



HAL
open science

**Isolement reproductif et architecture génomique de la
différenciation chez deux espèces du complexe Jaera
albifrons (isopodes marins) - Etude de populations
mixtes présentant des niveaux d'isolement
interspécifique contrastés**

Ambre Ribardière

► **To cite this version:**

Ambre Ribardière. Isolement reproductif et architecture génomique de la différenciation chez deux espèces du complexe Jaera albifrons (isopodes marins) - Etude de populations mixtes présentant des niveaux d'isolement interspécifique contrastés. Evolution [q-bio.PE]. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066397 . tel-01746147

HAL Id: tel-01746147

<https://theses.hal.science/tel-01746147>

Submitted on 29 Mar 2018

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.

Thèse de doctorat de l'Université Pierre et Marie Curie

Ecole Doctorale 227 UPMC-MNHN
Sciences de la Nature et de l'Homme : Evolution et Ecologie
Spécialité Diversité du Vivant

Présentée par

Ambre RIBARDIERE

En vue de l'obtention du grade de
DOCTEUR de L'UNIVERSITE PIERRE ET MARIE CURIE

**Isolement reproductif et architecture génomique de la différenciation chez
deux espèces du complexe *Jaera albifrons* (isopodes marins)**

-

**Etude de populations mixtes présentant des niveaux d'isolement
interspécifique contrastés**

Présentée et soutenue publiquement le 30 Novembre 2017,

Devant un jury composé de :

Dr. Richard CORDAUX Directeur de Recherche, CNRS, Poitiers	Rapporteur
Dr. Guillaume EVANNO Chargé de Recherche, INRA, Rennes	Rapporteur
Pr. Christophe DESTOMBE Professeur, UPMC, Paris	Examineur
Dr. Carole SMADJA Chargée de Recherche, CNRS, Montpellier	Examinatrice
Dr. Myriam VALERO Directrice de Recherche, CNRS, Roscoff	Examinatrice
Dr. Pierre-Alexandre GAGNAIRE Chargé de Recherche, CNRS, Montpellier	Membre invité
Dr. Thomas BROQUET Chargé de Recherche, CNRS, Roscoff	Directeur de thèse

A tous ceux qui ont un jour cru en moi,

Remerciements

Qu'il est parfois « simple » (hum...) d'écrire une partie d'un article et parfois si difficile de trouver les bons mots pour exprimer la gratitude qui nous habite envers quelques personnes...

Ce manuscrit n'aurait jamais eu lieu sans plusieurs personnes qui ont réussi à me pousser dans mes retranchements et à aller chercher le meilleur de moi-même en toutes circonstances.

Mes premiers remerciements iront tout naturellement à mon Directeur de Thèse, Thomas. Merci. C'est un mot court mais comme tu le sais j'aime faire court ! Bon d'accord, cette fois je vais essayer de développer plus ! Merci de m'avoir ouvert les portes de la Station, et m'avoir fait découvrir le monde des *Jaera albifrons*, j'ai vraiment aimé faire cette thèse ! Merci pour les discussions et brain-storming, pour nos excursions échantillonnages fort sympathiques en Normandie, ainsi que les repas (souvent copieux) partagés. Merci de m'avoir fait grandir dans le milieu qu'est la recherche ! Il me reste encore beaucoup à apprendre, mais les bases sont (j'espère) pas trop mauvaises maintenant, grâce à toi. Tu as été un super « Chef ». Merci d'avoir pris le temps de m'écouter, de m'expliquer, (parfois) de m'avoir obligée à prendre du temps pour me reposer et surtout de m'avoir soutenue jusqu'au bout de cette thèse.

Merci également à Fred, d'avoir accepté d'être responsable administrative de cette thèse, sans quoi je n'aurais pas pu écrire ce manuscrit de presque 300 pages!!

Je voudrais remercier le Jury d'avoir accepté d'évaluer mon travail de thèse et plus particulièrement Richard Cordaux et Guillaume Evanno, pour avoir accepté de rapporter mon travail. Et merci également aux membres de mon comité de thèse (Nico, Pierre-Alex et Didier), pour le temps que vous avez pris et vos commentaires constructifs.

Merci à toute la Team *Jaera* !!

