gauthier-Clerc M et al. (2008) King penguin population threatened by Southern Ocean warming. *Proc Natl Acad Sci U S A*, 105, 2493–2497.
410. Le Bohec C, Gauthier-Clerc M, Grémillet D et al. (2007) Population dynamics in a long-lived seabird: I. Impact of breeding activity on survival and breeding probability in unbanded king penguins. *Journal of Animal Ecology*, 76, 1149–1160.
411. Le Maho Y (1977) The Emperor Penguin: A Strategy to Live and Breed in the Cold: Morphology, physiology, ecology, and behavior distinguish the polar emperor penguin from other penguin species, particularly from its close relative, the king penguin. *American Scientist*, 680–693.
412. Le Maho Y, Saraux C, Durant JM et al. (2011) An ethical issue in biodiversity science: The monitoring of penguins with flipper bands. *C R Biol*, 334, 378–384.
413. Le Vaillant M, Ropert-Coudert Y, Le Maho Y, Le Bohec C (2015) Individual parameters shape foraging activity in breeding

- king penguins. *Behavioral Ecology*, 00, 1–11.
414. Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of statistical software*, 25, 1–18.
415. Lea M, Soper T (2005) Discovery of the first Emperor penguin *Aptenodytes forsteri* colony in Marie Byrd Land, Antarctica. *Marine Ornithology*, 33, 59–60.
416. Leaper R, Cooke J, Trathan P, Reid K, Rowntree V, Payne R (2006) Global climate drives southern right whale (*Eubalaena australis*) population dynamics. *Biology Letters*, 2, 289–292.
417. Lecomte N, Kuntz G, Lambert N et al. (2006) Alloparental feeding in the king penguin. *Animal behaviour*, 71, 457–462.
418. Lecomte N, Gauthier G, GIROUX JEAN, Milot E, Bernatchez L (2009) Tug of war between continental gene flow and rearing site philopatry in a migratory bird: the sex-biased dispersal paradigm reconsidered. *Molecular Ecology*, 18, 593–602.
419. Lefebvre W (2004) Influence of the Southern Annular Mode on the sea ice-ocean system. *J. Geophys. Res.*, 109, 1–12.
420. Legrand M, Feniet-Saigne C, Sattzman ES, Germain C, Barkov NI, Petrov VN (1991) Ice-core record of oceanic emissions of dimethylsulphide during the last climate cycle.
421. Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17, 183–189.
422. Lewis Smith RI, Tallowin JRB (1979) The distribution and size of king penguin rookeries on South Georgia. *Br Antarct Surv Bull*, 49, 259–276.
423. Lewison RL, Crowder LB, Read AJ, Freeman SA (2004) Understanding impacts of fisheries bycatch on marine megafauna. *Trends in Ecology & Evolution*, 19, 598–604.
424. Li C, Zhang Y, Li J et al. (2014) Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *Gigascience*, 3, 27.
425. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25, 1754–1760.
426. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature*, 475, 493–496.
427. Li H, Handsaker B, Wysoker A et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
428. Lindsay BG (1988) Composite likelihood methods. *Contemporary mathematics*, 80, 221–239.
429. Lisiecki LE, Raymo ME (2005) A Pliocene Pleistocene stack of 57 globally distributed benthic ^{18}O records. *Paleoceanography*, 20,
430. Liu J, Curry JA (2010) Accelerated warming of the Southern Ocean and its impacts on the hydrological cycle and sea ice. *Proceedings of the National Academy of Sciences*, 107, 14987–14992.
431. Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *Bmc Genetics*, 6, S26.
432. Liu N, Zhao H (2006) A non-parametric approach to population structure inference using multilocus genotypes. *Hum Genomics*, 2, 353.
433. Liu X, Fu Y-X (2015) Exploring population size changes using SNP frequency spectra. *Nature genetics*, 47, 555–559.

434. Livermore R, Eagles G, Morris P, Maldonado A (2004) Shackleton Fracture Zone: no barrier to early circumpolar ocean circulation. *Geology*, 32, 797–800.
435. Livermore R, Nankivell A, Eagles G, Morris P (2005) Paleogene opening of Drake passage. *Earth and Planetary Science Letters*, 236, 459–470.
436. Livezey BC, Zusi RL (2007) Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool J Linn Soc*, 149, 1–95.
437. Lovenduski NS, Gruber N (2005) Impact of the Southern Annular Mode on Southern Ocean circulation and biology. *Geophysical Research Letters*,
438. Lucas I, Palakodeti A, Jiang Y et al. (2007) High throughput mapping of origins of replication in human cells. *EMBO reports*, 8, 770–777.
439. Luna-Jorquera G, Fernández CE, Rivadeneira MM (2012) Determinants of the diversity of plants, birds and mammals of coastal islands of the Humboldt current systems: implications for conservation. *Biodiversity and Conservation*, 1–20.
440. Luna-Jorquera G, Simeone A, Aguilar R (2003) Ecofisiología de animales endotermos en un desierto cálido y un mar frío: el caso de las aves marinas de la corriente de Humboldt. *Fisiología ecológica y evolutiva. Teoría y casos de estudios en animales*, 297–316.
441. Lund AH, Duch M, Pedersen FS (1996) Increased cloning efficiency by temperature-cycle ligation. *Nucleic acids research*, 24, 800–801.
442. Lynch HJ, Naveen R, Trathan PN, Fagan WF (2012) Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology*, 93, 1367–1377.
443. Lynch M, Conery JS (2003) The origins of genome complexity. *science*, 302, 1401–1404.

M

444. Mackintosh NA (1946) The Antarctic Convergence and the distribution of surface temperatures in Antarctic waters.
445. Mahlstein I, Daniel JS, Solomon S (2013) Pace of shifts in climate regions increases with global temperature. *Nature Clim. Change*, 3, 739–743.
446. Majumdar A, Gries C, Walker J (2011) A non-stationary spatial generalized linear mixed model approach for studying plant diversity. *Journal of Applied Statistics*, 38, 1935–1950.
447. Mallory ML, Robinson SA, Hebert CE, Forbes MR (2010) Seabirds as indicators of aquatic ecosystem conditions: A case for gathering multiple proxies of seabird health. *Marine Pollution Bulletin*, 60, 7–12.
448. Mangin S, Gauthier-Clerc M, Frenot Y, Gendner J-P, Le Maho Y (2003) Ticks *Ixodes uriae* and the breeding performance of a colonial seabird, king penguin *Aptenodytes patagonicus*. *J Avian Biol*, 34, 30–34.
449. MARGO (2009) Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. *Nature Geoscience*, 2, 127–132.
450. Marshall GJ (2003) Trends in the Southern Annular Mode from observations and reanalyses. *Journal of Climate*, 16, 4134–4143.
451. Marshall J, Speer K (2012) Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nature Geoscience*, 5, 171–180.
452. Martínez Garcia A, Rosell Melé A, Geibert W et al. (2009) Links between iron supply, marine productivity, sea surface tem-

- perature, and CO₂ over the last 1.1 Ma. *Paleoceanography*, 24,
453. Martínez Meyer E, Townsend Peterson A, Hargrove WW (2004) Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography*, 13, 305–314.
454. Martinez Palma P (2014) Reporte Anual Regional del cumplimiento de las actividades del Plan de Conservación de la especie Pelecanoides garnotii (Yunco). *Departamento de Areas Silvestres Protegidas, region Coquimbo*.
455. Massom RA, Hill K, Barbraud C et al. (2009) Fast ice distribution in Adélie Land, East Antarctica: interannual variability and implications for emperor penguins *Aptenodytes forsteri*. *Marine Ecology Progress Series*, 374, 243–257.
456. Mathews G (1911) The Birds of Australia.
457. Matschiner M (2015) Fitchi: Haplotype genealogies based on Fitch distances.
458. Matschiner M, Hanel R, Salzburger W (2009) Gene flow by larval dispersal in the Antarctic notothenioid fish *Gobionotothen gibberifrons*. *Molecular Ecology*, 18, 2574–2587.
459. Mattern T, Ellenberg U, Luna-Jorquera G (2002) A South American marine otter *Lontra felina* preys upon chicks of the Peruvian diving petrel *Pelecanoides garnotii*. *Marine Ornithology*, 30, 95–96.
460. Matthiopoulos J, Harwood J, Thomas L (2005) Metapopulation consequences of site fidelity for colonially breeding mammals and birds. *Journal of Animal Ecology*, 74, 716–727.
461. Mayr E (1963) Animal species and evolution.
462. Mayr E (1970) Populations, species, and evolution: an abridgment of animal species and evolution.
463. Mayr E (1982) The growth of biological thought: diversity, evolution, and inheritance.
464. Mayr E, Provine WB (1998) The evolutionary synthesis: perspectives on the unification of biology.
465. McCauley DJ, DeSalles PA, Young HS et al. (2012) From wing to wing: the persistence of long ecological interaction chains in less-disturbed ecosystems. *Scientific Reports*, 2,
466. McCauley DJ, Pinsky ML, Palumbi SR, Estes JA, Joyce FH, Warner RR (2015) Marine defaunation: Animal loss in the global ocean. *Science*, 347, 1255641.
467. McElhany P, Ruckelshaus MH, Ford MJ, Wainwright TC, Bjorkstedt EP (2000) Viable salmonid populations and the recovery of evolutionarily significant units. US Dept. Commer. *NOAA Tech. Memo. NMFS-NWFSC*, 42, 156.
468. McGlone MS, Turney CSM, Wilmshurst JM, Renwick J, Pahnke K (2010) Divergent trends in land and ocean temperature in the Southern Ocean over the past 18,000 years. *Nature Geoscience*, 3, 622–626.
469. McGuigan K, Nishimura N, Currey M, Hurwit D, Cresko WA (2010) Quantitative genetic variation in static allometry in the threespine stickleback. *Integrative and comparative biology*, 50, 1067–1080.
470. McKay R, Naish T, Carter L et al. (2012) Antarctic and Southern Ocean influences on Late Pliocene global cooling. *Proceedings of the National Academy of Sciences*, 109, 6423–6428.
471. McKechnie AE, Freckleton RP, Jetz W (2006) Phenotypic plasticity in the scaling of avian basal metabolic rate. *Proceed-*

- ings of the Royal Society of London B: Biological Sciences*, 273, 931–937.
472. McMahon CR, Burton HR (2005) Climate change and seal survival: evidence for environmentally mediated changes in elephant seal, *Mirounga leonina*, pup survival. *Proceedings of the Royal Society of London B: Biological Sciences*, 272, 923–928.
473. McVean GAT, Cardin NJ (2005) Approximating the coalescent with recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1387–1393.
474. Meijers AJS (2014) The Southern Ocean in the Coupled Model Intercomparison Project phase 5. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 372, 20130296.
475. Meinshausen M, Smith SJ, Calvin K et al. (2011) The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic change*, 109, 213–241.
476. Melick D, Bremmers W (1995) A recently discovered breeding colony of emperor penguins (*Aptenodytes forsteri*) on the Budd Coast, Wilkes Land, East Antarctica. *Polar Record*, 31, 426–427.
477. Meredith MP, Locarnini RA, Van Scoy KA, Watson AJ, Heywood KJ, King BA (2000) On the sources of Weddell Gyre Antarctic bottom water. *Journal of Geophysical Research: Oceans (1978–2012)*, 105, 1093–1104.
478. Mesak F, Tatarenkov A, Earley RL, Avise JC (2014) Hundreds of SNPs vs. dozens of SSRs: which dataset better characterizes natural clonal lineages in a self-fertilizing fish. *Frontiers in Ecology and Evolution*, 2, 74.
479. Midgley GF, Bond WJ (2015) Future of African terrestrial biodiversity and ecosystems under anthropogenic climate change. *Nature Clim. Change*, 5, 823–829.
480. Millar CD, Dodd A, Anderson J et al. (2008) Mutation and evolutionary rates in Adélie penguins from the Antarctic. *PLoS Genetics*, 4, e1000209.
481. Miller JA (2012) Species distribution models: Spatial autocorrelation and non-stationarity. *Progress in Physical Geography*, 37, 161–177.
482. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research*, 17, 240–248.
483. Miller W, Schuster SC, Welch AJ et al. (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences*, 109, E2382–E2390.
484. Mills LS, Allendorf FW (1996) The one-migrant-per-generation rule in conservation and management. *Conservation Biology*, 10, 1509–1518.
485. Milot E, Weimerskirch H, Bernatchez L (2008) The seabird paradox: dispersal, genetic structure and population dynamics in a highly mobile, but philopatric albatross species. *Molecular ecology*, 17, 1658–1673.
486. Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol*, 25, 1459–1471.
487. Möbius KA (1877) Die Auster und die Austernwirthschaft.
488. Moe B, Stempniewicz L, Jakubas D et al. (2009) Climate change and phenological responses of two seabird species breed-

- ing in the high-Arctic. *Mar Ecol Prog Ser*, 393, 235–246.
489. Molfetti E, Vilaca ST, Georges JY et al. (2013) Recent demographic history and present fine-scale structure in the Northwest Atlantic leatherback (*Dermochelys coriacea*) turtle population. *PLoS One*, 8, e58061.
490. Møller AP, Fiedler W, Berthold P (2010) Effects of climate change on birds.
491. Møller AP, Rubolini D, Lehikoinen E (2008) Populations of migratory bird species that did not show a phenological response to climate change are declining. *Proceedings of the National Academy of Sciences*, 105, 16195–16200.
492. Montecino V, Strub PT, Chavez F, Thomas A, Tarazona J, Baumgartner T (2006) Bio-physical interactions off western South-America. *The sea*, 14, 329–390.
493. Moore GJ, Robertson G, Wienecke B (1998) Food requirements of breeding king penguins at Heard Island and potential overlap with commercial fisheries. *Polar Biology*, 20, 293–302.
494. Moore JK, Abbott MR, Richman JG (1999) Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data. *Journal of Geophysical Research*, 104,
495. Morin PA, Parsons KM, Archer FI et al. (2015) Geographic and temporal dynamics of a global radiation and diversification in the killer whale. *Mol Ecol*, 24, 3964–3979.
496. Moritz C, Agudo R (2013) The future of species under climate change: resilience or decline. *Science*, 341, 504–508.
497. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5, 621–628.
498. Moss RH, Edmonds JA, Hibbard KA et al. (2010) The next generation of scenarios for climate change research and assessment. *Nature*, 463, 747–756.
499. Mougin J-L, Van Beveren M (1979) Structure et dynamique de la population de Manchots empereurs *Aptenodytes forsteri* de la colonie de l'archipel de Pointe Géologie, Terre Adélie.
500. Moy CM, Seltzer GO, Rodbell DT, Anderson DM (2002) Variability of El Niño/Southern Oscillation activity at millennial timescales during the Holocene epoch. *Nature*, 420, 162–165.
501. Mulvaney R, Abram NJ, Hindmarsh RCA et al. (2012) Recent Antarctic Peninsula warming relative to Holocene climate and ice-shelf history. *Nature*, 489, 141–144.
502. Munafò MR, Clark TG, Flint J (2004) Assessing publication bias in genetic association studies: evidence from a recent meta-analysis. *Psychiatry research*, 129, 39–44.
503. Murphy EJ, Watkins JL, Trathan PN et al. (2007a) Spatial and temporal operation of the Scotia Sea ecosystem: a review of large-scale links in a krill centred food web. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 113–148.
504. Murphy EJ, Trathan PN, Watkins JL et al. (2007b) Climatically driven fluctuations in Southern Ocean ecosystems. *Proceedings of the Royal Society of London B: Biological Sciences*, 274, 3057–3067.
505. Murphy RC (1915) The penguins of South Georgia. *Brooklyn museum bulletin*, 5, 103–133.
506. Murphy RC (1936) Oceanic birds of South America: a study of species of the related coasts and seas, including the American quadrant of Antarctica, based upon the Brewster-Sanford collection in the American Museum of Natural History.
507. Murrell DJ, Travis JMJ, Dytham C (2002) The evolution of dispersal distance in

- spatially-structured populations. *Oikos*, 97, 229–236.
508. Myers RA, Worm B (2003) Rapid worldwide depletion of predatory fish communities. *Nature*,
509. Myers S, Fefferman C, Patterson N (2008) Can one learn history from the allelic spectrum? *Theoretical population biology*,
510. Myrcha A (2006) Eocene penguins of Seymour Island, Antarctica: Taxonomy. *Pol. Polar Res*, 27, 3–62.
- N
511. Nash SB (2011) Persistent organic pollutants in Antarctica: current and future research priorities. *Journal of Environmental Monitoring*, 13, 497–504.
512. Nathan R (2006) Long-distance dispersal of plants. *Science*, 313, 786–788.
513. Nathan R, Muller-Landau HC (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in ecology & evolution*, 15, 278–285.
514. Nathan R, Schurr FM, Spiegel O, Steinitz O, Trakhtenbrot A, Tsoar A (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology & Evolution*, 23, 638–647.
515. Navarro J, Cardador L, Brown R, Phillips RA (2015) Spatial distribution and ecological niches of non-breeding planktivorous petrels. *Sci Rep*, 5, 12164.
516. Navarro J, Votier SC, Aguzzi J, Chiesa JJ, Forero MG, Phillips RA (2013) Ecological segregation in space, time and trophic niche of sympatric planktivorous petrels. *PLoS One*, 8, e62897.
517. Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76, 5269–5273.
518. Neil HL, Carter L, Morris MY (2004) Thermal isolation of Campbell Plateau, New Zealand, by the Antarctic Circumpolar Current over the past 130 kyr. *Paleoceanography*, 19,
519. Nesterova AP, Le Bohec C, Beaune D, Pettex E, Le Maho Y, Bonadonna F (2010) Do penguins dare to walk at night? Visual cues influence king penguin colony arrivals and departures. *Behavioral Ecology and Sociobiology*, 64, 1145–1156.
520. Nicol S, Clarke J, Romaine SJ, Kawaguchi S, Williams G, Hosie GW (2008) Krill (*Euphausia superba*) abundance and Adélie penguin (*Pygoscelis adeliae*) breeding performance in the waters off the Béchervaise Island colony, East Antarctica in 2 years with contrasting ecological conditions. *Deep Sea Research Part II: Topical Studies in Oceanography*, 55, 540–557.
521. Nicol S, Pauly T, Bindoff NL et al. (2000) Ocean circulation off east Antarctica affects ecosystem structure and sea-ice extent. *Nature*, 406, 504–507.
522. Nicotra AB, Atkin OK, Bonser SP et al. (2010) Plant phenotypic plasticity in a changing climate. *Trends in plant science*, 15, 684–692.
523. Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PloS one*, 7, e37558.
524. Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet*, 12, 443–451.
525. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. *Genome research*, 15, 1566–1575.