Claire, ça a vraiment été instructif et agréable de travailler avec toi ! J'espère te revoir bientôt, peut-être lors d'un fest-noz ?! Jérôme et Céline, on a finalement réussi à tout déterminer et fixer !! Gildas, merci de ne pas avoir désespéré lors des recherches de coquilles dans les scripts ! Merci à tous ceux qui ont participé de près ou de loin à cette aventure : Komlan Avia pour m'avoir formée à JoinMap, Sebastien Colin pour les belles photos 3D, Denis Roze pour l'échantillonnage aux Scilly et Olivier Timsit pour celui en Normandie, Alan Brelsford pour le protocole RAD, Sébastien Henry et Stéphane Loisel pour les entretiens d'élevage et les échantillonnages, Glenn Yannick pour la relecture du papier, et Christophe pour avoir essayé de comprendre mes arbres phylogénétiques. Merci aux stagiaires, Julia, Arnaud, Elise pour leur implication, et plus particulièrement Elsa pour ces 6 mois sympathiques.

Merci aux sétois pour leur accueil et les discussions toujours hyper intéressantes. Merci particulièrement à Nico et Pierre-Alex. Pierre merci pour ton aide tout au long de ces 3 années, pour ta sympathie et pour les after-works. J'ai encore loupé la soirée plage cette année, peut-être l'année prochaine ?

Merci Richard pour les discussions autour de *Wolbachia*, et pour avoir accepté de relire ce manuscrit. Merci à Marius pour ses échanges sur les *Jaera* et sa gentillesse, et enfin merci Alan pour les soirées repas/dadi !

Je voudrais également remercier toute l'équipe DIVCO actuelle et ancienne, ainsi que l'équipe BEDIM pour leurs sourires au détour du couloir ou autour d'un goûter !

Merci à François pour la mission sur Carese, qui ne m'aura finalement (et malheureusement) pas pris beaucoup de temps.. ainsi qu'à Céline, pour ton aide dans les différents couacs administratifs.

Merci au service administratif et particulièrement à Nathalie ! Merci à Brigitte pour son accueil toujours sympathique et à Franck, Barbara et les membres du Gulf Stream pour les bons petits plats qu'ils proposent.

Merci à ma famille, et plus particulièrement mes parents et ma sœur pour leur soutien (même à plusieurs milliers de kms), leurs efforts à lire mes publications et à comprendre ce que je fais, et tout simplement leur présence. Vous êtes mon socle. Merci également à ma belle-famille pour tout leur soutien, merci de m'avoir aussi bien acceptée dans la famille. Merci également à mon beau-frère et ma belle-sœur, pour leur accueil toujours sympa à Montpellier ! Quelques personnes nous ont quittés ces dernières années, mais leurs encouragements et gentillesse m'aident encore aujourd'hui à vouloir faire toujours mieux. Enfin merci à ma plus vieille amie : Merci Marion d'avoir toujours été là et de l'être encore.

La thèse ça a également été l'AJC, et l'organisation de la JJC donc merci aux copains avec qui j'ai vécu cette aventure, pas toujours évidente ! ;) Dans l'ensemble, merci aux copains, toujours présents quand il s'agit d'aller boire un verre et de relâcher la pression ! Alexis, presque 4 ans en coloc, c'est intense, mais je suis contente d'avoir partagé ça avec toi. Merci également aux filles de mon bureau Marine et Eloïse, ainsi qu'aux autres copains de la Station (en vrac et non pas par ordre d'importance: Pym, Margot, Solène, Charles, Laure, Peche, Ulysse, Emilie, Perrine, Damien, Camille, Erwann, Florian..) ainsi que les copains de mariage (Virginie, Benoit, Karine) et bien sûr les copains de discussions, jeux et initiation au surf: Théophile, Justine, Hugo, Laura, Tristan et Doriane (et Gaspard, le petit dernier !). Et Meci à toi Morgane, pour nos petites confidences de covoiturage, et les bons moments passés à la zumba et au pound !

Laura, merci de m'avoir entraînée à la danse africaine et d'avoir passé du temps avec moi sur cette fin de thèse à rendre ce manuscrit agréable ! Doriane et Anaïs, merci pour les petits et grands moments partagés ensemble (et les bouquets !! ;)) Anaïs, tu es un vrai bol d'air, je suis très heureuse de pouvoir te compter parmi mes amies ! Merci pour votre présence, à Romain et toi, durant la « dernière ligne droite » qui aura décidément été bien longue !

Mes derniers mots iront pour mon « pacsé ». Merci Pierre-Marie de rendre ma vie si lumineuse et heureuse. Ta présence (merci internet et le téléphone), ton soutien et ton amour inconditionnels me rendent chaque jour un peu plus forte.