526. Nielsen SHH, Koç N, Crosta X (2004) Holocene climate in the Atlantic sector of the Southern Ocean: Controlled by insolation or oceanic circulation. *Geology*, 32, 317–320.
527. Nikula R, Fraser CI, Spencer HG, Waters JM (2010) Circumpolar dispersal by rafting in two subantarctic kelp-dwelling crustaceans. *Mar Ecol Prog Ser*, 405, 221–230.
528. Nikula R, Spencer HG, Waters JM (2013) Passive rafting is a powerful driver of transoceanic gene flow. *Biology letters*, 9, 20120821.
529. Nong GT, Najjar RG, Seidov D, Peterson WH (2000) Simulation of ocean temperature change due to the opening of Drake Passage. *Geophysical Research Letters*, 27, 2689–2692.
530. Norberg J, Urban MC, Vellend M, Klausmeier CA, Loeuille N (2012) Eco-evolutionary responses of biodiversity to climate change. *Nature Clim. Change*, 2, 747–751.
531. Nunn GB, Stanley SE (1998) Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. *Molecular Biology and Evolution*, 15, 1360–1371.
532. Nussey DH, Postma E, Gienapp P, Visser ME (2005) Selection on heritable phenotypic plasticity in a wild bird population. *Science*, 310, 304–306.
- O
533. O'Connell J, Gurdasani D, Delaneau O et al. (2014) A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet*, 10, e1004234.
534. O'Roak BJ, Vives L, Girirajan S et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, 485, 246–250.
535. Olsson O (1998) Divorce in king penguins: asynchrony, expensive fat storing and ideal free mate choice. *Oikos*, 574–581.
536. Olsson O, Bonnedahl J, Anker-Nilssen P (2001) Mate switching and copulation behaviour in king penguins. *Journal of avian biology*, 32, 139–145.
537. Olsson O, Brodin A (1997) Changes in King Penguin breeding cycle in response to food availability. *Condor*, 994–997.
538. Olsson O, van der Jeugd HP (2002) Survival in king penguins *Aptenodytes patagonicus*: temporal and sex-specific effects of environmental variability. *Oecologia*, 132, 509–516.
539. Onley D, Scofield P (2013) Albatrosses, petrels and shearwaters of the world.
540. Orsi AH, Whitworth T, Nowlin WD (1995) On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep Sea Research Part I: Oceanographic Research Papers*, 42, 641–673.
541. Osborne PE, Foody GM, Suárez-Seoane S (2007) Non-stationarity and local approaches to modelling the distributions of wildlife. *Diversity and Distributions*, 13, 313–323.
542. Overeem RL, Peucker AJ, Austin CM, Dann P, Burrige CP (2008) Contrasting genetic structuring between colonies of the World's smallest penguin, *Eudyptula minor* (Aves: Spheniscidae). *Conservation genetics*, 9, 893–905.
- P
543. Pachauri RK, Allen MR, Barros VR et al. (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

544. Pacifici M, Foden WB, Visconti P et al. (2015) Assessing species vulnerability to climate change. *Nature Climate change*, 5, 215–224.
545. Paetkau D, Slade R, Burden M, Estoup A (2003) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, 13, 55–65.
546. Pakhomov EA, Perissinotto R, McQuaid CD (1996) Prey composition and daily rations of myctophid fishes in the Southern Ocean. *Marine ecology progress series. Oldendorf*, 134, 1–14.
547. Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JPA (2005) Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Medicine*, 2, 1309.
548. Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.
549. Park YH, Durand I, Kestenare E et al. (2014) Polar Front around the Kerguelen Islands: An up to date determination and associated circulation of surface/subsurface waters. *Journal of Geophysical Research: Oceans*, 119, 6575–6592.
550. Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42.
551. Parsons M, Mitchell I, Butler A et al. (2008) Seabirds as indicators of the marine environment. *ICES Journal of Marine Science*, 65, 1520–1526.
552. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6, 288–295.
553. Pearce JM, Blums P, Lindberg MS (2008) Site fidelity is an inconsistent determinant of population structure in the hooded merganser (*Lophodytes cucullatus*): evidence from genetic, mark-recapture, and comparative data. *The Auk*, 125, 711–722.
554. Pearce JM (2007) Philopatry: A return to origins. *The Auk*, 124, 1085–1087.
555. Pearce JM, Talbot SL, Petersen MR, Rearick JR (2005) Limited genetic differentiation among breeding, molting, and wintering groups of the threatened Steller's eider: the role of historic and contemporary factors. *Conservation Genetics*, 6, 743–757.
556. Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful. *Global ecology and biogeography*, 12, 361–371.
557. Pelletier F, Reale D, Garant D, Coltman DW, Festa-Bianchet M (2007) Selection on heritable seasonal phenotypic plasticity of body mass. *Evolution*, 1969–1979.
558. Pereira HM, Leadley PW, Proença V et al. (2010) Scenarios for global biodiversity in the 21st century. *Science*, 330, 1496–1501.
559. Péron C, Weimerskirch H, Bost C-A (2012) Projected poleward shift of king penguins' (*Aptenodytes patagonicus*) foraging range at the Crozet Islands, southern Indian Ocean. *Proceedings of the Royal Society of London B: Biological Sciences*, 279, 2515–2523.
560. Peter HU, Buesser C, Mustafa O, Pfeiffer S (2008) Risk assessment for the Fildes Peninsula and Ardley Island and the development of management plans for designation as Antarctic Specially Protected or Managed Areas. *Federal Environmental Agency, Texte*, 20, 508.
561. Peterson RG, Whitworth T (1989) The Subantarctic and Polar Fronts in relation to deep water masses through the southwest-

- ern Atlantic. *Journal of Geophysical Research: Oceans* (1978–2012), 94, 10817–10838.
562. Petit J-R, Jouzel J, Raynaud D et al. (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, 399, 429–436.
563. Petit RJ, Hu FS, Dick CW (2008) Forests of the past: a window to future changes. *science*, 320, 1450–1452.
564. Peucker AJ, Dann P, BurrIDGE CP (2009) Range-wide phylogeography of the little penguin (*Eudyptula minor*): evidence of long-distance dispersal. *The Auk*, 126, 397–408.
565. Pfuhl HA, McCave IN (2005) Evidence for late Oligocene establishment of the Antarctic Circumpolar Current. *Earth and Planetary Science Letters*, 235, 715–728.
566. Piatt JF, Nettleship DN (1985) Diving depths of four alcids. *The Auk*, 293–297.
567. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*, 8, e1002967.
568. Pistorius PA, Baylis A, Crofts S, Pütz K (2012) Population development and historical occurrence of king penguins at the Falkland Islands. *Antarctic Science*, 24, 435–440.
569. Pithan F, Mauritsen T (2014) Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, 7, 181–184.
570. Poloczanska ES, Brown CJ, Sydeman WJ et al. (2013) Global imprint of climate change on marine life. *Nature Clim. Change*, 3, 919–925.
571. Ponchon A, Garnier R, Grémillet D, Boulinier T (2015) Predicting population responses to environmental change: the importance of considering informed dispersal strategies in spatially structured population models. *Diversity and Distributions*, 21, 88–100.
572. Poulin E, González-Wevar C, Díaz A, Gérard K, Hüne M (2014) Divergence between Antarctic and South American marine invertebrates: What molecular biology tells us about Scotia Arc geodynamics and the intensification of the Antarctic Circumpolar Current. *Global and Planetary Change*, 123, 392–399.
573. Prévost J (1961) L'Écologie du manchot empereur.
574. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38, 904–909.
575. Primeau F (2005) Characterizing transport between the surface mixed layer and the ocean interior with a forward and adjoint global ocean transport model. *Journal of Physical Oceanography*, 35, 545–564.
576. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
577. Prugnolle F, De Meeûs T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity*, 88, 161–165.
578. Przybylo R, Sheldon BC, Merilä J (2000) Climatic effects on breeding and morphology: evidence for phenotypic plasticity. *Journal of animal ecology*, 69, 395–403.
579. Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81, 559–575.
580. Purkey SG, Johnson GC (2010) Warming of global abyssal and deep South-

ern Ocean waters between the 1990s and 2000s: contributions to global heat and sea level rise budgets. *Journal of Climate*, 23, 6336–6351.

581. Pütz K, Cherel Y (2005) The diving behaviour of brooding king penguins (*Aptenodytes patagonicus*) from the Falkland Islands: variation in dive profiles and synchronous underwater swimming provide new insights into their foraging strategies. *Marine Biology*, 147, 281–290.

582. Pütz K (2002) Spatial and temporal variability in the foraging areas of breeding king penguins. *The Condor*, 104, 528–538.

583. Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, 155, 1429–1437.

Q

584. Queiroz AD (2005) The resurrection of oceanic dispersal in historical biogeography. *Trends in ecology & evolution*, 20, 68–73.

585. Quilty PG (2007) Origin and evolution of the sub-Antarctic islands: the foundation. *Papers and Proceedings of the Royal Society of Tasmania*, 141, 35–58.

586. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.

R

587. Rahmstorf S (2007) A semi-empirical approach to projecting future sea-level rise. *Science*, 315, 368–370.

588. Raine JJ, Askin RA (2001) Terrestrial palynology of Cape Roberts Project Drill-

hole CRP-3, Victoria Land Basin, Antarctica. *Terra Antartica*, 8, 389–400.

589. Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, 197, 573–589.

590. Rambaut A, Drummond A (2005) Tracer version 1.3: a program for analyzing results from Bayesian MCMC programs such as BEAST and MrBayes. *Distributed by the authors at: <http://evolove.zoo.ox.ac.uk/software.html>*.

591. Rambaut A, Grass NC (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer applications in the biosciences*, 13, 235–238.

592. Rasmussen MD, Hubisz MJ, Gronau I, Siepel A (2014) Genome-wide inference of ancestral recombination graphs.

593. Raybaud V, Beaugrand G, Dewarumez J-M, Luczak C (2015) Climate-induced range shifts of the American jackknife clam *Ensis directus* in Europe. *Biological Invasions*, 17, 725–741.

594. Raybaud V, Beaugrand G, Goberville E et al. (2013) Decline in kelp in west Europe and climate. *PloS one*, 8, e66044.

595. Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of heredity*, 86, 248–249.

596. Réale D, McAdam AG, Boutin S, Berteaux D (2003) Genetic and plastic responses of a northern mammal to climate change. *Proceedings of the Royal Society of London B: Biological Sciences*, 270, 591–596.

597. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature*, 461, 489–494.

598. Reich DE, Cargill M, Bolk S et al. (2001) Linkage disequilibrium in the human genome. *Nature*, 411, 199–204.
599. Rex MA, Stuart CT, Hessler RR, Allen JA, Sanders HL, Wilson GDF (1993) Global-scale latitudinal patterns of species diversity in the deep-sea benthos. *Nature*, 365, 636–639.
600. Reynolds J, Weir B, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105, 767–779.
601. Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. *Journal of climate*, 15, 1609–1625.
602. Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in ecology & evolution*, 29, 165–176.
603. Riesch R, Barrett-Lennard LG, Ellis GM, Ford JKB, Deecke VB (2012) Cultural traditions and the evolution of reproductive isolation: ecological speciation in killer whales. *Biological Journal of the Linnean Society*, 106, 1–17.
604. Rintoul S, Hughes C, Olbers D (2001) The Antarctic circumpolar current system. In: *Ocean Circulation and Climate*, (eds Siedler G, Church J, Gould J), pp. 271–302.
605. Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research*, 67, 175–186.
606. Ritz MS, Millar C, Miller GD et al. (2008) Phylogeography of the southern skua complex-rapid colonization of the southern hemisphere during a glacial period and reticulate evolution. *Mol Phylogenet Evol*, 49, 292–303.
607. Rivers NM, Butlin RK, Altringham JD (2005) Genetic population structure of Natterer's bats explained by mating at swarming sites and philopatry. *Molecular Ecology*, 14, 4299–4312.
608. Roeder AD, Marshall RK, Mitchellson AJ et al. (2001) Gene flow on the ice: genetic differentiation among Adélie penguin colonies around Antarctica. *Molecular Ecology*, 10, 1645–1656.
609. Rogers AD (2007) Evolution and biodiversity of Antarctic organisms: a molecular perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 2191–2214.
610. Rolland C, Danchin E, Fraipont MD (1998) The evolution of coloniality in birds in relation to food, habitat, predation, and life-history traits: a comparative analysis. *The American Naturalist*, 151, 514–529.
611. Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. *Science*, 301, 508–510.
612. Romiguier J, Gayral P, Ballenghien M et al. (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515, 261–263.
613. Rosenthal Y, Boyle EA, Labeyrie L (1997) Last glacial maximum paleochemistry and deepwater circulation in the Southern Ocean: Evidence from foraminiferal cadmium. *Paleoceanography*, 12, 787–796.
614. Rounsevell DE, Copson GR (1982) Growth rate and recovery of a king penguin, *Aptenodytes patagonicus*, population after exploitation. *Wildlife Research*, 9, 519–525.
615. Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, 15, 174–175.
616. Ruzzante DE, Hansen MM, Meldrup D (2002) Distribution of individual inbreeding coefficients, relatedness and influence of stocking on native anadromous brown trout

(*Salmo trutta*) population structure. *Molecular Ecology*, 10, 2107–2128.

617. Ryman N, Palm S, André C et al. (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology*, 15, 2031–2045.

S

618. Saether BE, Lande R, Engen S et al. (2005) Generation time and temporal scaling of bird population dynamics. *Nature*, 436, 99–102.

619. Sæther B-E, Engen S, Lande R (1999) Finite metapopulation models with density dependent migration and stochastic local dynamics. *Proceedings of the Royal Society of London B: Biological Sciences*, 266, 113–118.

620. Saji NH, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. *Nature*, 401, 360–363.

621. Saji NH, Yamagata T (2003) Possible impacts of Indian Ocean dipole mode events on global climate. *Climate Research*, 25, 151–169.

622. Saraux C, Le Bohec C, Durant JM et al. (2011a) Reliability of flipper-banded penguins as indicators of climate change. *Nature*, 469, 203–206.

623. Saraux C, Viblanc VA, Hanuise N, Le Maho Y, Le Bohec C (2011b) Effects of Individual Pre-Fledging Traits and Environmental Conditions on Return Patterns in Juvenile King Penguins. *PloS one*, 6, e20407.

624. Saraux C, Friess B, Le Maho Y, Le Bohec C (2012) Chick-provisioning strategies used by king penguins to adapt to a multiseasonal breeding cycle. *Animal Behaviour*, 84, 675–683.

625. Sarmiento JL, Gruber N, Brzezinski MA, Dunne JP (2004) High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature*, 427, 56–60.

626. Scheffer M, Carpenter S, Foley JA, Folke C, Walker B (2001) Catastrophic shifts in ecosystems. *Nature*, 413, 591–596.

627. Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. *Nature genetics*, 46, 919–927.

628. Schloss CA, Nuñez TA, Lawler JJ (2012) Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. *Proceedings of the National Academy of Sciences*, 109, 8606–8611.

629. Schlosser JA, Dubach JM, Garner TWJ et al. (2009) Evidence for gene flow differs from observed dispersal patterns in the Humboldt penguin, *Spheniscus humboldti*. *Conservation Genetics*, 10, 839–849.

630. Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in zoology*, 4, 1–13.

631. Seehausen O, Butlin RK, Keller I et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics*, 15, 176–192.

632. Serrano D, Forero MG, Donázar JA, Tella JL (2004) Dispersal and social attraction affect colony selection and dynamics of lesser kestrels. *Ecology*, 85, 3438–3447.

633. Serreze MC, Francis JA (2006) The Arctic amplification debate. *Climatic Change*, 76, 241–264.

634. Shackleton EH (1919) South: The Endurance Expedition.

635. Sheehan S, Harris K, Song YS (2013) Estimating variable effective population sizes from multiple genomes: a sequentially Markov conditional sampling distribution approach. *Genetics*, 194, 647–662.

636. Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature biotechnology*, 26, 1135–1145.
637. Shepherd LD, Lambert DM (2005) Mutational bias in penguin microsatellite DNA. *J Hered*, 96, 566–571.
638. Shirihai H, Kirwan GM (2008) Complete guide to Antarctic wildlife. *agris.fao.org*,
639. Shu Q, Song Z, Qiao F (2015) Assessment of sea ice simulations in the CMIP5 models. *The Cryosphere*, 9, 399–409.
640. Simeone A, Hiriart-Bertrand L, Reyes-Arriagada R et al. (2009) Heterospecific Pairing and Hybridization between Wild Humboldt and Magellanic Penguins in Southern Chile. *The Condor*, 111, 544–550.
641. Simeone A, Luna-Jorquera G, Bernal M et al. (2003) Breeding distribution and abundance of seabirds on islands off north-central Chile. *Revista Chilena de Historia Natural*, 76, 323–333.
642. Simmonds I, Jacka TH (1995) Relationships between the interannual variability of Antarctic sea ice and the Southern Oscillation. *Journal of Climate*, 8, 637–647.
643. Simpson GG (1975) Fossil penguins.
644. Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195, 693–702.
645. Slack KE, Jones CM, Ando T et al. (2006) Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution*, 23, 1144–1155.
646. Smagulova F, Gregoretto IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV (2011) Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature*, 472, 375–378.
647. Smetacek V, Nicol S (2005) Polar ocean ecosystems in a changing world. *Nature*, 437, 362–368.
648. Smith AL, Monteiro L, Hasegawa O, Friesen VL (2007) Global phylogeography of the band-rumped storm-petrel (*Oceanodroma castro*; Procellariiformes: Hydrobatidae). *Molecular phylogenetics and evolution*, 43, 755–773.
649. (2007) The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.
650. Song YS, Hein J (2005) Constructing minimal ancestral recombination graphs. *Journal of Computational Biology*, 12, 147–169.
651. Soong K, Lang JC (1992) Reproductive integration in reef corals. *The Biological Bulletin*, 183, 418–431.
652. Soulé ME, Estes JA, Berger J, Del Rio CM (2003) Ecological effectiveness: conservation goals for interactive species. *Conservation Biology*, 17, 1238–1250.
653. Staab PR, Zhu S, Metzler D, Lunter G (2015) scrm: efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics*, 31, 1680–1682.
654. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.
655. Stammerjohn SE, Martinson DG, Smith RC, Yuan X, Rind D (2008) Trends in Antarctic annual sea ice retreat and advance and their relation to El Niño–Southern Oscillation and Southern Annular Mode variability. *J. Geophys. Res.*, 113, 1–20.
656. Stearns SC (1992) The evolution of life histories.

657. Steele JH (1985) A comparison of terrestrial and marine ecological systems. *Nature*, 313, 355–358.
658. Steiner UK (2005) Reproductive consequences of natal dispersal in a highly philopatric seabird. *Behavioral Ecology*, 16, 634–639.
659. Stenseth NC, Durant JM, Fowler MS et al. (2015) Testing for effects of climate change on competitive relationships and co-existence between two bird species. *Proceedings of the Royal Society of London B: Biological Sciences*, 282, 20141958.
660. Stevens EL, Heckenberg G, Robertson ED, Baugher JD, Downey TJ, Pevsner J (2011) Inference of relationships in population data using identity-by-descent and identity-by-state. *PLoS Genet*, 7, e1002287.
661. Stocker T, Qin D, Plattner GK, Tignor M, Allen SK (2014) Climate change 2013: The physical science basis.
662. Stonehouse B (1960) The King penguin: *Aptenodytes patagonica*, of South Georgia.
663. Stonehouse B, Glenister TW (1953) The Emperor Penguin, *Aptenodytes Forsteri*.
664. Stuecker MF, Timmermann A, Jin F-F, McGregor S, Ren H-L (2013) A combination mode of the annual cycle and the El Niño/Southern Oscillation. *Nature Geoscience*, 6, 540–544.
665. Subramanian S, Lingala SG, Swaminathan S, Huynen L, Lambert D (2014) Second generation DNA sequencing of the mitogenome of the Chinstrap penguin and comparative genomics of Antarctic penguins. *Mitochondrial DNA*, 25, 271–272.
666. Subramanian S, Beans-Picón G, Swaminathan SK, Millar CD, Lambert DM (2013) Evidence for a recent origin of penguins. *Biology letters*, 9, 20130748.
667. Sydeman WJ, Thompson SA, Kitaysky A (2012) Seabirds and climate change: roadmap for the future. *Marine Ecology Progress Series*, 454, 107–117.
- T
668. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595.
669. Taschetto AS, Wainer I (2008) The impact of the subtropical South Atlantic SST on South American precipitation. *Annales geophysicae: atmospheres, hydrospheres and space sciences* 26, 3457.
670. Tasker ML, Camphuysen CJ, Cooper J, Garthe S, Montevecchi WA, Blaber SJM (2000) The impacts of fishing on marine birds. *ICES Journal of Marine Science: Journal du Conseil*, 57, 531–547.
671. Taylor CM, Norris DR (2007) Predicting conditions for migration: effects of density dependence and habitat quality. *Biology Letters*, 3, 280–284.
672. Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498.
673. Taylor SA, Zavalaga CB, Luna-Jorquera G, Simeone A, Anderson DJ, Friesen VL (2010) Panmixia and high genetic diversity in a Humboldt Current endemic, the Peruvian Booby (*Sula variegata*). *Journal of Ornithology*, 1–8.
674. Tchernia P, Jeannin PF (1980) Observations on the Antarctic East Wind Drift using tabular icebergs tracked by satellite Nimbus F (1975–1977). *Deep Sea Research Part A. Oceanographic Research Papers*, 27, 467–474.
675. Techow NMSM, O’Ryan C, Phillips RA et al. (2010) Speciation and phylogeography of giant petrels *Macronectes*. *Molecu-*

- lar phylogenetics and evolution, 54, 472–487.
676. Terauds A, Chown SL, Morgan F et al. (2012) Conservation biogeography of the Antarctic. *Diversity and Distributions*, 18, 726–741.
677. Thatje S, Hillenbrand C-D, Mackensen A, Larter R (2008) Life hung by a thread: endurance of Antarctic fauna in glacial periods. *Ecology*, 89, 682–692.
678. Thiebot J-B, Lescroël A, Barbraud C, Bost C-A (2013) Three-dimensional use of marine habitats by juvenile emperor penguins *Aptenodytes forsteri* during post-natal dispersal. *Antarctic Science*, 25, 536–544.
679. Thiel M, Macaya EC, Acuna E et al. (2007) The Humboldt Current System of northern and central Chile: oceanographic processes, ecological interactions and socioeconomic feedback. *Oceanography and Marine Biology*, 45, 195–344.
680. Thomas CD, Cameron A, Green RE et al. (2004) Extinction risk from climate change. *Nature*, 427, 145–148.
681. Thomas DN, Dieckmann GS (2008) Sea ice: an introduction to its physics, chemistry, biology and geology.
682. Thomas DB, Ksepka DT, Fordyce RE (2011) Penguin heat-retention structures evolved in a greenhouse Earth. *Biology Letters*, 7, 461–464.
683. Thouzeau C, Le Maho Y, Froget G et al. (2003) Spheniscins, avian β -defensins in preserved stomach contents of the king penguin, *Aptenodytes patagonicus*. *Journal of Biological Chemistry*, 278, 51053.
684. Thuiller W, Richardson DM, Py EK, Petr, Midgley GF, Hughes GO, Rouget M (2005) Niche based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, 11, 2234–2250.
685. Todd FS, Adie S, Splettstoesser JF (2004) First ground visit to the emperor penguin *Aptenodytes forsteri* colony at Snow Hill Island, Weddell Sea, Antarctica. *Marine Ornithology*, 32,
686. Toggweiler JR, Samuels B (1998) On the ocean's large-scale circulation near the limit of no vertical mixing. *Journal of Physical Oceanography*, 28, 1832–1852.
687. Tolstoy L (1869) War and Peace.
688. Tomczak M, Godfrey JS (2003) Regional oceanography: an introduction.
689. Trathan PN, Fretwell PT, Stonehouse B (2011) First recorded loss of an emperor penguin colony in the recent period of Antarctic regional warming: implications for other colonies. *PloS one*, 6, e14738.
690. Travis JMJ, Dytham C (1999) Habitat persistence, habitat availability and the evolution of dispersal. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266, 723–728.
691. Trivelpiece WZ, Hinke JT, Miller AK, Reiss CS, Trivelpiece SG, Watters GM (2011) Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. *Proceedings of the National Academy of Sciences*, 108, 7625.
692. Trucchi E, Facon B, Gratton P et al. (2015) Long live the alien: studying the fate of the genomic diversity along the long-term dynamics of an extremely successful invader, the crested porcupine. *bioRxiv*, 016493.
693. Trucchi E, Gratton P, Whittington JD et al. (2014) King penguin demography since the last glaciation inferred from genome-wide data. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20140528.
694. Turchin P (1998) Quantitative analysis of movement: measuring and modeling

population redistribution in animals and plants.

695. Turner J, Barrand NE, Bracegirdle TJ et al. (2014) Antarctic climate change and the environment: an update. *Polar Record*, 50, 237–259.

696. Turner J, Bracegirdle TJ, Phillips T, Marshall GJ, Hosking JS (2013) An initial assessment of Antarctic sea ice extent in the CMIP5 models. *Journal of Climate*, 26, 1473–1484.

697. Tynan CT (1998) Ecological importance of the southern boundary of the Antarctic Circumpolar Current. *Nature*, 392, 708–710.

V

698. Valverde Romero MB (2006) First record of the endangered Peruvian diving-petrel *Pelecanoides garnotii* breeding on Corcovado island, Peru. *Marine Ornithology*, 34, 75–76.

699. Van Buskirk J, Steiner UK (2009) The fitness costs of developmental canalization and plasticity. *Journal of evolutionary biology*, 22, 852–860.

700. Van der Putten N, Verbruggen C, Ochyra R, Verleyen E, Frenot Y (2010) Subantarctic flowering plants: pre glacial survivors or post glacial immigrants. *Journal of biogeography*, 37, 582–592.

701. Van Heerwaarden B, Hoffmann AA (2007) Global warming: fly populations are responding rapidly to climate change. *Current Biology*, 17, R16–R18.

702. VanDerWal J, Murphy HT, Kutt AS et al. (2013) Focus on poleward shifts in species' distribution underestimates the fingerprint of climate change. *Nature Clim. Change*, 3, 239–243.

703. Viblanc VA, Gineste B, Stier A, Robin J-P, Groscolas R (2014) Stress hor-

mones in relation to breeding status and territory location in colonial king penguin: a role for social density? *Oecologia*, 1–10.

704. Vilina YA (1992) Status of the peruvian diving petrel, *Pelecanoides garnotii*, on Chañaral Island, Chile. *Colonial Waterbirds*, 137–139.

705. Viot CR, Jouventin P, Bried J (1993) Population genetics of southern seabirds. *Marine Ornithology*, 21, 1–25.

706. Virkkala R, Heikkinen RK, Leikola N, Luoto M (2008) Projected large-scale range reductions of northern-boreal land bird species due to climate change. *Biological Conservation*, 141, 1343–1353.

707. Visser ME, Both C (2005) Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings of the Royal Society of London B: Biological Sciences*, 272, 2561–2569.

708. Visser ME, Van Noordwijk AJ, Tinbergen JM, Lessells CM (1998) Warmer springs lead to mistimed reproduction in great tits (*Parus major*). *Proceedings of the Royal Society of London B: Biological Sciences*, 265, 1867–1870.

709. Vogl C, Karhu A, Moran G, Savolainen O (2002) High resolution analysis of mating systems: inbreeding in natural populations of *Pinus radiata*. *Journal of Evolutionary Biology*, 15, 433–439.

710. Von Haartman L (1949) Der Trauerfliegenschnäpper. *Acta Zoologica Fennica*, 56, 104.

711. Vucetich JA, Waite TA (2000) Is one migrant per generation sufficient for the genetic management of fluctuating populations. *Animal Conservation*, 3, 261–266.

W

712. Wagner A (2008) Neutralism and selectionism: a network-based reconciliation. *Nature Reviews Genetics*, 9, 965–974.
713. Wagner EL, Boersma PD (2011) Effects of fisheries on seabird community ecology. *Reviews in Fisheries Science*, 19, 157–167.
714. Wahlund S (1928) Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, 11, 65–106.
715. Wainer I, Prado LF, Khodri M, Otto-Bliesner B (2014) Reconstruction of the South Atlantic Subtropical Dipole index for the past 12,000 years from surface temperature proxy. *Sci Rep*, 4, 5291.
716. Wall JD, Tang LF, Zerbe B et al. (2014) Estimating genotype error rates from high-coverage next-generation sequence data. *Genome research*, 24, 1734–1739.
717. Wallberg A, Han F, Wellhagen G et al. (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat Genet*, 46, 1081–1088.
718. Walther G-R, Post E, Convey P et al. (2002) Ecological responses to recent climate change. *Nature*, 416, 389–395.
719. Wang C, Zhang L, Lee S-K, Wu L, Mechoso CR (2014) A global perspective on CMIP5 climate model biases. *Nature Climate Change*, 4, 201–205.
720. Wang J (2011) COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol Ecol Resour*, 11, 141–145.
721. Wang J (2004) Application of the one migrant per generation rule to conservation and management. *Conservation Biology*, 18, 332–343.
722. Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89, 438–450.
723. Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology*, 15, 1419–1439.
724. Warnock RCM, Parham JF, Joyce WG, Lyson TR, Donoghue PCJ (2015) Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proceedings of the Royal Society of London B: Biological Sciences*, 282, 20141013.
725. Warnock RCM, Yang Z, Donoghue PCJ (2012) Exploring uncertainty in the calibration of the molecular clock. *Biology letters*, 8, 156–159.
726. Warren R, VanDerWal J, Price J et al. (2013) Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss. *Nature Clim. Change*, 3, 678–682.
727. Waser PM, Jones WT (1983) Natal philopatry among solitary mammals. *Quarterly review of biology*, 58, 355.
728. Waters JM (2008) Driven by the West Wind Drift? A synthesis of southern temperate marine biogeography, with new directions for dispersalism. *Journal of Biogeography*, 35, 417–427.
729. Watson JEM, Iwamura T, Butt N (2013) Mapping vulnerability and conservation adaptation strategies under climate change. *Nature Climate change*, 3, 989–994.
730. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical population biology*, 7, 256–276.
731. Weber ME, Clark PU, Kuhn G et al. (2014) Millennial-scale variability in

- Antarctic ice-sheet discharge during the last deglaciation. *Nature*, 510, 134–138.
732. Webster MT, Axelsson E, Ellegren H (2006) Strong regional biases in nucleotide substitution in the chicken genome. *Molecular biology and evolution*, 23, 1203–1216.
733. Weimerskirch H, Stahl JC, Jouventin P (1992) The breeding biology and population dynamics of king penguins *Aptenodytes patagonica* on the Crozet Islands. *Ibis*, 134, 107–117.
734. Weimerskirch H, Tarroux A, Chastel O, Delord K, Chérel Y, Descamps S (2015) Population-specific wintering distribution of adult south polar skua over three oceans. *Marine Ecology Progress series*, 538, 229–237.
735. Weimerskirch H (2013) Seabirds-in-dividuals in colonies. *Science*, 341, 35–36.
736. Weimerskirch H, Catard A, Prince PA, Chérel Y, Croxall JP (1999) Foraging white-chinned petrels (*Procellaria aequinoctialis*) at risk: from the tropics to Antarctica. *Biological Conservation*, 87, 273–275.
737. Weimerskirch H, Louzao M, De Grissac S, Delord K (2012) Changes in wind pattern alter albatross distribution and life-history traits. *Science*, 335, 211–214.
738. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *evolution*, 1358–1370.
739. Weir BS (1990) Genetic data analysis. Methods for discrete population genetic data.
740. Welch AJ, Yoshida AA, Fleischer RC (2011) Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. *Molecular Ecology*, 20, 1364–1377.
741. Westerhold T, Bickert T, Röhl U (2005) Middle to late Miocene oxygen isotope stratigraphy of ODP site 1085 (SE Atlantic): new constraints on Miocene climate variability and sea-level fluctuations. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 217, 205–222.
742. Wheelwright NT, Mauck RA (1998) Philopatry, natal dispersal, and inbreeding avoidance in an island population of Savannah Sparrows. *Ecology*, 79, 755–767.
743. White WB, Peterson RG (1996) An Antarctic circumpolar wave in surface pressure, wind, temperature and sea-ice extent. *Nature*, 380, 699–702.
744. Whitehead H, Rendell L (2014) The cultural lives of whales and dolphins.
745. Whitehead JM (2003) Pliocene summer sea surface temperature reconstruction using silicoflagellates from Southern Ocean ODP Site 1165. *Paleoceanography*, 18, 1–11.
746. Whitlatch RB, Lohrer AM, Thrush SF et al. (1998) Scale-dependent benthic recolonization dynamics: life stage-based dispersal and demographic consequences. In: *Recruitment, Colonization and Physical-Chemical Forcing in Marine Biological Systems*, pp. 217–226. Springer,
747. Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity*, 82, 117–125.
748. Wienecke B (2009a) Emperor penguin colonies in the Australian Antarctic Territory: how many are there. *Polar Record*, 45, 304–312.
749. Wienecke B (2009b) The history of the discovery of emperor penguin colonies. *Polar Record*, 46, 271–276.
750. Wienecke B (2010) The history of the discovery of emperor penguin colonies. *Polar Record*,
751. Wienecke B (2012) Emperor penguins at the West Ice Shelf. *Polar biology*, 35, 1289–1296.

752. Wienecke B, Robertson G (2002) Foraging areas of king penguins from Macquarie Island in relation to a marine protected area. *Environmental management*, 29, 662–672.
753. Wienecke BC, Lawless R, Rodary D et al. (2000) Adélie penguin foraging behaviour and krill abundance along the Wilkes and Adélie land coasts, Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, 47, 2573–2587.
754. Wiens JA, Stralberg D, Jongsomjit D, Howell CA, Snyder MA (2009) Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences*, 106, 19729–19736.
755. Wiens JJ, Ackerly DD, Allen AP et al. (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology letters*, 13, 1310–1324.
756. Williams JW, Jackson ST, Kutzbach JE (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences*, 104, 5738–5742.
757. Williams TD (1995) The penguins.
758. Willing EM, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, 7, e42649.
759. Wilson GJ (1983) Distribution and abundance of Antarctic and sub-Antarctic penguins: a synthesis of current knowledge.
760. Wilson PR, Ainley DG, Nur N et al. (2001) Adélie penguin population change in the pacific sector of Antarctica: relation to sea-ice extent and the Antarctic Circumpolar Current. *Marine ecology progress series*, 213, 301–309.
761. Winters JB, Waser PM (2003) Gene dispersal and outbreeding in a philopatric mammal. *Molecular Ecology*, 12, 2251–2259.
762. Wiuf C, Hein J (1999) Recombination as a point process along sequences. *Theoretical population biology*, 55, 248–259.
763. Woehler EJ (2001) A statistical assessment of the status and trends of Antarctic and Subantarctic seabirds.
764. Woehler EJ, Croxall JP (1997) The status and trends of Antarctic and sub-Antarctic seabirds. *Marine Ornithology*, 25, 43–66.
765. Wolff EW, Rankin AM, Röthlisberger R (2003) An ice core indicator of Antarctic sea ice production. *Geophysical Research Letters*, 30,
766. Wolff EW, Fischer H, Fundel F et al. (2006) Southern Ocean sea-ice extent, productivity and iron flux over the past eight glacial cycles. *Nature*, 440, 491–496.
767. Wright S (1931) Evolution in Mendelian populations. *Genetics*, 16, 97–159.
768. Wright S (1943) Isolation by distance. *Genetics*, 28, 114.
769. Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics*, 31, 39.
770. Wright S (1950) Genetical structure of populations. *Nature*, 166, 247–249.
771. Wright S (1965) The interpretation of population structure by F -statistics with special regard to systems of mating. *Evolution*, 19, 395–420.
772. Wright S (1978) Evolution and the Genetics of Populations: A Treatise in Four Volumes.

X

773. Xu S, Song M, Liu J et al. (2013) Simulation of sea ice in FGOALS-g2: Climatology and late 20th century changes. *Advances in Atmospheric Sciences*, 30, 658–673.

Y

774. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24, 1586–1591.

775. Yeakel JD, Guimarães PR, Bocherens H, Koch PL (2013) The impact of climate change on the structure of Pleistocene food webs across the mammoth steppe. *Proceedings of the Royal Society of London B: Biological Sciences*, 280, 20130239.

776. Younger JL, Clucas GV, Kooyman G et al. (2015) Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. *Glob Chang Biol*, 21, 2215–2226.

Z

777. Zarnetske PL, Skelly DK, Urban MC (2012) Biotic multipliers of climate change. *Science*, 336, 1516–1518.

778. Zavalaga CB, Alfaro-Shigueto J, Dell Omo G (2010) First GPS-track of Peruvian diving-petrels and inca terns in Southern Peru.

779. Zavalaga CB, Jahncke J (1997) Maximum dive depths of the Peruvian diving-petrel. *Condor*, 1002–1004.

780. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18, 821–829.

781. Zhang G, Lambert D, Wang J, (2011a) Genomic data from the Adélie penguin (*Pygoscelis adeliae*). *10K Genome Project*.

782. Zhang G, Lambert D, Wang J, (2011b) Genomic data from the Emperor penguin (*Aptenodytes forsteri*). *10K Genome Project*.

783. Zhang G, Li C, Li Q et al. (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346, 1311–1320.

784. Zhang J (2007) Increasing Antarctic Sea Ice under Warming Atmospheric and Oceanic Conditions. *J. Climate*, 20, 2515–2529.

785. Zhang Y-X, Perry K, Vinci VA, Powell K, Stemmer WPC, del Cardayré SB (2002) Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature*, 415, 644–646.

786. Zhao Z, Boerwinkle E (2002) Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. *Genome research*, 12, 1679–1686.

787. Zhou Q, Zhang J, Bachtrog D et al. (2014) Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*, 346, 1246338.

788. Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, 17, 2107–2121.