AVANT-PROPOS

Cette thèse a été effectuée à la Station Biologique de Roscoff, et a été soutenue par l'implication de différentes personnes.

Equipe d'accueil : Equipe DIVCO, UMR 7144, Station Biologique de Roscoff

Claire Daguin-Thiébaud, Ingénieure de Recherche, a participé à toutes les manipulations en laboratoire (développement de marqueurs microsatellites, constructions de banques RAD, et recherche de bactéries endosymbiotiques). De plus elle a participé à l'encadrement des deux stagiaires (Julia Centanni et Arnaud Davo) impliqués dans la recherche de *Wolbachia* au sein de notre complexe d'espèces.

Jérôme Coudret, Assistant Ingénieur, a contribué à certains échantillonnages, à l'entretien des différents élevages et à l'identification spécifique des individus élevés tout au long de cette thèse.

Stéphane Loisel, Assistant Ingénieur, a participé à certains échantillonnages.

Collaborations hors équipe d'accueil :

Gildas Le Corguillé, Ingénieur d'Etudes, Plateforme ABIMS, FR2424, Station Biologique de Roscoff, a été présent pour superviser le nettoyage des séquences d'ADN obtenues lors de différentes banques RAD.

Céline Houbin, Assistante Ingénieure, Service Mer et Observation, FR2424, Station Biologique de Roscoff, a participé à certains échantillonnages, à la détermination et au phénotypage des individus.

Pierre-Alexandre Gagnaire, Chargé de Recherche CNRS, équipe Spéciation, Evolution, et Adaptation en milieu marin, Isem, Université Montpellier 2, a participé aux chapitres 5 et 6 de cette thèse (histoire démographique / analyse de l'architecture génomique), et m'a notamment particulièrement aidée dans l'obtention des spectres joints de fréquences alléliques.

Komlan Avia post-doc dans l'équipe Biologie Intégrative des Modèles Marins, UMR 8227, Station Biologique de Roscoff, a participé au chapitre 6 de cette thèse, avec une implication particulière dans la cartographie.

Richard Cordaux, Directeur de Recherche CNRS, équipe Ecologie et Biologie des Intéractions, UMR 7267, Université de Poitiers, a participé au chapitre 4 de cette thèse, concernant la réflexion autour de la recherche de *Wolbachia*.

Sébastien Colin, Ingénieur de Recherche, équipe Evolution des Protistes et Ecosystèmes Pélagiques, UMR 7144 Station Biologique de Roscoff, a réalisé l'acquisition et le traitement des images 3D des périopodes (Figure I.16 page 30).

Olivier Timsit, Chargé d'études, Groupe d'Etude des Milieux Estuariens et Littoraux de Normandie, et Caroline Broudin, Ingénieure d'Etudes, FR2424, Station Biologique de Roscoff, ont participé ponctuellement aux échantillonnages en Normandie.

Enfin ce projet a été financé par l'ANR ISOBAR, et a été suivi tout du long par Thomas Broquet, Chargé de Recherche CNRS, équipe DIVCO - UMR 7144, Station Biologique de Roscoff, Directeur de Thèse.

TABLE DES MATIERES

INTRODUCTION GENERALE.....	1
1. De la spéciation à l'isolement reproductif complet	3
1.1 Spéciation : définition et mécanismes	3
1.2 Spéciation primaire vs contact secondaire	6
1.3 Histoires démographiques	7
1.4 Barrières à la reproduction.....	7
2. Les zones hybrides : des laboratoires naturels.....	15
3. Deux taxons du complexe d'espèce <i>Jaera albifrons</i> (Crustaceae) comme modèles d'étude	21
3.1 Taxonomie	21
3.2 Ecologie	22
3.3 Comportement sexuel	24
3.4 Différenciation génétique	26
3.5 Isolement reproductif et hybridation	29
4. Objectifs de la thèse et structure du document	31
CHAPITRE 1 : Test de l'hypothèse d'hybridation entre <i>J. albifrons</i> et <i>J. praehirsuta</i>, différenciation génétique et comparaison de zones avec et sans hybridation entre ces deux taxons.....	35
1. Développement de marqueurs microsatellites pour le complexe d'espèces <i>Jaera albifrons</i>	37
2. Analyse de la différenciation génétique entre deux espèces du complexe <i>Jaera albifrons</i> , en zone d'isolement reproductrice et en zone d'hybridation .	44
CHAPITRE 2 : Structure génétique intraspécifique et interspécifique au sein du complexe <i>Jaera albifrons</i>.....	71
1. Introduction	73
2. Matériel et Méthodes	75
2.1 Zones de sympatrie et limites de la zone hybride.....	75
2.2 Le complexe <i>Jaera albifrons</i> : Différenciation génétique entre les quatre espèces présentes sur les côtes européennes	78
2.3 Différenciation génétique entre <i>Jaera albifrons</i> et <i>Jaera praehirsuta</i>	82
3. Résultats.....	83
3.1 Zones de sympatrie et limites de la zone hybride	83