1 Annexe 5 : Introduction (traduction française)

2 *Populations, philpatrie et dispersion.*

3 *§-1 Les espèces: type et répétition.* La description systématique des espèces peut être conçue comme
4 un cas extrême de médiation de l'*individual* et du *collectif*. Un échantillon singulier, l'*holotype*, est
5 d'ordinaire choisi pour référence, et légitime seul l'attribution du nom de l'espèce à chaque autre
6 specimen - au delà de cet holotype, le reste n'est que *répétition*, dans le temps et dans l'espace.
7 Mais aucune répétition n'est parfaite (Kierkegaard 1843) : la variation, ou *diversité*, joue un rôle
8 essentiel tant dans le temps que dans l'espace (Mayr 1982). La variabilité phénotypique intraspé-
9 cifique entre dans le champs de la biologie sous le nom de « *différences individuelles* » dans l'œuvre
10 de Darwin, célèbre pour y avoir trouvé « *le matériau sur lequel la sélection naturelle agit et s'accu-*
11 *mule* » (Darwin 1859): les fondements génétiques de cette diversité sont désormais au centre de la
12 *théorie synthétique de l'évolution* (Mayr & Provine 1998; Seehausen *et al.* 2014). Cette variation
13 ne sert aucun but prédéfini : elle peut pourtant être *a posteriori* organisée le long d'un spectre, qui
14 s'étend de la diversité *adaptative* à la diversité *neutre* (Kimura 1983; Wagner 2008). A l'échelle
15 moléculaire, la variation adaptative se définit comme l'apparition de mutations aux effets phéno-
16 typiques avantageux, qui sont positivement sélectionnés dans la population (Mayr 1963), tandis
17 que la variation neutre est faite de mutations ne semblant offrir aucune prise à la sélection na-
18 turelle (Kimura 1983). Nous ne prendrons pas position quant à l'importance relative de la varia-
19 tion neutre et adaptative dans l'évolution moléculaire (un sujet au cœur de continuels débats,

1 voir par ex. Wagner 2008) : mais c'est essentiellement sur la variation neutre que s'appuiera notre
2 travail, car elle offre une perspective inégalée sur les processus écologiques non-moléculaires.

3 L'organisation de la variation moléculaire dans l'espace et le temps est au coeur de la génétique
4 des populations (Wright 1978; Kingman 2000) - une discipline que l'on pourrait définir comme
5 l'étude de la diversité génétique à l'intérieur des limites spatio-temporelles conventionnelles de
6 l'espèce - cette « *combinaison tout artificielle, faite par commodité* » (Darwin 1859). L'ampleur et
7 l'organisation de cette diversité reflètent certains caractères essentiels de l'espèce : la diversité
8 génétique elle-même a été expliquée par plusieurs paramètres, tels que la complexité de l'organ-
9 isme (Lynch & Conery 2003), la taille et l'histoire de ses populations (Amos & Harwood 1998),
10 ou sa biodémographie (Romiguier *et al.* 2014) - quant à son organisation dans le temps et l'es-
11 pace, elle est avant tout le fruit de l'histoire démographique et migratoire de l'espèce (un point
12 abordé en détail au §41). Ainsi l'étude des espèces en tant que système de populations nous
13 permet-elle de saisir non seulement leur inhérente diversité taxonomique, mais aussi la variété de
14 leurs dynamiques écologiques et évolutives.

15 Un aspect remarquable (et lourd de conséquences méthodologiques) de la façon dont la géné-
16 tique des populations aborde la diversité intraspécifique est - pour ce qui est du moins des méta-
17 zoaires - la place nécessairement ambivalente qu'y occupe l'*individu*. La génétique des popula-
18 tions reste prétrie de l'idée d'une irréductible singularité : comme l'écrit Ernst Mayr, « *où que nous*
19 *regardions, nous trouvons l'unique, et l'unique est diversité* » (Mayr 1982). L'individu, en tant qu'il
20 est combinaison singulière (et vraisemblablement unique) d'allèles, est la seule forme sur laquelle
21 peut s'opérer l'évaluation des *valeurs sélectives*¹, et, de ce fait, la seule forme *efficace* de la diversité
22 génétique. En ce sens, une population, ou *ressource génétique*², n'est rien d'autre que la représenta-
23 tion abstraite de ce qui n'est, en réalité, qu'un assemblage d'individus singuliers. Et bien que les
24 concepts d'espèce et de population (voir §2) soient des outils essentiels à la compréhension des
25 processus comportementaux, démographiques ou évolutifs, la seule unité *atomique* observable et

1. Nous traduisons ainsi d'anglais *fitness*.

2. Selon la traduction de l'anglais *gene pool* par la Convention sur la Diversité Biologique.

1 non-interprétée de la vie des animaux demeure l'individu. Dès lors, l'on comprend que toute
2 étude empirique qui d'appuie sur l'échantillonnage de groupes ou de populations ne fait, en réal-
3 ité, qu'assembler une théorie d'individus uniques : leur attribution à tel ou tel groupe n'est qu'une
4 décision postérieure de l'observateur (ce qui pose entre autres de réelles questions pour la con-
5 ception d'expériences - voir entre autres *Fine scale structure*, p.135).

6 Et pourtant, l'*individu* se dissout soit *verticalement*, soit *transversalement*, dans toute la théorie de
7 la génétique des populations. La notion essentielle de *fréquence allélique* (voir §45), par exemple,
8 n'existe que par l'analyse transversale (c'est à dire, à un point donné du temps) d'un grand nom-
9 bre de chromosomes, à une position génomique donnée, sans plus retenir aucune information
10 quant aux associations particulières des allèles au sein des individus : on peut ainsi la concevoir
11 comme l'abolition des frontières *synchroniques* entre individus - une idée qui trouve d'ailleurs sa
12 consécration mathématique dans l'approximation de la dérive génétique par les équations de
13 diffusion (voir §39). La théorie du *coalescent* (voir §40), quant à elle, se fonde sur l'analyse verti-
14 cale (c'est à dire transgénérationnelle) des séquences : dès lors, l'avènement et la disparition de
15 l'individu (ou la *génération*) devient un accident, et une simple mesure du temps. Ainsi la notion
16 même de coalescence réclame l'abolition des frontières *diachroniques* entre individus : un concept
17 que formalise la représentation en temps continu du coalescent, qui écarte totalement les notions
18 d'individu ou de génération (voir §45).

19 Cette représentation paradoxale du singulier dans le cadre de la génétique des populations forme
20 un contraste frappant avec la forme que prend d'ordinaire l'individu tant dans le cadre de la sys-
21 tématique (où la singularité est réduite à la dispersion autour de l'holotype dans l'espace
22 morphologique), que dans celui de l'écologie (où la diversité est soit remplacée par des paramètres
23 démographiques homogènes tels que la densité - voir Bolnick *et al.* 2011 - ou au contraire de-
24 vient le point focal de l'étude des choix d'histoire de vie - voir Le Bohec 2008; Weimerskirch
25 2013). L'un des défis de la théorie synthétique de l'évolution reste donc la réconciliation de ces
26 différents paradigmes, en dépit de la divergence de leurs fondations. En particulier, la différence
27 entre les concepts de *taille de population* ou *taux de migration* dans le paradigme écologique et

1 celui de la génétique des populations (voir §41 et Fig. 1) rend ardue l'intégration de sources in-
2 terdisciplinaire (voir *The Emperor synnome* p. 211). Non moins ardue est la réconciliation des
3 différentes échelles de temps dans lesquelles ces paradigmes s'inscrivent: par bien des aspects,
4 la génétique des populations occupe une position intermédiaire entre la systématique (qui opère à
5 l'échelle «géologique» des spéciations) et l'écologie (qui s'occupe de la vie discrète des individus).
6 Si la synthèse de ces différents modes dépasse le cadre de ce travail, les questions qui nous occu-
7 pent ici, parce qu'elles sont au croisement de l'écologie et de la génétique des populations, pren-
8 nent tout leur sens à travers cette inévitable ambivalence.

9 *§-2 Population, dème, colonie.* Tout comme la distribution des positions individuelles dans les es-
10 paces morphologiques, comportementaux ou génétiques définit la forme de la diversité spatiale et
11 temporelle d'une espèce, l'étendue du recouvrement entre la zone de mouvement d'un individu,
12 et la répartition totale de l'espèce définit un spectre de structures spatiales allant de formes ho-
13 mogènes et *panmictiques* à des ensembles hétérogènes et fragmentés. L'architecture particulière
14 d'une espèce intègre des processus qui se déploient à des échelles très diverses : et sa description
15 change selon le cadre théorique dans lequel l'observateur se place (Esler *et al.* 2006). La définition
16 d'une *population* - un concept pourtant central en génétique des populations - change radicale-
17 ment selon qu'on se positionne dans le paradigme écologique ou génétique (Waples & Gaggiotti
18 2006). Dans le paradigme écologique, les critères retenus sont d'ordinaire la géographie (par ex-
19 emple la co-occurrence d'individus *au même endroit*, voir Krebs 2013), les interactions entre indi-
20 vidus (comme le partage de ressources en nourriture, voir Huffaker & Gutierrez 1984), ou, de
21 façon plus importante, l'indépendance démographique des groupes étudiés (c'est à dire, si la mi-
22 gration entre ces groupes suffit à les ramener à une unique trajectoire démographique commune,
23 ou à influencer le risque d'extinction d'un de ces groupes - voir McElhany *et al.* 2000). Waples
24 and Gaggiotti (2006) résument de la façon suivante la définition écologique : « *un groupe d'indi-*
25 *vidus de la même espèce qui se trouvent au même point de l'espace et du temps, et ont l'occasion d'in-*
26 *teragir les uns avec les autres* » - une définition qui se base sur l'étude de la distribution, la démo-

1 graphie, et le comportement d'une espèce à une échelle de temps observable - c'est à dire du
2 même ordre de grandeur que la vie de l'observateur.

3 Dans le paradigme génétique (ou « évolutionniste » - voir Waples & Gaggiotti 2006), une popula-
4 tion est définie comme une communauté reproductrice, c'est à dire comme un groupe d'indi-
5 vidus qui ont une plus grande probabilité de se reproduire entre eux, plutôt qu'avec un individu
6 d'un autre groupe (Hartl *et al.* 1988), soit, par conséquent, un groupe d'individus dont les
7 génotypes sont corrélés. L'aspect quantitatif de la corrélation minimale nécessaire pour obtenir le
8 statut de « population » est quant à lui matière à débat : mais la plupart des indicateurs retenus se
9 fondent sur la comparaison des influences relatives de la migration et de la dérive dans l'évolution
10 de la diversité génétique d'un groupe (voir §39 et §41). Par exemple, un seuil d'un *migrant effi-*
11 *cace* par génération a été proposé dans plusieurs études empiriques (voir Mills & Allendorf 1996;
12 Vucetich & Waite 2000; Wang 2004) : si deux groupes échangent plus d'un migrant efficace par
13 génération, ils ne peuvent être considérés séparés. Des seuils moins restrictifs ont été proposés -
14 par exemple un maximum de 5 ou de 25 migrants efficaces (Waples & Gaggiotti 2006 en don-
15 nent un résumé). Puisque ces critères se concentrent sur des paramètres évalués par générations,
16 ils s'appliquent uniquement en moyenne, sur une fenêtre de plusieurs générations - c'est à dire, à
17 une échelle bien plus longue que la définition écologique.

18 La relation entre les concepts de population dans le paradigme écologique et évolutionniste n'est
19 pas simple. Quoiqu'il puisse sembler que l'isolation comportementale et géographique d'un
20 groupe d'individus doive naturellement mener à une réduction du flux génique et à l'isolation
21 génétique, le passage d'échelles de temps courtes à des échelles plus longues n'est pas nécessaire-
22 ment direct : la cohérence des groupes peut être hétérogène dans le temps, et donner lieu à l'ap-
23 parente contradiction de groupes très structurés d'un point de vue comportemental, mais par-
24 faitement mélangés d'un point de vue génétique (c'est par exemple le cas de l'Arlequin plongeur
25 *Histrionicus histrionicus*, voir Esler *et al.* 2006, ou de l'Albatros hurleur *Diomedea exulans*, voir
26 Milot *et al.* 2008). Pour distinguer clairement ces différents paradigmes, dans cette étude, nous

1 aurons recours à trois termes apparentés, mais non équivalents : le *dème*, la *population*, et la
2 *colonie*.

3 Un *dème* répond à la définition génétique d'un groupe d'individus. On pourrait le décrire comme
4 « *le plus grand espace, ou le plus grand ensemble d'individus, à l'intérieur duquel la reproduction est,*
5 *en moyenne, aléatoire* » (Hamilton 2011). Cette définition n'inclut pas nécessairement de com-
6 posante spatiale, puisque plusieurs groupes d'individus se reproduisant de façon homogène peu-
7 vent co-exister en sympatrie (par exemple les orques *Orcinus orca*, voir Riesch *et al.* 2012). Par
8 ailleurs, un *dème* n'a pas nécessairement de frontières (spatiales ou comportementales), et peut
9 être défini comme une «fenêtre coulissante», comme c'est typiquement le cas des systèmes
10 d'« isolation par la distance », où chaque individu est le centre d'un groupe ou «voisinage géné-
11 tique » (Crawford 1984), à l'intérieur duquel les fréquences alléliques sont fortement corrélées, et
12 qui se dégrade progressivement avec la distance. Dans ce cas, le *dème* est l'approximation spatiale
13 d'un plus vaste système d'autocorrélation (Double *et al.* 2005; Epperson 2005). Un seuil de 5%
14 sur la probabilité de se reproduire a été proposé pour fixer les limites d'un *dème* (Hamilton 2011)
15 - mais ici encore, ce seuil est arbitraire, et le concept fait surtout sens pour des espèces à distribu-
16 tion discrète, avec de fortes discontinuités dans leur distribution de probabilité de reproduction.
17 Ainsi, lorsque l'on définit un *dème*, considère-t-on *a priori* que dans l'ensemble en considération,
18 la mutation et la dérive, et non la migration, sont les principaux garants de la diversité génétique.

19 Une *colonie* - et tout particulièrement pour les oiseaux de mer, voir §18 - est le pendant con-
20 ceptuel du *dème* : il s'agit en effet d'un concept purement géographique et comportemental, qui
21 décrit l'aggrégation d'individus reproducteurs pendant une période de temps donnée, et ne pose
22 aucune hypothèse quant aux structures génétiques ou processus évolutifs sous-jacents. Le terme
23 est couramment utilisé pour de nombreux organismes, allant du polype (par exemple les coraux
24 ou les siphonophores) aux insectes (comme les fourmis, voir Giraud *et al.* 2002) - avec, à chaque
25 fois, des caractéristiques différentes (en particulier en ce qui concerne l'intégration fonctionnelle,
26 qui est extrême chez les siphonophores - voir Dunn 2005, mais minime chez les coraux - voir
27 Soong & Lang 1992). La colonialité est aussi très répandue chez les vertébrés terrestres, qu'il

1 s'agisse des mammifères (comme le spermophile, voir Armitage 1981; Johnson 1981), ou des
2 oiseaux (voir Rolland *et al.* 1998). La colonialité de certains oiseaux s'étend à tous les aspects de
3 leur vie (comme c'est le cas de l'étourneau *Sturnus vulgaris*), mais dans la plupart des cas, elle se
4 limite à la période reproductrice (Lack 1968). C'est en particulier le cas des oiseaux de mer (voir
5 §18) qui passent le plus clair de leur temps à s'alimenter au large, en petits groupes, mais se
6 rassemblent en grandes colonies sur la terre ferme au début de la saison de reproduction. Para-
7 doxalement, la plupart des oiseaux de mer a une forte tendance à la philopatrie (§5) : bien que
8 chaque oiseau vagabonde au large durant ses voyages d'alimentation, et ne revient, de fait, jamais
9 à la côte, tous gardent un lien intérieur avec une parcelle de terre bien particulière, et y revien-
10 nent régulièrement pour s'y reproduire. Pour certaines espèces dont l'espace d'alimentation est
11 limité, cette a-parcelle est souvent la seule terre disponible (c'est par exemple souvent le cas du
12 Guillemot à Miroir *Cepphus grylle*, Ewins 1986). Mais pour les espèces qui voyagent loin au large,
13 comme les grands Albatros (*Diomedea sp.*) ou les grands Manchots (*Aptenodytes sp.*), la colonialité
14 devient une idée tout abstraite : la zone d'alimentation entre saisons de reproduction recouvre
15 largement celle d'autres colonies (et peut d'ailleurs, dans certains cas, s'étendre tout autour du
16 monde - voir Croxall *et al.* 2005), et pourtant chaque oiseau revient avec une étonnante régular-
17 ité à son lieu de naissance. L'apprentissage individuel, et le transfert horizontal, ou *culturel* (*sensu*
18 Whitehead & Rendell 2014) d'information entre oiseaux qui se reproduisent au même endroit
19 peut même donner naissance à des idiosyncrasies locales - par exemple différentes zones d'alimen-
20 tation pour des colonies pourtant proches (Weimerskirch 2013). Ainsi la colonie peut-elle être
21 conçue comme la représentation la plus extrême du concept de population dans le paradigme
22 écologique : c'est une aggrégation d'individus, stable dans le temps comme dans l'espace, que l'on
23 peut aisément observer, définir et délimiter (quoique non sans une série de parti-pris conceptuels,
24 voir §41) - mais sans caractéristiques génétiques particulières.

25 Notre emploi du mot *population*, dans ce travail, se trouve à l'intersection du *dème* et de la
26 *colonie*. Une population y sera définie comme un ensemble délimité à la fois dans les espaces géo-
27 graphiques, démographiques, et génétiques - en d'autres termes une colonie, ou un groupe de

1 colonies, constituant un dème. Il y a là plus en effet qu'une coïncidence anecdotique : parler de
2 *population* implique que les processus observés à l'échelle écologique (par exemple le regroupe-
3 ment spatial d'individus, l'utilisation de ressources communes, ou la philopatrie) sont assez puis-
4 sants et assez stables dans le temps pour influencer les mécanismes évolutifs - de sorte que les par-
5 adigmes évolutifs et génétiques en viennent à se recouper. Ainsi, à l'opposé d'un dème, notre
6 définition d'une population a une composante spatiale : et contrairement à la colonie, elle de-
7 mande aussi une composante génétique.

8 Ainsi deux forces antagonistes interagissent-elles dans la définition de la diversité des architectures
9 possibles des espèces : la *dispersion* (ou capacité active ou passive qu'ont les individus à se déplacer
10 à travers leur habitat), et la *philopatrie*, ou tendance qu'a un individu à demeurer ou revenir là où
11 il est né (lorsqu'elle est absolument passive, cette tendance s'appelle *inertie*). Ensemble, ces deux
12 forces déterminent la *fluidité* et la *viscosité* d'une espèce.

13 §-3 *Qu'est-ce qu'un Synnome ?* L'architecture particulière que nous retrouvons chez les deux es-
14 pèces du genre *Aptenodytes* échappent à tout concept déjà défini. Comme nous le présenterons
15 par la suite en détails (voir en particulier *The King synnome*, et *The Emperor synnome*), les deux
16 grands manchots nous forcent à opérer une médiation difficile entre la définition de population
17 dans les paradigmes écologiques et génétiques. La structure coloniale est poussée à l'extrême, faite
18 de concentrations locales extrêmement discontinues, et qui semblent stables à travers le temps - du
19 moins à l'échelle des siècles. La philopatrie est forte (§19) et l'on considère souvent qu'elle ne se
20 limite pas à la colonie, mais jusqu'au lieu de naissance précis à l'intérieur de la colonie (voir *Fine*
21 *scale structure*). Les colonies font montre en apparence d'un fort degré d'indépendance démo-
22 graphique - comme semblent l'indiquer les trajectoires divergentes des différentes colonies de l'île
23 de la Possession, dans l'Archipel des Crozets (voir Delord *et al.* 2004). Cependant, aucun des
24 critères génétiques n'est satisfait : la fluidité de l'espèce, sur le long terme, semble contrecarrer
25 toute trace de structure génétique, et la migration apparaît comme une force majeure dans l'évo-
26 lution de l'espèce. Une telle structure se distingue d'un système de métapopulations (Hanski
27 1998), dans lequel les populations sont, dans une certaine mesure, séparées génétiquement - mais

1 il se distingue aussi d'un système panmictique, car l'extrême discontinuité dans la distribution des
 2 individus et leur forte philopatrie contredisent l'hypothèse de la reproduction aléatoire à l'échelle
 3 d'une génération. Quoique de nombreuses recherches soient encore nécessaires afin d'en com-
 4 prendre els mécanismes sous-jacents, il est possible que des processus démographiques
 5 hétérogènes soient impliqués : par exemple, une dispersion pulsatile, qui impliquerait des événe-
 6 ments d'exode local catastrophiques durant certaines générations, équilibrés par des générations
 7 où la philopatrie totale serait de règle (voir *Empirical evidence of heterogeneous dispersal* pour un
 8 possible exemple de ce phénomène). Afin de décrire cette structure paradoxale, dans cette étude,
 9 nous proposons le terme *synnome*, que nous dérivons du grec σύννομος, une «*pâturage commune*» -
 10 qui a en particulier été utilisé pour décrire les vols d'oiseaux¹, et de façon plus générale pour
 11 désigner les *réunions*, les *rassemblements*, et même la *parentèle* (vois §123). Le concept central de *pâ-*
 12 *ture commune* rend compte avec précision de l'importance de la prospection alimentaire cen-
 13 tripète² dans la structure coloniale de l'espèce, tout en la distinguant d'un système de métapopu-
 14 lations fragmentées. Ces différentes significations, prises ensemble, capturent la particularité du
 15 système pbservé chez les grands manchots, que nous définirons ici comme «*un ensemble d'indi-*
 16 *vidus unique et homogène, distribué de façon extrêmement hétérogène et discontinue dans son aire de ré-*
 17 *partition. Les concentrations locales, ou colonies, sont très cohérentes à l'échelle de quelques générations,*
 18 *et à cette échelle la philopatrie est la norme. Pourtant, la migration entre ces zones est assez forte pour*
 19 *maintenir un parfaite homogénéité dans les ressources génétiques de l'espèce, de sorte qu'à une échelle*
 20 *micro-évolutive, la seule unité pertinente est l'espèce entière*».

21 *§-4 Dispersion et migration.* Comme nous l'avons mentionné plus haut, une population peut être
 22 définie dans deux paradigmes divergents - l'un écologique, l'autre génétique (voir §2). Dans ces
 23 deux paradigmes, une population est un ensemble *délimité* - géographiquement ou génétique-
 24 ment, elle est définie par son aspect autarcique. Toute forme de mouvement entre groupes œuvre
 25 à l'encontre de cette autarcie. Démographiquement, le mouvement couple les trajectoires des

1. Voir en particulier *Les Oiseaux* d'Aristophane, v. 1755 sqq.

2. A défaut d'une meilleure traduction pour l'expression *central place foraging*.

1 populations, et contrecarre l'évolution de l'indépendance démographique. Génétiquement, le
2 mouvement cause le transfert d'allèles entre dèmes, et travaille à homogénéiser la variation géné-
3 tique et les fréquences alléliques. Cependant, la distinction entre les paradigmes écologiques et
4 génétiques s'applique aussi aux mouvements entre groupes - là encore, cette distinction a motivé
5 le choix de termes différents. Une difficulté particulière, et qu'il s'agit de clarifier d'emblée, est le
6 choix convergent du mot « migration » pour désigner deux phénomènes très différents.

7 Le sens le plus courant du mot « migration », et ce tout particulièrement dans le contexte de la bi-
8 ologie des oiseaux, est le mouvement pendulaire saisonnier, souvent de colonies entières, entre
9 des terrains d'hiver et des terrains de reproduction estivale. Les espèces migratrices, en ce sens,
10 ont développé des adaptations particulières (voir par ex. Berthold 1991) qui leur permettent de
11 synchroniser leurs trajets, afin de suivre tout au long de l'année les environnements les plus favor-
12 ables. Il est à noter que le comportement de migration pose un défi pour la définition des popula-
13 tions selon le paradigme écologique, puisque chaque groupe peut changer ses frontières démo-
14 graphiques et géographiques entre ses aires d'hiver et d'été. Cependant, la migration entendue en
15 ce sens ne sera pas examinée dans notre étude, tout simplement parce qu'elle n'est pas observable
16 de façon classique dans notre système (quoique les voyages de prospection alimentaires des man-
17 chots, entre leurs périodes de reproduction, puisse être comparés à des migrations).

18 Un mouvement plus simple, et unidirectionnel, est la *dispersion* des individus qui quittent leur
19 groupe d'origine. La dispersion fut originellement décrite par Howard (1960) comme « *le mouve-
20 ment que fait un animal de son point de naissance au point où il se reproduit* ». Du point de vue de
21 chaque individu, c'est donc « *la distance maximale à laquelle il transmet ses caractéristiques géné-
22 tiques, plutôt que la distance maximale qu'il lui arrivera de parcourir au cours de sa vie depuis son
23 lieu de naissance* ». La dispersion peut être définie tant dans le paradigme écologique que dans le
24 paradigme génétique. Écologiquement, la dispersion est la force qui pousse au couplage démo-
25 graphique des populations - un excès de juvéniles à un endroit peut par exemple compenser un

1 faible recrutement ailleurs. Génétiquement, un événement de dispersion est «l'unité» du flux
2 génique - chaque individu dispersant apporte ses allèles d'un lieu à un autre.

3 Le second sens de *migration*, que nous utiliserons dans ce travail, a été défini par Dingle & Drake
4 (2007) dans un contexte biogéographique comme «*l'expansion de la distribution des faunes et des*
5 *espèces*», comme «*l'extension vers le Nord des aires de répartition après la retraite des glaciers à la fin*
6 *des périodes glaciaires*». Plus spécifiquement, dans le contexte de la génétique des populations, le
7 paramètre migratoire M est défini par les mêmes auteurs comme «*l'échange de gènes entre popula-*
8 *tions par n'importe quel moyen, incluant, mais de façon non exclusive, la migration au sens où nous*
9 *l'entendons ici*». C'est en ce sens que le terme est utilisé dans le cadre de la théorie du coalescent
10 par Beerli et al. (Beerli 2004; Beerli & Palczewski 2010, voir aussi §41). En ce sens, la migration
11 est la moyenne, sur le long terme, de la fréquence des événements de dispersion. Et tandis que la
12 dispersion est un phénomène à l'échelle de l'individu et de la génération, qui peut être directe-
13 ment observé, la migration est une moyenne dans le temps, à l'échelle des populations, et unique-
14 ment observable par des méthodes indirectes, comme par exemple la reconstruction du flux
15 génique. Ainsi, tandis que la dispersion appartient à la théorie des traits d'histoire de vie, et est
16 variable selon les caractéristiques de l'individu et du moment, la migration est une caractéristique
17 structurelle du système étudié, et devrait varier lentement avec les changements d'organisation de
18 l'espèce.

19 §-5 *Philopatrie et fidélité au site*. La philopatrie est définie pour la première fois sous le nom
20 d'*Ortstreue* (ou «fidélité au lieu») par von Haartman (1949), comme la tendance qu'à le Gobe-
21 mouche à collier *Ficedula hypoleuca* à se reproduire près de son lieu de naissance. Le concept est
22 généralisé comme *philopatrie* par Huntington (1951) comme «*la tendance qu'à un animal à*
23 *revenir à son lieu de naissance ou de reproduction*». Ainsi, quoique le terme est été occasionnelle-
24 ment employé pour décrire la pure fidélité au lieu de nidification (Anderson *et al.* 1992), il décrit
25 au sens strict la fidélité des adultes à leur propre lieu de naissance, ou «philopatrie natale» - et
26 c'est ce sens premier que nous emploierons ici (Pearce 2007). La philopatrie *stricto sensu* a des
27 conséquences importantes sur les processus génétiques qui informent l'architecture des espèces.

1 La tendance à se reproduire près de son lieu de naissance contredit l'idée de panmixie, puisque la
2 probabilité de se reproduire, pour deux individus choisis au hasard, devient inversement propor-
3 tionnelle à leur distance à la naissance. A un faible degré, ceci est à l'origine d'isolation par la dis-
4 tance : la parenté entre deux individus est fonction de leur distance géographique (Wright 1943;
5 Wright 1946). A plus forte indensité, la philopatrie peut mener à la consanguinité et à la frag-
6 mentation des populations (Mayr 1963; Avise *et al.* 2000). Les bénéfices de la philopatrie sont
7 pour autant considérables. Ces bénéfices sont comportementaux - comme une meilleure connais-
8 sance des sites et partenaires de meilleure qualité (Wheelwright & Mauck 1998; Heg *et al.* 2011;
9 Arnaud *et al.* 2012), une meilleure valeur sélective d'un comportement défensif de proximité
10 (Dunford 1977) ou de l'alloparentalité (Lecomte *et al.* 2006) par sélection de parentèle. Mais ils
11 sont aussi génétiques : par exemple, la promotion de la microadaptation locale (Richardson *et al.*
12 2014). Un comportement philopatric est à la base de la colonialité (Bowler & Benton 2005).
13 Cependant, là encore, les échelles de temps génétiques et écologiques divergent. A une échelle
14 écologique, le taux de retour moyen des juvéniles à leur colonie natale est d'ordinaire considéré
15 comme un estimateur de la philopatrie. Chez plusieurs espèces, en particulier d'oiseaux de mer,
16 les très forts taux de retour que l'on observe semblent incohérents avec le manque de structure
17 génétique entre colonies - le soi-disant *paradoxe des oiseaux marins* de Milot *et al.* (2008). Cepen-
18 dant, des études *in silico* ont montré que de faibles taux de migration (de 5 à 10%) suffisaient à
19 contrecarrer entièrement les effets de la dérive génétique locale à de plus longues échelles de
20 temps (Waples & Gaggiotti 2006). De plus, la notion que la philopatrie est un processus station-
21 naire chez les espèces coloniales, et que le taux de retour moyen doit, à long term, converger vers
22 la philopatrie générale de l'espèce ne prend pas en compte la possibilité d'événements de disper-
23 sion pulsatile, susceptibles de redistribuer entièrement les ressources génétiques de l'espèce. La
24 philopatrie, par opposition à la probabilité instantanée de fidélité au site (F) ou de dispersion (1-
25 F), n'est donc pas un paramètre directement estimable (Kendall & Nichols 2004; Pearce 2007),
26 mais plutôt un trait comportemental de long terme de l'espèce, qui peut, ou pas, amener l'indi-
27 vidu à un choix de fidélité au site, en fonction des conditions environnementales prévalentes.

1 §-6 *Philopatry et dispersion dans les systèmes océaniques*. En ce qui concerne la dispersion, la dy-
2 namique des structures océaniques diffère radicalement des structures terrestres (Steele 1985).
3 Bien qu'aux plus bas niveaux du système trophique, la dispersion passive sur de longues distances
4 soit également prévalente dans les écosystèmes marins et terrestres (les courants atmosphériques y
5 jouent en effet le même rôle que les courants marins pour la dispersion des graines et propagules -
6 voir Cain *et al.* 2000; Nathan & Muller-Landau 2000; Cowie & Holland 2006; Nathan 2006;
7 Nathan *et al.* 2008; Nikula *et al.* 2013), les différences d'organisation spatiale deviennent évi-
8 dentes pour les organismes de plus grande taille, en particulier les vertébrés (Carr *et al.* 2003). Les
9 écosystèmes terrestres montrent de façon typique un plus fort degré de fragmentation, moins de
10 dispersion et de flux génique, et des systèmes globalement plus fermés (Waser & Jones 1983;
11 Turchin 1998; Carr *et al.* 2003) : l'interaction entre les gradients latitudinaux et altitudinaux a
12 pour résultat une distribution complexe des habitats, en contraste avec les habitats pélagiques
13 volontiers zonaux (Burrows *et al.* 2014). Par ailleurs, les océans peuvent être considérés comme
14 un *paysage fluide* dans lequel la dispersion (active ou passive) prend place à grande échelle
15 (Queiroz 2005; Cowie & Holland 2006; Nikula *et al.* 2013). La fragmentation des habitats,
16 quoique documentée (Acosta 1999; González-Wevar *et al.* 2010) s'y restreint d'ordinaire aux
17 zones côtières ou aux espèces benthiques aux niches très spécialisées (Rex *et al.* 1993; Whitlatch *et*
18 *al.* 1998; Poulin *et al.* 2014). La dispersion passive de longue distance (mue par le courant) a été
19 observée chez des organismes allant du plancton (par exemple les stades vagiles des laminaires,
20 voir Fraser *et al.* 2009; Nikula *et al.* 2010) aux vertébrés supérieurs, en particulier dans leurs pre-
21 miers stades de développement (comme c'est le cas chez les tortues marines, Gaspar *et al.* 2012).
22 La dispersion active des vertébrés est courante à des échelles transocéaniques (Bowen & Siniff
23 1999; Le Boeuf *et al.* 2000).

24 Dans les systèmes océaniques caractérisés par de forts courants (comme le Courant Circumpo-
25 laire Antarctique ou ACC, voir §7 - ou le courant de Humboldt, voir §147), la dispersion active
26 comme passive sont des caractères généraux des écosystèmes. Dans l'ACC, une forte dispersion a
27 été mise en évidence pour des espèces allant des algues et petits invertébrés (Nikula *et al.* 2010)

1 aux invertébrés benthiques (Arango *et al.* 2011) et aux poissons (Matschiner *et al.* 2009; Damer-
2 au *et al.* 2012). Cependant, la diversité génétique le long des distributions circumpolaires de-
3 meure variable entre taxons, et semble dépendre largement du potentiel de dispersion et des traits
4 d'histoire de vie (Rogers 2007). Dans de nombreux cas, des discontinuités telles que la Péninsule
5 Antarctique (qui s'étend au Nord jusqu'à presque 62°S) rompt ces systèmes de dispersion ho-
6 mogène, et apparaît comme une province biogéographique distincte, en terme de climate (Mul-
7 vaney *et al.* 2012) comme de biogéographie (Terauds *et al.* 2012).

8 La pouvoir de dispersion active qu'ont les vertébrés a néanmoins une conséquence paradoxale :
9 leur plus grande capacité de mouvement leur permet de *ne pas* disperser avec le courant. Tandis
10 que les espèce sessiles résistent à la dispersion par la pure force de leur inertie, les nageurs actifs et
11 les oiseaux font preuve de philopatrie (§5 - voir aussi Frederiksen & Petersen 1999; Steiner 2005;
12 Bicknell *et al.* 2012; Fernández-Chacón *et al.* 2013), qui s'oppose à la grande fluidité des envi-
13 ronnements océaniques par un mouvement actif. La philopatrie permet donc la perpétuation de
14 dèmes relativement stables (par exemple chez l'otarie, Bonin *et al.* 2013), de groupes familiaux
15 (comme chez l'orque, Hoelzel 1998), ou de colonies (comme chez la plupart des oiseaux marins,
16 Friesen *et al.* 2007). De fait, dans la plupart des cas, le comportement philopatric est assez fort
17 pour mener à la différenciation génétique entre des population qui se recouvrent pourtant large-
18 ment dans leurs aires d'alimentation (Dearborn *et al.* 2003; Friesen *et al.* 2007; Smith *et al.*
19 2007). De telles structures ont été observées chez les mammifères marins (cétacés ou pinnipèdes,
20 voir Hoelzel 1998; Baker *et al.* 2008; Bonin *et al.* 2013), chez les tortues marines (par ex. Molfet-
21 ti *et al.* 2013), et chez la plupart des oiseaux de mer (Friesen *et al.* 2007; Bicknell *et al.* 2012). Les
22 déplacements libres et rapides permis par les systèmes océaniques ont donc la conséquence inat-
23 tendue de permettre de forts taux de retour au lieux de naissance pour la plupart des vertébrés,
24 plutôt que de favoriser le brassage génétique.

25 *L'Antarctique et l'Océan Austral*

1 §-7 *Océanographie et géographie*. On nomme Océan Austral la vaste masse d'eau qui entoure le
2 continent Antarctique, et qui constitue l'un des points d'échange thermique majeurs entre
3 l'océan, l'atmosphère et la cryosphère (Hofmann & Maqueda 2011; Meijers 2014). Bien que sa
4 définition officielle fixe sa limite septentrionale à 60°S (International Hydrographic Organiza-
5 tion 2000), sa définition fonctionnelle s'étend d'habitude au moins jusqu'au Front Polaire
6 Antarctique (APF), aux alentours de 55°S (Moore *et al.* 1999; Gersonde *et al.* 2005), latitude à
7 laquelle il se fond dans les Océans Atlantique Sud, Indien, et Pacifique Sud (voir Fig. 6).

8 L'océanographie de l'Océan Austral est dominée par l'influence des vents planétaires de la cellule
9 de Ferrell australe (De Boer *et al.* 2008; Kohfeld *et al.* 2013) qui soufflent continuellement aux
10 latitudes intermédiaires, sans être interrompus par aucune terre dans l'hémisphère Sud, et qui
11 mettent en mouvement les eaux de surface. Le Courant d'Ouest (*West Wind Drift* ou Courant
12 Circumpolaire Antarctique, ACC) qui en résulte est l'un des courants les plus puissants sur terre,
13 avec un transport moyen de 105 à 140 Svedrups - et c'est le seul qui connecte directement les
14 trois bassins océaniques majeurs (Rintoul *et al.* 2001). Le principal effet des vents d'Ouest est de
15 causer la dérive vers l'Est des eaux de surface dans l'ACC : mais ils sont aussi responsables du
16 transport d'Ekman (dans l'hémisphère Sud, une déviation antihoraire allant jusqu'à 90° du
17 courant de surface par rapport à la direction du vent, voir Toggweiler & Samuels 1998), qui
18 pousse les eaux de surface vers l'Equateur. Dans la région antarctique, le refroidissement dû au
19 vent et à l'évaporation, causés par les vents katabatiques, et la surconcentration de sel dû au rejet
20 de la saumure durant la congélation de la surface autour des polynyes (voir Thomas & Dieck-
21 mann 2008 et §18) rendent les eaux de surface particulièrement froides et salées - et par con-
22 séquent denses. Durant leur transport vers le Nord, ces masses d'eaux rencontrent des eaux plus
23 chaudes et moins denses : les isopycnaux¹ se redressent, et les eaux de surface antarctiques coulent
24 vers le fond, sur la zone de convergence de l'APF, pour former les Eaux Profondes Antarctiques
25 (AABW, voir Rintoul *et al.* 2001). Au sud de l'APF, par conséquent, le transport d'Ekman est di-
26 vergent : il provoque la résurgence des eaux profondes plus chaudes (*Warm Deep Water*, WDW, et

1. Isolignes de densité constante

1 *Circumpolar Dee Water*, CDW - voir Sarmiento *et al.* 2004; Anderson *et al.* 2009 pour plus de
2 détails) au sud de l'APF.

3 Au Nord de l'APF s'étend la zone subantarctique, limitée au Nord par la convergence subtropi-
4 cale (STC) où les eaux subantarctiques et subtropicales se stratifient sous les eaux tropicales de
5 surface, aux alentours de 40°S. Quoiqu'elle soit loin d'être aussi nette que l'APF, la STC peut être
6 définie par la disparition du thermocline¹ permanent (Tomczak & Godfrey 2003) : au Nord de la
7 STC, la température chute rapidement de la surface vers le fond, jusqu'à environ 1000 m. Au
8 Sud de la STC, par contre, le gradient de température entre la surface et le fond s'estompe, avec
9 parfois moins d'1°C de différence dans toute la colonne d'eau en Antarctique. L'APF, cependant,
10 conserve un thermocline saisonnier à la limite inférieure de la couche mixte (le SMC, voir Char-
11 rassin & Bost 2001). Au Sud de l'ACC, plus près du continent Antarctique, le Courant Côtier
12 Antarctique, ou dérive d'Est, forme un contrecourant sur l'étroit plateau continental (Tchernia &
13 Jeannin 1980; Fahrbach *et al.* 1994). Ces deux courants se combinent dans les grandes gyres de
14 l'Océan Austral, dans les mers de Ross et de Weddell (Gouretski 1999; Meredith *et al.* 2000), qui
15 ont des caractéristiques océanographiques particulières - par exemple la résurgence d'eaux
16 chaudes profondes en mer de Weddell, qui y entretient une polynye à chaleur sensible (c'est à
17 dire une polynye causée par la température de l'eau, et non par le stress éolien - voir Thomas &
18 Dieckmann 2008).

19 L'importance de l'Océan Austral et de l'ACC dépasse largement leurs effets régionaux. Avec le
20 Pacifique Nord, l'ACC est la plus grande zone de résurgence d'eaux profondes anciennes
21 (Primeau 2005; Marshall & Speer 2012), et de ce fait un moteur majeur de la circulation ther-
22 mohaline mondiale (ou MOC, de l'anglais *Meridional Overturning Circulation* - voir Marshall &
23 Speer 2012), ce qui lui donne un rôle central dans les échanges thermiques globaux. Dans la
24 mesure où c'est l'une des rares zones où les eaux profondes affleurent directement à la surface,

1. Le *thermocline* est la zone de transition entre la couche mixte de surface, où les radiations solaires et le mixage dû au vent et à la marée résultent en des températures relativement homogènes et chaudes, et les eaux profondes plus froides. Le thermocline est caractérisé par un gradient serré de température dans la colonne d'eau.

1 l'ACC joue aussi un rôle essentiel dans les échanges gazeux entre l'océan profond et l'atmosphère
2 : une augmentation de l'activité de résurgence a en effet été reliée à une augmentation de CO₂ at-
3 mosphérique dans l'Océan Austral (Anderson *et al.* 2009). La résurgence de l'Océan Austral a
4 aussi été identifiée comme l'une des voies majeures de retour à la surface des nutriments après
5 leur plongée aux basses latitudes - ce qui en fait une pièce essentielle de la pompe biologique, et
6 rend compte de près des trois quarts de la productivité au nord de 30°S (Sarmiento *et al.* 2004).
7 Enfin, puisque c'est la seule masse d'eau zonalement illimitée sur terre, et qu'il se mêle aux trois
8 grands bassins océaniques, l'Océan Austral joue le rôle de mécanisme de couplage central entre
9 des masses d'eau autrement presque indépendantes, et permet la téléconnexion entre les différents
10 bassins (White & Peterson 1996; Rintoul *et al.* 2001). L'Océan Austral se prolonge aussi au Nord
11 dans les courants de limite Est de l'hémisphère Sud (le courant de Humboldt dans le Pacifique -
12 voir §160, le courant de Benguela en Atlantique, et le plus marginal courant d'Australie
13 occidentale).

14 L'influence locale de l'ACC et de l'APF sur les processus biologiques est conséquente. La résur-
15 gence d'eau profondes chargées de nutriments, qui s'oxygène dans la couche mixte, est à l'orig-
16 ine d'une productivité primaire massive dans la zone de l'APF (jusqu'à 20% de la productivité
17 primaire marine mondiale, voir Laubscher *et al.* 1993; Bathmann *et al.* 1997; Carr *et al.* 2006).
18 Cette productivité nourrit de vastes populations de krill (*Euphausia sp.*, voir Murphy *et al.* 2007a)
19 et de poisson (Pakhomov *et al.* 1996) - ceux-ci dominés par les myctophidés (en particulier *Elec-*
20 *trona sp.*, *Protomyctophum sp.*, *Gymnoscopelus sp.* and *Krefflichthys sp.*, voir Collins *et al.* 2008).
21 Plusieurs prédateurs supérieurs tirent avantage de cette concentration de proies sur l'APF - par ex-
22 emple l'Elephant de mer *Mirounga leonina* (Boyd & Arnborn 1991), ou le Manchot royal (Char-
23 rassin & Bost 2001). A la limite sud de l'ACC, la couverture de glace hivernale assure des flo-
24 raçons de productivité saisonnière : les années de grande glace ont les plus fortes biomasses de
25 krill (Nicol *et al.* 2008), avec des conséquences directs sur les populations de prédateurs
26 supérieurs (comme le Manchot Adélie *Pygoscelis adeliae*, Nicol *et al.* 2008), ce qui fait de la zone
27 des glaces flottantes un aire de grande productivité (Tynan 1998).

1 §-8 *Variabilité climatique de l'Océan Austral*. La structure complexe et les connexions planétaires
2 de l'Océan Austral font qu'il répond fortement aux principaux modes de variabilité dans l'hémis-
3 phère austral : le Mode Annulaire Austral (SAM), et les dipôles subtropicaux - dont le plus im-
4 portant est l'ENSO (*El Niño Southern Oscillation*), mais qui incluent aussi le dipôle Indien
5 (SIOD, *Southern Indian Ocean Dipole*) et le SASD (*South Atlantic Subtropical Dipole*).

6 Le SAM est la différence moyenne de pression entre les latitudes moyennes et le centre de
7 l'Antarctique. La circulation atmosphérique qui en résulte est la cause principale de variabilité cli-
8 matique dans l'Océan Austral (Marshall 2003; Abram *et al.* 2014), avec une forte cyclicité décen-
9 nale, et de plus lentes oscillations séculaires (Abram *et al.* 2014). Le SAM a des effets inversés
10 dans les zones polaires et subtropicales. Durant les phases positives du SAM, le déplacement vers
11 le pôle des vents d'Ouest renforce le transport d'Ekman dans l'ACC, et a pour résultat un dé-
12 placement vers le Nord de l'APF, une activité de résurgence renforcée au sud de l'APF, et une aug-
13 mentation de la productivité dans la zone frontale (Lovenduski & Gruber 2005) - tandis que
14 dans la zone subtropicale, les anomalies positives de température de surface ont pour résultat une
15 diminution de la productivité marine. En revanche, durant les phases négatives (c'est à dire
16 lorsque les vents d'Ouest se déplacent vers le Nord), la productivité de l'Océan Austral décroît,
17 avec des effets importants sur les prédateurs supérieurs (Forcada & Trathan 2009; Bost *et al.*
18 2015).

19 Tandis que l'impact du SAM se fait principalement sentir dans l'Océan Austral, chaque bassin
20 océanique est aussi sujet aux effets plus locaux - quoique téléconnectés - des dipôles subtropi-
21 caux. Bien que l'ENSO diffère significativement du SIOD et du SASD, en particulier par son
22 ampleur, la structure générale de ces trois systèmes est l'oscillation de température entre l'Est et
23 l'Ouest de la zone subtropicale de chaque bassin. Un gradient important de température garantit
24 de forts alizés et une importante circulation de Walker¹. Dans le cas de l'ENSO, la diminution
25 périodique de la circulation atmosphérique le long de la côte Ouest d'Amérique du Sud diminue

1. C'est à dire, une circulation zonale des masses d'air sous les tropiques, mue principalement par des différences est-ouest de température et de pression.

1 la résurgence d'eaux froides profondes, et par conséquent la différence de température entre l'Est
2 et l'Ouest du Pacifique Sud : la circulation de Walker s'en trouve réduite, ce qui ramène les pré-
3 cipitations vers l'Amérique du Sud, et élève la température de surface dans tout le Pacifique Sud
4 (Tomczak & Godfrey 2003; Stuecker *et al.* 2013). Récemment, des phénomènes similaires ont
5 été mis en évidence dans l'Océan Indien, avec le SIOD (Saji *et al.* 1999) et l'Atlantique Sud avec
6 le SASD (Taschetto & Wainer 2008; Wainer *et al.* 2014). Leur lien avec l'ENSO et leur télé-
7 connexions globales sont cependant encore mal compris (Saji & Yamagata 2003; Ashok *et al.*
8 2004; Abram *et al.* 2008). Ensemble, le SAM, l'ENSO, le SIOD et le SASD contribuent à la
9 variabilité de la glace de mer autour d'Antarctique (Simmonds & Jacka 1995; Lefebvre 2004;
10 Stammerjohn *et al.* 2008).

11 Les effets combinés du SAM, de l'ENSO et des dipôles ont pour résultat des modes de variabilité
12 complexes dans l'Océan Austral (Fogt & Bromwich 2006; Ciasto & Thompson 2008; Fogt *et al.*
13 2011), qui influencent profondément les communautés biologiques. La productivité primaire de
14 l'Océan Austral réagit aussi bien au SAM (Lovenduski & Gruber 2005) qu'à l'ENSO (Beaufort *et*
15 *al.* 2001; Behrenfeld *et al.* 2001), la plupart du temps à travers les changements de température
16 de surface, de glace de mer, et de disponibilité des nutriments. Les changements dans les pro-
17 priétés physiques des eaux de surface, et dans les interactions trophiques, se répercutent à tous les
18 niveaux trophiques au dessus du phytoplancton - par exemple sur le krill (Murphy *et al.* 2007b) :
19 à leur tour, ces effets se transfèrent aux niveaux plus élevés, par exemple aux baleines mysticètes
20 (Leaper *et al.* 2006), aux éléphants de mer (McMahon & Burton 2005) ou aux manchots (Bost *et*
21 *al.* 2015). Quoique la variabilité climatique soit un phénomène tout à fait naturel, et ne doive pas
22 menacer la résilience des espèces, la grande intégration des mécanismes climatiques de l'Océan
23 Austral le rendent susceptible de rétroactions importantes, et y amplifie les effets du changement
24 climatique (voir §12). Il est donc de toute première importance de comprendre ces processus
25 eux-mêmes, ainsi que leurs effets sur les communautés biologiques.

26 *§-9 Histoire climatique de l'Océan Austral.* L'Océan Austral est supposé avec pris son aspect actuel
27 au début du Pliocène (il y a ~5.2 Ma). La séparation physique entre l'Antarctique et l'Amérique

1 du Sud se produit entre l'Eocène moyen (~41 Ma, Livermore *et al.* 2005) et la transition
2 Oligocène-Miocène (~23 Ma, Pfuhl & McCave 2005), avec le début des conditions glaciales
3 (dont l'ouverture du passage de Drake a peut être été l'une des causes - Nong *et al.* 2000; Liver-
4 more *et al.* 2004). L'ACC (voir §7) se met en place au Miocène supérieur (Pfuhl & McCave
5 2005), mais les frontières actuelles des masses d'eau semblent s'être établies plus tard, peut-être
6 aux alentours de la Transition climatique du Miocène moyen (MMCT, ~16-11.5 Ma, voir Hay-
7 wood *et al.* 2008; Knorr & Lohmann 2014). Cette idée semble confirmée par la date de radiation
8 évolutive de plusieurs taxa, inférée génétiquement, et dont les clades Antarctiques et Subantar-
9 tiques semblent avoir été isolés depuis la MMCT (Poulin *et al.* 2014) - ce qui coïncide aussi avec
10 la radiation évolutive des manchots modernes (voir §21).

11 Le Miocène marque la transition entre les conditions de serre, et les conditions glaciales actuelles.
12 La calotte glaciaire antarctique replace la toundra jusque là dominante au début du Miocène
13 (Raine & Askin 2001), et se développe jusqu'à atteindre ses dimensions actuelles à la fin de la
14 période (Westerhold *et al.* 2005). Malgré une forte variabilité (encore mal comprise), les condi-
15 tions en mer sont globalement plus chaudes au Pliocène qu'à présent (Haywood *et al.* 2008) -
16 mais avec néanmoins une forte productivité. La température de surface est environs 5°C
17 supérieures aux moyennes actuelles dans l'Océan Austral (Whitehead 2003) et l'APF se situe 6°
18 au Sud de sa position présente au milieu du Pliocène (Barron 1996). La circulation thermoha-
19 line semble pourtant avoir été plus forte qu'à présent (Haywood *et al.* 2008). A la fin du Pliocène
20 (~2.5 Ma), cependant, le refroidissement s'accroît en Antarctique, et les vents d'Ouest accrus
21 accompagnent une migration septentrionale de l'APF, accompagnée d'une possible réduction de
22 la circulation thermohaline (McKay *et al.* 2012).

23 La plupart des îles subantarctiques se forment durant cette période. Mis à part les archipels conti-
24 nentaux de Géorgie du Sud et des Malouines (qui sont des reliques du Gondwana, tout comme
25 l'Amérique du Sud voisine), et du plateau subcontinental des Kerguelen (formé durant
26 l'Oligocène), les îles subantarctiques sont principalement de jeunes formations volcaniques, qui
27 s'élèvent directement au-dessus de la plaine abyssale. Crozet, Heard et Macquarie apparaissent en-

1 tre le Pliocène et le Pléistocène (voir un exposé complet dans Quilty 2007). Ainsi, que ce soit en
2 termes d'océanographie ou de géographie, il semble que l'Océan Austral n'ait pris sa forme
3 actuelle qu'au début du Pleistocène.

4 Quoique mieux connue, l'histoire de l'Océan Austral au Pléistocène (~2.6 Ma - 11.7 Ka) et à
5 l'Holocène (~11.7 Ka - présent) est toujours incertaine. Généralement, la reconstruction des
6 processus atmosphériques et océanographiques à ces périodes relativement récentes repose sur
7 deux approches différentes. D'abord, des indices directs ou indirects peuvent être obtenus par le
8 biais de carottes, soit dans le sédiment marin, soit dans la calotte glaciaire, soit dans la tourbe des
9 îles subantarctiques. Par ailleurs, les modèles de circulation générale (GCM) développés pour la
10 période historique peuvent donner de bons résultats pour le passé récent, puisque les caractéris-
11 tiques les plus importantes de la circulation océan-atmosphère peuvent être considérées stables
12 durant cette période : ainsi le Coupled Model Intercomparison Project phase 5 (CMIP5, voir
13 §50) inclut-il des expériences dédiées au Pléistocène et à l'Holocène moyen (voir §50 pour le dé-
14 tail de ces expériences). Les indices sédimentologiques nous fournissent des marqueurs chim-
15 iques, physiques et biostratigraphiques qui servent de révélateurs pour un certain nombre de vari-
16 ables (Gersonde *et al.* 2005; Armand & Leventer 2010; Hodgson *et al.* 2014), soit directement
17 au site de la carotte (comme c'est le cas pour les carottes de sédiment terrestre ou marin - voir
18 Gersonde *et al.* 2005; Martínez Garcia *et al.* 2009; Armand & Leventer 2010), ou pour une com-
19 binaison de sources locales et distantes (Wolff *et al.* 2003; Wolff *et al.* 2006). Les assemblages de
20 diatomées et de foraminifères dans le sédiment marin offrent un indice précis de l'état des masses
21 d'eau de surface, et peuvent être utilisés pour identifier la position de l'APF (comme cela a été fait
22 sur le Plateau de Campbell, voir Neil *et al.* 2004, ou dans l'Océan Indien et l'Atlantique, voir
23 Kemp *et al.* 2010), ou l'étendue de la glace de mer (Hodell *et al.* 2001; Gersonde *et al.* 2005). Les
24 pollens fossiles dans le sédiment marin et terrestre sont aussi de bons révélateurs des conditions
25 climatiques locales (McGlone *et al.* 2010). Des indicateurs chimiques peuvent être extraits de
26 carottes de glace comme de sédiment. Ces différents moyens sont utilisés soit directement, soit
27 comme conditions initiales pour un modèle numérique de climat, pour reconstituer les condi-

1 tions globales. Cependant, la réconciliation des différentes sources disponibles demeure problé-
2 matique, tant les signaux peuvent être confondus entre les environnements glaciaires, marins et
3 terrestres (voir Armand & Leventer 2010; McGlone *et al.* 2010).

4 Malgré des variations locales et des conditions divergentes dans les différents bassins océaniques,
5 il est possible d'identifier quatre périodes majeures dans l'histoire récente de l'Océan Austral -
6 périodes qui semblent cohérentes à l'échelle globale. (i) les conditions du Quaternaire (du Pléis-
7 tocène moyen à 22 Ka), (ii) le Dernier Maximum Glaciaire (LGM - 21-18 Ka), (iii) la déglacia-
8 tion de la fin du Pléistocène et l'Optimum Climatique de l'Holocène (17-9 Ka), et (iv) les condi-
9 tions hypsithermales et néoglaciales de la fin de l'Holocène (8-0 Ka) - voir §101 pour plus de
10 détails sur ces périodes.

11 *§-10 Biogéographie de l'Océan Austral.* L'Océan Austral actuel a une biogéographie relativement
12 simple, faite de systèmes généralement zonaux, mais qui abritent pourtant une biodiversité éton-
13 nante (Smetacek & Nicol 2005; Chown *et al.* 2015). Quatre régions distinctes se distinguent : (i)
14 la zone *subantarctique*, qui dépasse l'Océan Austral au sens strict, mais lui appartient d'un point
15 de vue biogéographique, et est limitée au sud par l'APF, et au nord par la STC, (ii) la zone de
16 *l'océan antarctique*, au sud de la zone subantarctique, et limitée au sud par la limite sud de
17 l'ACC, (iii) la zone *côtière antarctique* qui inclut le contrecourant antarctique et le plateau conti-
18 nental, ainsi que la zone des glaces côtières, et (iv) la *péninsule antarctique* (Convey *et al.* 2012;
19 Terauds *et al.* 2012). Dans cette étude, nous nous concentrerons surtout sur les trois premières
20 zones concentriques. La région Subantarctique se caractérise par des températures plus modérées,
21 d'importantes précipitations, aucune glace de mer, et de forts vents d'Ouest. Les îles sont cou-
22 vertes de tourbières et de végétation rase, ou un habitat périglaciaire. Les îles subantarctiques sont
23 rares et espacées (du Pacifique vers l'Atlantique : Macquarie, Heard-et-McDonald, Kerguelen,
24 Crozet, Marion-et-Prince-Edward, Bouvet, Sandwichs du Sud, Géorgie du Sud, Terre de Feu et
25 Malouines), et concentrent oiseaux et pinnipèdes. L'Océan antarctique, au sud de l'APF, se carac-
26 térise par des eaux de surface beaucoup plus froides, et une importante couverture de glace hiver-
27 nale - les îles y sont rares et couvertes de glace (Peter I, Scott, Balleny et les Orcades du Sud). La

1 végétation est rase et limitée aux lichens, hépatiques et mousses (Convey *et al.* 2012). La zone
2 antarctique côtière (Tynan 1998) est quant à elle un environnement de plateau continental, car-
3 actérisé par une couverture de glace quasi-permanente, et une forte influence de la calotte
4 glaciaire antarctique. La végétation y est presque absente, et y fait preuve d'adaptations extrêmes,
5 et la faune vertébrée se limite à une poignée d'espèces : les manchots Empereur et Adélie, le pétrel
6 des neiges, le pétrel géant, le pétrel antarctique, le fulmar antarctique, le damier du cap, l'océanite
7 de Wilson, le labbe antarctique, et quelques espèces de pinnipèdes. De toutes ces espèces, seul le
8 Manchot Empereur se reproduit l'hiver austral.

9 Les ressources alimentaires proviennent principalement de deux sources. Comme mentionné plus
10 haut (§7), les vastes zones de résurgence associées à l'APF sont d'importants contributeurs à la
11 productivité marine globale (Laubscher *et al.* 1993; Bathmann *et al.* 1997; Carr *et al.* 2006), et
12 attirent de nombreuses espèces comme le Manchot royal (Péron *et al.* 2012). La zone de glaces
13 flottantes concentre aussi une vaste productivité primaire (Arrigo *et al.* 1997; Arrigo *et al.* 1998),
14 et constituent une ressource alimentaire importante pour les espèces côtières antarctiques comme
15 le Pétrel des neiges (Barbraud & Weimerskirch 2001) ou le Manchot Adélie (Wienecke *et al.*
16 2000). Les zones de résurgence côtière d'eau libre dans les polynyes qui bordent la côte (voir §18)
17 sont particulièrement propices à la productivité de plancton, et constituent des zones d'alimenta-
18 tion privilégiées pour plusieurs espèces (Ancel *et al.* 1992; Ancel *et al.* 1999). Ainsi l'une des car-
19 actéristiques les plus remarquables de la biogéographie de l'Océan Austral est-elle son lien presque
20 immédiat avec les caractéristiques climatiques générales. Contrairement aux environnements
21 terrestres, où les complexes interactions de la latitude et de l'altitude ont pour résultat des motifs
22 très hétérogènes (Burrows *et al.* 2011), la structure zonale de l'Océan Austral en accord avec ses
23 grandes structures de circulation annulaires et circumpolaires. La conséquence en est que la distri-
24 bution de ces zones est en lien direct avec les modes de circulation généraux (voir §18).

25 *Réponse des espèces aux changements climatiques*

1 §-11 *Le Climat change*. Les reconstitutions paléoclimatiques nous apprennent que le changement
2 climatique est, en lui-même, un phénomène naturel et inévitable à l'échelle millénaire (voir §9),
3 et est partiellement lié à des forçages extérieurs (comme les changements dans l'activité solaire, les
4 cycles orbitaux dits de Milankovitch - et leurs rétroactions positives, comme la boucle de la « terre
5 boule de neige », voir Hyde *et al.* 2000; Berger 2013), ou à des modifications biogéniques in-
6 ternes dans la composition de l'atmosphères (voir Frolking & Roulet 2007). Cependant, ces
7 changements se font lentement. Même le dernier maximum glaciaire l'a pas change significative-
8 ment la structure ou la composition des communautés en Europe (voir Yeakel *et al.* 2013). Les
9 changements climatiques actuels, par contre, se caractérisent par des changements du même or-
10 dre que la transition Pléistocène-Holocène, mais sur des intervalles de temps de l'ordre de la dé-
11 cennie, et non du millénaire. Le changement climatique a donc été défini de façon non ambiguë
12 par l'Intergovernmental Panel on Climate Change (IPCC) comme « *un changement dans l'état du*
13 *climat qui peut être identifié (par exemple par des tests statistiques) par des changements dans l'état*
14 *moyen et/ou la variance de ses propriétés, et qui persiste pour une longue période de temps, typiquement*
15 *des décennies ou plus. Le changement climatique peut être dû à des processus naturels internes, ou des*
16 *forçages externes comme des modulations des cycles solaires, des éruptions volcaniques, ou des change-*
17 *ments anthropogéniques persistants dans la composition de l'atmosphère et l'utilisation du territoire* »
18 (Solomon *et al.* 2007). Durant les dernières décennies du XXème siècle, les preuves accumulées
19 montrent clairement l'aspect unique des rapides changements dont nous sommes actuellement
20 témoins, et qui commencent peu de temps après la Révolution industrielle du XIXème siècle. La
21 combinaison de déforestation et d'utilisation à grande échelle de combustibles fossiles a mené
22 l'augmentation de la production, et à la diminution de la fixation, des gaz à effet de serre (GHG),
23 en particulier du dioxyde de carbone et du méthane, au point que « *les concentrations atmo-*
24 *sphériques de dioxyde de carbone, et méthane et de protoxyde d'azote sont sans précédent depuis au*
25 *moins 800,000 ands* » (Pachauri *et al.* 2014). Une concentration accrue de GHG a pour résultat
26 une augmentation de l'effet de serre, c'est à dire du forçage radiatif efficace sur le climat global.
27 De sorte que tous les indices actuels montrent qu'il est « *extrêmement probable que l'influence hu-*

1 *maine a été la cause déterminante du réchauffement observé depuis le milieu du XXème siècle*» (Stock-
2 *er et al.* 2014).

3 Des efforts intenses et coordonnés sont en œuvre à présent pour comprendre les mécanismes
4 sous-jacents, et les issues possibles des changements actuels (voir §50 pour une présentation des
5 efforts de l'IPCC) - malheureusement, il n'en va pas de même des efforts politiques. Le résultat de
6 cet effort scientifique est une connaissance croissante de l'articulation entre trois différents do-
7 maines : le changement climatique, le changement environnemental, et la météorologie. Quoique
8 profondément connectés, ces trois paradigmes ne devraient pas être confondus. Le Changement
9 climatique est seul l'objet de la définition de l'IPCC, et se limite aux effets visibles dans la circula-
10 tion océan-atmosphère - c'est à dire aux changements de long terme de la température, des pré-
11 cipitations, du régime des vents, etc. Ces phénomènes peuvent être précisément décrits et saisis
12 avec les méthodes de la physiques, comme les modèles de circulation générale océan-atmosphère
13 (AO-GCM, voir §50), dont les limites résident surtout dans notre compréhension de la nature
14 des processus physiques à l'œuvre, ainsi que dans les hypothèses socio-économiques qui gouver-
15 nent nos projections des futures concentrations de GHG. Quoique les AO-GCMs soient de plus
16 en plus efficaces pour prédire l'environnement physique de la Terre au XXIème siècle, la réponse
17 des communautés biologiques, ou « changement environnemental », est un sujet plus complexe
18 (voir §13 sqq. pour un aperçu de ces difficultés), à la fois par manque d'un exemple passé
19 (quoique le réchauffement de l'Holocène soit souvent utilisé comme exemple, son rythme est en
20 effet beaucoup plus lent - voir Petit *et al.* 2008; Blois *et al.* 2013; Moritz & Agudo 2013). Les
21 divers mécanismes par lesquels ces changements affectent les écosystèmes les rendent particulière-
22 ment difficiles à prévoir. Ces effets vont du dérèglement des processus développementaux des
23 organismes de bas niveau trophique par le rachauffement de l'eau (Gregg *et al.* 2003; Behrenfeld
24 *et al.* 2006) ou son acidification (Harvey *et al.* 2013), qui se répercutent aux niveaux trophiques
25 supérieurs (McMahon & Burton 2005), jusqu'au changement des aires de distributions et des
26 réductions de biodiversité qui en résultent (par exemple chez les copépodes arctiques, voir
27 Beaugrand *et al.* 2002). Les écosystèmes marins, tout comme les écosystèmes terrestres, sont déjà

1 sévèrement affectés par l'exploitation humaine, qui diminue leur résilience aux changements cli-
2 matiques (Halpern *et al.* 2008; Watson *et al.* 2013; McCauley *et al.* 2015). L'effondrement de la
3 biodiversité prédit par les scénarios les plus sombres (voir §50) pourrait ainsi toucher de 50 à
4 70% des océans plus fortement que tous les changements qui ont eu lieu depuis le Pléistocène
5 (Beaugrand *et al.* 2015).

6 Par ailleurs, de nombreuses études ont étudié les effets du changement climatique en prenant
7 pour modèle la variabilité cyclique du climat, qui est particulièrement prévalente dans les sys-
8 tèmes océaniques (voir §8). Ces études ont souvent exploité le fort contraste entre les phases posi-
9 tives et négatives de l'ENSO (par exemple, Le Bohec *et al.* 2008), du SAM (Weimerskirch *et al.*
10 2012) ou de l'Oscillation d'Atlantique Nord (NAO, Frederiksen *et al.* 2004). C'est souvent une
11 approximation raisonnable, et, de façon importante, souvent la seule possible - cependant, ces os-
12 cillations météorologiques ne devraient pas être complètement assimilées au changement clima-
13 tique. La météorologie, qui est surtout en jeu dans ce cas, est seulement l'interface entre le
14 changement climatique et les espèces. Les phases chaudes de l'ENSO reproduisent peut-être pré-
15 cisément les conditions d'une planète réchauffée dans l'Océan Austral, mais pas les mécanismes
16 de téléconnexion plus vastes, par exemple avec l'hémisphère Nord : et ce malgré le fait que le cou-
17 plage Arctique-Antarctique soit sans conteste un aspect central du changement climatique au sens
18 strict (Barbante *et al.* 2006). Si l'approximation météorologique est souvent une nécessité
19 méthodologique, il convient de garder à l'esprit la différence essentielle entre changement
20 météorologique et changement climatique.

21 *§-12 Particularités du changement climatique dans les régions polaires.* Les effets du changement cli-
22 matique sont particulièrement visibles dans les régions polaires, un fait dû à leurs caractéristiques
23 climatologiques, mais aussi à la structure de leurs communautés biologiques. D'un point de vue
24 purement climatologique (ou « abiotique », voir §49), les régions polaires apparaissent différentes
25 du reste de la planète. Les singularités de la circulation Océan-Atmosphère dans l'Arctique - et
26 tout particulièrement l'abrupt changement d'albedo selon l'étendue de la glace - donne lieu à un
27 phénomène d'*amplification arctique* (Serreze & Francis 2006; Pithan & Mauritsen 2014).

1 L'Antarctique, d'un autre côté, semble plus ambiguë : tandis que la partie orientale paraît « résis-
2 ter » au changement (potentiellement grâce à l'effet tampon de l'Océan Austral, voir Mulvaney *et*
3 *al.* 2012 - quoique les incertitudes liées aux méthodes d'observation puissent masquer la tendance
4 réelle, voir Hanna *et al.* 2013), la partie occidentale et la Péninsule sont au nombre des régions
5 connaissant le réchauffement le plus rapide sur Terre (Mulvaney *et al.* 2012; Bromwich *et al.*
6 2013). Dans l'Océan Austral, le réchauffement accéléré des eaux de surface (Liu & Curry 2010)
7 comme des eaux profondes (Purkey & Johnson 2010) a été directement observé, et ses effets de
8 l'influx d'eau douce en provenance de la calotte antarctique sur l'activité de résurgence a de même
9 déjà été documentée (de Lavergne *et al.* 2014). Les connexions globales fondamentales de
10 l'Océan Austral le rendent aussi particulièrement sensible aux changements dans la circulation
11 tropicale, comme l'ENSO (voir §8 et Collins *et al.* 2010). De sorte qu'un effet globalement accru
12 du changement climatique est à craindre.

13 Outre ses caractéristiques physiques, la biogéographie de l'Océan Austral le rend aussi partic-
14 ulièrement sensible aux changements climatiques. Tandis que les habitats devraient généralement
15 se déplacer vers les pôles en réponse au réchauffement (Williams *et al.* 2007; Fraser *et al.* 2012 et
16 §15), les écosystèmes polaires ont relativement peu de marge pour se déplacer, et les hautes lati-
17 tudes devraient devenir des zones de « subduction climatique » (Burrows *et al.* 2014), où les habi-
18 tats se contractent et finissent par disparaître, plutôt que de se déplacer plus avant (§15). Ce
19 phénomène est encore compliqué par le fait que l'activité humaine a fortement influencé l'Océan
20 Austral depuis le bas de la chaîne trophique (c'est à dire *via* le changement climatique anthro-
21 pogène et la réduction de productivité primaire associée), comme depuis le haut (par le massacre
22 des prédateurs supérieurs - baleines, phoques ou manchots - voir Ainley *et al.* 2007), ce qui
23 donne lieu à des effets combinés et complexes (Smetacek & Nicol 2005 - voir aussi l'exemple
24 concret et spectaculaire donné par Trivelpiece *et al.* 2011).

25 Par conséquent, les régions polaires, en apparence intactes, concentrent désormais en réalité les
26 effets du changement climatique. De nombreuses années durant, l'opinion générale fut que les
27 écosystèmes polaires étaient plus simples, et basés sur des chaînes trophiques plus courtes, que les

1 écosystèmes des basses latitudes - menant ainsi à des bouleversements plus facilement intelligi-
2 bles. Cependant, il est à présent admis que la diversité taxonomique et la complexité du réseau
3 trophique sont aussi élevés que dans les régions tempérées (Smetacek & Nicol 2005; Chown *et al.*
4 2015 et Fig. 6), et il est probable que la particularité apparente des régions polaires ne soit due à
5 rien d'autre qu'à leur destruction plus tardive, garantie par leur inaccessibilité - et non à quelque
6 caractéristique particulière qui y favoriserait la mégafaune, par exemple. Smetacek & Nicol 2005
7 ont ainsi proposé que ces régions soient considérées comme un vestige des écosystèmes marins
8 préhistoriques - un « *Serengetis des mers* ». De sorte que les processus que nous voyons à l'œuvre
9 dans les régions polaires ont déjà eu lieu depuis longtemps sous les tropiques, et sont les mêmes
10 qui sont à l'œuvre en Afrique continentale, avec le déclin de la macrofaune mammifère et son
11 remplacement par l'exploitation humaine omniprésente. Une question cependant reste en suspens
12 - le changement climatique est-il spécial en lui-même dans les régions polaires, au-delà de son ry-
13 thme accéléré ; ou ses effets sur les communautés biologiques sont-ils surtout le résultat de la rela-
14 tive intégrité de ces communautés encore peu anthropisées ?

15 *§-13 Réponse des espèces au changement climatique.* L'importance des changements globaux actuels
16 sur les communautés biologiques est reconnue par l'ensemble de la communauté scientifique (voir
17 §11 et Thomas *et al.* 2004; Warren *et al.* 2013; Pacifici *et al.* 2015), mais le détail de leurs im-
18 pacts est encore matière à débat (Pereira *et al.* 2010; He & Hubbell 2011; Pacifici *et al.* 2015).
19 Les différents niveaux auxquels les impacts se produisent (qui vont de la physiologie à l'organisa-
20 tion des communautés, voir §11), la complexité des interactions biotiques (Davis *et al.* 1998;
21 Norberg *et al.* 2012; Midgley & Bond 2015), et le caractère idiosyncratique de la sensibilité et de
22 l'exposition de chaque espèce (Moritz & Agudo 2013; Comte *et al.* 2014; Dickinson *et al.* 2014)
23 contribuent à l'apparente hétérogénéité des réponses au niveau des écosystèmes (Parmesan &
24 Yohe 2003). La plupart des études, cependant, s'accordent sur l'idée que trois mécanismes fonda-
25 mentaux interagissent dans la détermination de la réponse d'une espèce au changement : (i) la
26 microévolution adaptative par sélection naturelle (Bradshaw & Holzapfel 2006; Hoffmann &
27 Sgrò 2011), (ii) la plasticité phénotypique et en particulier les changements de phénologie (Char-

1 mantier *et al.* 2008; Hoffmann & Sgrò 2011) et (iii) les changements de distribution pour suivre
2 les isohabitats (Walther *et al.* 2002; Chen *et al.* 2011; VanDerWal *et al.* 2013).

3 §-14 *Micro-évolution et plasticité phénotypique.* La micro-évolution et la plasticité phénotypique
4 partagent une structure commune : afin de s'adapter à son environnement local, une espèce
5 change ses caractéristiques physiologiques, phénologies, ou généralement comportementales
6 (Hoffmann & Sgrò 2011). Les mécanismes impliqués, par contre, sont très différents. Une
7 réponse micro-évolutive nécessite un certain nombre de générations, par sélection des traits héri-
8 tables conférant un meilleur avantage (et déterminés soit génétiquement, soit épigénétiquement)
9 : c'est donc un processus de population. La plasticité phénotypique, par contre, est un
10 phénomène individuel. Soit par des modifications développementales, soit par des processus
11 purement comportementaux, chaque individu optimise sa valeur sélective dans les conditions
12 présentes - mais son nouvel état n'est pas héritable, si nous laissons de côté la relativement mar-
13 ginale (dans l'état actuel des connaissances) transmission culturelle des comportements (White-
14 head & Rendell 2014).

15 Des réponses évolutives au changement climatique anthropogénique ont déjà été mises en évi-
16 dence chez plusieurs espèces à temps de génération court, comme les moustiques (Gienapp *et al.*
17 2008; Hoffmann & Sgrò 2011). Cependant, aucune étude conduite sur des vertébrés n'a encore
18 pu mettre au jour de processus évolutif clair (soit Møller *et al.* 2010 pour un exposé complet chez
19 les oiseaux). Le décalage entre le rythme rapide du changement, les temps de génération plus
20 longs des vertébrés, et leurs stratégies reproductives généralement de type K, rend la tâche ardue
21 pour la sélection naturelle.

22 La plasticité, d'un autre côté, joue un rôle important dans la réponse de court terme de nom-
23 breuses espèces (Charmantier *et al.* 2008; Hoffmann & Sgrò 2011). Elle affecte en priorité la
24 phénologie, afin de conserver une synchronisation avec les conditions météorologiques, et en par-
25 ticulier le pic de nutriments, comme cela a été démontré chez la mésange charbonnière *Parus ma-*
26 *gor* (Charmantier *et al.* 2008; Husby *et al.* 2010) ou le Gobemouche à collier (*Ficedula hypoleuca*,

1 Dunn *et al.* 2010), mais aussi chez certaines plantes (Anderson *et al.* 2012). Mais elle affecte aussi
2 les traits métaboliques (McKechnie *et al.* 2006; Nicotra *et al.* 2010) ou la morphologie (Przybylo
3 *et al.* 2000). La structure hautement dynamique de l'Océan Austral fait de la phénologie un
4 mécanisme central pour la survie de la plupart des espèces. Les variations saisonnières de la tem-
5 pérature causent un déplacement cyclique de l'APF (voir §7), et changent ainsi la distance entre
6 les terrains d'alimentation et les aires de reproduction pour de nombreuses espèces (Massom *et al.*
7 2009; Bost *et al.* 2015). Elles déterminent la force des résurgences et par conséquent la productiv-
8 ité marine, jusque dans le courant de Humboldt (Thiel *et al.* 2007). Ainsi, tout comme la syn-
9 chronisation avec le pic d'abondance des insectes est un facteur de succès reproducteur crucial
10 pour de nombreux oiseaux terrestres (Dunn *et al.* 2010), les variations saisonnières des aires d'ali-
11 mentation dans les systèmes océaniques sont probablement une contrainte majeure pour leur
12 phénologie, et son évolution devrait amener un changement dans la phénologie de plusieurs es-
13 pèces. De tels changements, cependant, ont un coût (Van Buskirk & Steiner 2009), et aug-
14 mentent la probabilité d'une désynchronisation entre différentes contraintes écologiques (Both *et*
15 *al.* 2006) - la plasticité a donc ses limites (Visser & Both 2005; Both *et al.* 2006; Møller *et al.*
16 2008).

17 En résumé, les réponses micro-évolutives, qui nécessitent une sélection, à une échelle multi-
18 générationnelle, de variantes génétiques pré-existantes, sont un processus de temps long. Elles ont
19 été mises en évidence chez plusieurs espèces à courte durée de vie (Van Heerwaarden & Hoff-
20 mann 2007; Hoffmann & Willi 2008; Hoffmann & Sgrò 2011; Krehenwinkel *et al.* 2015), mais
21 l'extrême vélocité des changements actuels (Mahlstein *et al.* 2013; Poloczanska *et al.* 2013; Bur-
22 rows *et al.* 2014) est un défi majeur pour les espèces à longues générations (Hoffmann & Sgrò
23 2011) et faible diversité génétique (Norberg *et al.* 2012), et ce d'autant plus que ces deux carac-
24 tères sont souvent corrélés (Romiguier *et al.* 2014). La micro-évolution et la plasticité sont, de
25 plus, interconnectés : et seule l'ampleur de l'héritabilité permet actuellement de les distinguer
26 (Réale *et al.* 2003), d'autant plus que la plasticité elle-même est un trait héritable (Nussey *et al.*
27 2005; Pelletier *et al.* 2007).

1 Dans cette étude, cependant, le rythme des changements et les traits d'histoire de vie des oiseaux
2 de mer excluent, a priori, toute réponse évolutive forte. Une réponse plastique est possible, et
3 même souhaitable. Mais son étude dépasse le cadre de ce travail, puisqu'elle nécessite des données
4 transgénérationnelles actuellement indisponibles. Jusqu'à présent, des indices convaincants ont
5 été trouvés chez des espèces à courtes générations (Charmantier *et al.* 2008) - mais des décennies
6 d'observation phénologiques et comportementales seront nécessaires pour construire une image
7 claire chez les oiseaux marins.

8 *§-15 Déplacement d'habitat et de distributions.* Les déplacements de distributions semblent de plus
9 en plus centraux dans les réponses de court terme au changement climatique (Davis & Shaw
10 2001; Chen *et al.* 2011; Burrows *et al.* 2014). Quoique l'expression principale de la réponse plas-
11 tique se concentre sur la phénologie (§14), c'est à dire sur le cadre *temporel* du cycle reproducteur
12 des espèces ou des populations, les déplacements de distributions représentent un changement de
13 son cadre *spatial*. Par opposition à la micro-évolution et à la plasticité, ce déplacement implique
14 uniquement un changement géographique, a priori sans contraintes physiologiques ou autres -
15 cependant, les cas de déplacements « pur », non associés à des changements de phénologie, sont
16 rares (Møller *et al.* 2010). Les déplacements de distributions sont l'expression de facteurs allant
17 de l'échelle de l'individu à celle de l'espèce. A un niveau fonctionnel, le *conservatisme de niche*
18 peut être défini comme la tendance, pour un individu, à demeurer à l'intérieur de sa zone de con-
19 fort physiologique et écologique (Wiens *et al.* 2010). Selon la largeur de cette niche, et le po-
20 tentiel d'évolution de l'espèce, le conservatisme de niche peut se vérifier pendant de très longues
21 périodes. Par exemple, les oiseaux du Nouveau Monde sont un exemple frappant de conser-
22 vatismisme de niche à l'échelle géologique, dans la mesure où les taxa les plus apomorphiques occu-
23 pent les niches les plus spécialisées (Hawkins *et al.* 2006), tandis que des structures similaires
24 s'observent aussi chez les poissons d'eau douce (Comte *et al.* 2014) ou les mammifères terrestres
25 (Martínez Meyer *et al.* 2004). Ce conservatisme de niche dans le temps long signifie que les
26 préférences d'habitat ne évoluent lentement, et peuvent être héritées phylogénétiquement dans
27 un taxon, ralentissant la colonisation de nouveaux habitats.

1 A des échelles de temps plus courtes, au contraire, le conservatisme de niche est la force prin-
2 cipale qui pousse les individus à suivre leur habitat plutôt qu'à s'adapter à de nouvelles condi-
3 tions. En ce sens, les espèces *suivent leur habitat* - un mouvement par lequel une population
4 entière se déplace avec ses conditions environnementales. Le suivi d'habitat, observé depuis une
5 perspective humaine, résulte dans le *déplacement des distributions*, une modification globale de la
6 distribution géographique de l'espèce en réponse aux tendances de son habitat. Le réchauffement
7 climatique, par exemple, a pour résultat typique (quoique schématique) un déplacement des
8 habitats et des espèces vers les pôles, ou les sommets (Jump *et al.* 2009; Fraser *et al.* 2012).

9 Dans un système idéal, on peut s'attendre à ce que les espèces suivent leur habitat de façon
10 linéaire (Scheffer *et al.* 2001). Cependant, le caractère anisotropique des biozones implique que
11 les habitats ne se transforment pas de façon homogène (Diffenbaugh & Field 2013), et que les
12 changements sont vraisemblablement doublement hétérogènes - entre écosystèmes, et entre es-
13 pèces à l'intérieur des écosystèmes. Une étude récente sur les oiseaux australiens (VanDerWal *et*
14 *al.* 2013), par exemple, a montré que durant les 60 dernières années, les déplacements de distrib-
15 utions ont suivi des trajectoires majoritairement idiosyncratiques à l'échelle du continent, et qui
16 ne reflètent que partiellement les déplacements d'habitats. Un tel exemple montre l'importance
17 des interactions biotiques dans la détermination de la distribution des espèces : plusieurs relations
18 interspécifiques peuvent être affectées simultanément par le changement climatique (par exemple
19 le comportement alimentaire - voir Charmantier *et al.* 2008, ou la compétition - voir Stenseth *et*
20 *al.* 2015), avec pour résultat des réponses non linéaires et parfois paradoxales (comme c'est le cas
21 pour la Bernache nonnette *Branta leucopsis*, voir Bauer *et al.* 2008; Eichhorn *et al.* 2009). Ainsi,
22 la compréhension des distributions présentes et futures est presque impossible si l'on se limite à
23 l'étude d'une seule espèce (par exemple dans une approche de type «enveloppe climatique», voir
24 Pearson & Dawson 2003), et doit au contraire être étendue pour intégrer les interactions bio-
25 tiques les plus importantes (un point développé au §49).

26 L'hétérogénéité et l'idiosyncrasie des déplacements de distributions est exacerbé à chaque fois que
27 les discontinuités de l'environnement provoquent des *convergences d'habitat* (Burrows *et al.* 2014)

1 ou «subductions», par exemple lorsque des corridors climatiques où les déplacements d'habitats
2 pourraient, en théorie, se faire de façon linéaire, sont interrompus par des barrières soit naturelles,
3 soit anthropogéniques. La fragmentation des habitats, qui est souvent le résultat des influences
4 humaines (mais qui peut être naturelle, voir par ex *The King synnome*) interagit donc avec le
5 changement climatique pour donner lieu à des conséquences souvent imprévisibles. Ceci est par-
6 ticulièrement vrai autour des «gouffres climatiques» où les conditions locales disparaissent tout
7 de bon (Burrows *et al.* 2014). De plus, les espèces elles-mêmes ne répondent pas nécessairement
8 de façon linéaire aux déplacements d'habitats : des capacités de dispersion insuffisantes, par exem-
9 ple, peuvent conduire à des déplacements décalés (Bertrand *et al.* 2011; Schloss *et al.* 2012), une
10 compétition accrue autour des zones de gouffre peut empêcher l'expansion de la distribution de
11 certaines espèces (Norberg *et al.* 2012), et des mécanismes compensatoires peuvent masquer pour
12 un temps les effets du changement jusqu'à atteindre un point de bascule (Doak & Morris 2010).
13 Ces effets retardés peuvent nous pousser à sous-estimer l'importance des changements à l'œuvre,
14 et placer les espèces en situation de dette climatique, laissant finalement place à des
15 effondrements d'écosystèmes (Scheffer *et al.* 2001; Hoegh-Guldberg & Bruno 2010; Barnosky *et*
16 *al.* 2012).

17 Ainsi, les régions polaires, qui traversent une période de changements accélérés (Serreze & Francis
18 2006; Turner *et al.* 2014; Chown *et al.* 2015) et ont une marge de manœuvre très limitée pour les
19 déplacements d'habitats sont un exemple emblématique de ces problèmes (Williams *et al.* 2007).
20 Détecter les réponses délayées avant d'atteindre leur point de bascule est une priorité si nous
21 voulons développer des stratégies de mitigation appropriées, et prédire avec précision les distribu-
22 tions d'espèces futures pour y préparer des réserves environnementales (Keppel *et al.* 2012; Hope
23 *et al.* 2013; Watson *et al.* 2013), en particulier pour les prédateurs supérieurs, qui déterminent
24 dans une large mesure les propriétés de résilience des écosystèmes (Soulé *et al.* 2003; Heithaus *et*
25 *al.* 2008) afin de développer des stratégies de conservation et de gestion efficaces. Ainsi, la prédic-
26 tion des changements de distribution apparaît-elle encore comme un défi majeur dans le domaine
27 de la modélisation biologique.

1 *Objectifs de cette étude*

2 §-16 *Les Echelles du changement.* L'un des aspects les plus déroutants de la biologie du change-
3 ment climatique est le labyrinthe d'échelles dans lesquelles il se déploie. Les études actuelles oscil-
4 lent souvent entre deux extrêmes : une approche de « temps long », dans lequel les changements
5 climatiques « géologiques » (tels que la transition Pléistocène-Holocène) sont prises pour exemple
6 des changements anthropogéniques actuels, et une de « temps court », dans lequel la stochasticité
7 météorologique comme ses phénomènes cycliques (par exemple l'Oscillation d'Atlantique Nord,
8 ou NAO) sont utilisés à la même fin. Dans ce contexte, ce qui rend particulièrement difficile la
9 compréhension des changements climatiques actuels est qu'ils se déroulent à une échelle pour
10 laquelle nous n'avons que peu de concepts, parce qu'ils n'ont pas de précédent dans l'histoire de la
11 Terre : ils nous forcent à trouver un intermédiaire entre le temps écologique et le temps évolu-
12 tionniste. Une question centrale est donc de savoir si les observations que nous pouvons obtenir
13 sur le temps long (par exemple en paléodémographie) sont de même nature que celles que nous
14 obtenons sur le temps court (par exemple du suivi de terrain) - et si l'une ou l'autre de ces obser-
15 vations peut expliquer et prédire les événements résultat des changements climatiques anthro-
16 pogéniques. Par bien des aspects, il semble que les observations quasi-instantanées que nous pou-
17 vons faire sur des populations d'étude, et qui s'étendent, au mieux, sur quelques générations, sont
18 dominées par plusieurs niveaux de stochasticité et de forçages mixtes - la stochasticité et les cycles
19 de la météorologie autour des grandes tendances du climat, la stochasticité de la démographie au-
20 tour des trajectoires générales des populations et des espèces, la restructuration lente mais chao-
21 tique des écosystèmes alors que leurs éléments évoluent dans des directions différentes - tous ces
22 signaux contradictoires peuvent masquer pour quelques temps la tendance inévitablement direc-
23 tionnelle du changement climatique. De l'autre côté de ce spectre, pourtant, la paléodémogra-
24 phie, qui lisse tous ces événements stochastiques, et ces cycles de hautes fréquences, n'en conserve
25 que les motifs les plus lents, comme la lente pulsation des glaciations du Pléistocène, qui laissent

1 à l'évolution le temps de se réaliser. Mais elle pourrait, de ce fait, être totalement impuissante à
2 saisir le mouvement relativement rapide des changements actuels.

3 Pour ces raisons, ce qui se passe réellement dans une situation telle que les changements clima-
4 tiques anthropogéniques actuels peut être conçu comme une disruption du rythme et de la
5 manière selon laquelle ces différentes échelles sont articulées, et une probabilité croissante de dé-
6 calage entre les différents niveaux de variation spatiale et temporelle. La durée d'une génération,
7 par exemple, peut de façon simpliste être considérée comme le laps de temps durant lequel les
8 conditions environnementales sont intégrées par un individu qui est soumis à la sélection na-
9 turelle : cette durée est suffisante pour faire la moyenne des saisons, voire des modes climatiques -
10 mais assez courte pour qu'un grand nombre de générations s'écoulent afin de laisser place à l'évo-
11 lution au long des changements géologiques. Un changement au rythme plus serré a de fortes
12 chances de mettre à mal cette synchronisation. De même, l'organisation dans l'espace des indi-
13 vidus et des espèces reflète l'étendue au sein de laquelle les conditions les rendent les mieux adap-
14 tés (voir §49), et la taille et la mobilité des individus est adaptée à cet espace : là encore, la con-
15 vergence ou la divergence des habitats brise cet équilibre. Ainsi, par bien des aspects, il semble
16 que la clef de la compréhension des changements climatiques contemporains soit la façon dont
17 les différents paradigmes chronologiques et spatiaux s'articulent pour les écosystèmes, les espèces
18 et les individus.

19 *§-17 Comprendre les échelles d'organisation des oiseaux de mer.* Le but principal de notre travail sera
20 de comprendre l'organisation des différentes échelles spatiales et temporelles dans la démographie
21 et la dynamique des espèces d'oiseaux de mer - un groupe aux premières lignes du changement
22 climatique (voir §18) - en se basant sur le contraste de trois espèces : le manchot Royal (*Apten-*
23 *odytes patagonicus*), le manchot Empereur (*Aptenodytes forsteri*) et la puffinure de Garnot (*Pele-*
24 *canoides garnotii*). Si notre étude se place principalement dans une perspective évolutionniste, en
25 s'appuyant sur les outils de la génétique des populations, nous explorerons aussi une large gamme
26 d'approches complémentaires - comme le suivi par marquage individuel, ou la modélisation
27 d'habitats. Pour chacune des deux espèces du genre *Aptenodytes*, nous présentons deux études

1 complémentaires - l'une centrée sur des processus globaux, et ancrée dans les données de la
2 génomique (voir *The King Synnome* et *The Emperor Synnome*), et l'une centrée sur les phénomènes
3 de corut terme, à l'échelle de la génération (voir *Fine-scale structure* et *Empirical evidence for het-*
4 *erogeneous dispersal*). Ces deux approches sont combinées dans l'étude plus exploratoire de la
5 Puffinure de Garnot (*Unexpected philopatry*).

6 (i) La biogéographie comparée de ces trois espèces sert de base à l'étude. Une première tâche sera
7 donc de comprendre comment la distribution de chaque espèce s'organise, à des échelles allant de
8 la diversité intra-coloniale, à l'île, à la distribution complète - et d'établir ce qui semble être
9 l'unité d'étude démographique la plus pertinente pour l'étude des changements passés, présents et
10 futurs. Cette première direction impliquera aussi une révision des estimations de taille de popula-
11 tion, de structure, et, dans le cas du manchot Royal, de la diversité taxonomique.

12 (ii) Des prédictions plutôt sombres ont été formulées pour ces trois espèces, face au changement
13 environnemental. Le manchot Empereur a récemment été déclaré condamné dans toutes son aire
14 de distribution (Jenouvrier *et al.* 2014), la colonie principale, et la mieux étudiée, du manchot
15 Royal semble montrer des perspective pessimistes (Le Bohec *et al.* 2008; Péron *et al.* 2012), et la
16 Puffinure de Garnot a été classée *Menacée* en 1994 sur la liste rouge de l'IUCN. Cependant, afin
17 d'obtenir une meilleure compréhension de la dynamique des populations d'oiseaux de mer, nous
18 devons procéder à une ré-évaluation critique de ces prédictions, et à une intégration des
19 différentes méthodes qui y ont conduit. Dans ce but, nous utilisons toute l'information
20 disponible sur la mieux connue de ces trois espèces, le manchot Royal, et proposons une interpré-
21 tation alternative de la réponse du manchot Empereur aux changements de l'Océan Austral.
22 Dans ces deux cas, nous tenterons de proposer une projection consensuelle, basée à la fois sur des
23 observations de long terme et de court terme, comme une tentative de solution au délicat prob-
24 lème de l'intégration des données écologiques et évolutives dans le contexte du changement
25 climatique.

Structure and dynamics of the penguin synnomes

Résumé

L'Océan austral est l'un des pivot des écosystèmes et du climat de notre planète, qui concentre plus de 20% de la productivité primaire marine mondiale. La complexité de ses réseaux trophiques et son inaccessibilité rendent plus encore qu'ailleurs nécessaire l'utilisation d'espèces bio-indicatrices. Plusieurs espèces de manchots (comme le Manchot Royal et le Manchot Empereur) sont ainsi l'objet de programmes de suivi à long terme.

Dans cette étude, nous utilisons les données offertes par la génomique des populations (« RAD-sequencing » couvrant le génome de centaines d'individus issus couvrant la distribution de ces deux espèces) et les représentations numériques du climat de l'IPCC-CMIP5 pour calibrer dans le temps long les analyses démographique plus précises réalisées à l'échelle de quelques générations dans le cadre de suivis démographiques, et mieux comprendre la réponse des manchots au changement climatique.

Au-delà de ses conséquences immédiates pour l'étude des Manchots en tant que sentinelles de l'Océan Austral, cette étude montre l'intérêt d'une plus forte intégration de la génomique des populations dans les études démographiques et comportementales.

Mots-clés : *génétique des populations, RAD-sequencing, océan austral, sphéniscidés, changement climatique.*

Résumé en anglais

The Southern Ocean plays a central role in the regulation of the Earth's climate and ecosystems, and accounts for more than 20% of the world's marine productivity. The complexity of its trophic networks and its sheer inaccessibility make the use of bioindicator species more necessary there than anywhere else. Several penguin species (such as the King and the Emperor penguin) are therefore the focus of long-term monitoring programs.

In this study, we use the information from population genomics (« RAD-sequencing » data covering the genome of hundreds of individuals from the two species' full distribution) and from IPCC-CMIP5 numerical climate models to calibrate in the long time the more precise demographic analyses realised in the framework of field surveys, and understand penguin responses to climate change.

Beyond its implications for the study of penguins as sentinels of the Southern Ocean, our work demonstrates the interest of a stronger integration of population genomics in demographic and behavioural investigation.

Keywords: *population genetics, RAD-sequencing, Southern Ocean, spheniscids, climate change.*