3.2 Le complexe <i>Jaera albifrons</i> : Différenciation génétique entre espèces des côtes européennes	84
3.3 Différenciation génétique entre <i>Jaera albifrons</i> et <i>Jaera praehirsuta</i>	87
4. Discussion.....	89
4.1 Distribution de <i>J. albifrons</i> / <i>J. praehirsuta</i>	89
4.2 Structure génétique de la paire <i>J. albifrons</i> / <i>J. praehirsuta</i>	90
4.3 Contexte de relations inter-populationnelles	92
4.4 Méthode RAD-seq.....	92
4.5 Résultats marquants et perspectives	93
CHAPITRE 3 : Isolement reproductif.....	95
1. Introduction.....	97
2. Matériel et méthodes	99
2.1 Croisements contrôlés	99
2.2 Préférences femelles	104
3. Résultats	105
3.1 Croisements contrôlés.....	105
3.2 Préférences femelles	112
4. Discussion.....	112
4.1 Isolement sexuel.....	112
4.2 Isolement post-zygotique.....	113
4.3 Perspectives.....	116
CHAPITRE 4 : La bacterie <i>Wolbachia</i> impliquée dans l'isolement reproductif ?	119
1. Introduction.....	121
2. Matériel & Méthodes	124
2.1 Analyse du sex-ratio chez le complexe <i>Jaera albifrons</i>	124
2.2 Détection moléculaire de <i>Wolbachia</i>	125
3. Résultats	131
3.1 Sex-ratio des adultes provenant de populations naturelles	131
3.2 Sex-ratio chez les juvéniles élevés en laboratoire	131
3.3 Détection moléculaire de <i>Wolbachia</i>	134
4. Discussion.....	140
CHAPITRE 5 : Histoire démographique sous-jacente à la divergence entre <i>J. albifrons</i> et <i>J. praehirsuta</i>.....	145
1. Introduction.....	147
2. Matériel et méthodes	150

2.1 Echantillonnage	150
2.2 Génotypage	151
2.3 Histoire démographique : données	152
2.4 Histoire démographique : modèles	154
3. Résultats	158
3.1 Analyse de la diversité génétique au sein de la Bretagne et de la Normandie	158
3.2 Histoire démographique : données observées	162
3.3 Histoire démographique : inférences	163
4. Discussion	170
CHAPITRE 6 : Architecture génomique de l'isolement reproductif	175
1. Introduction	177
2. Matériel et méthodes	179
2.1 Cartes génétiques et polymorphisme chromosomique	179
2.2 Architecture génomique de l'isolement reproductif	185
3. Résultats	186
3.1 Polymorphisme chromosomique	186
3.2 Scans génomiques	196
4. Discussion	202
4.1 Structure des groupes de liaison	202
4.2 Distribution génomique de la différenciation génétique entre espèces	203
4.3 Perspectives	207
CONCLUSION & PERSPECTIVES	209
Thème 1 : Evènements passés et flux de gènes actuels	211
Thème 2 : Barrières à la reproduction	214
Thème 3 : Architecture génomique et forces évolutives en action	217
Conclusion générale	221
ANNEXES	223
ANNEXE 1 : Construction d'une banque ddRAD	224
ANNEXE 2 : Analyses bio-informatiques des séquences d'ADN	237
ANNEXE 3 : Scripts relatifs à DADI	250
ANNEXE 4 : Comparaison de spectres joints dépliés vs repliés	253
REFERENCES	259

LISTE DES FIGURES

<i>Figure I.1 : Barrières à la reproduction.</i>	8
<i>Figure I.2 : Représentation schématique de différents types de réarrangements chromosomiques.</i>	12
<i>Figure I.3 : Représentation schématique de DMI</i>	12
<i>Figure I.4 : Représentation des différents phénotypes induits par Wolbachia</i>	14
<i>Figure I.5 : Représentation schématique de l'hybridation, pouvant mener à une introgression adaptative ou à une spéciation hybride.</i>	16
<i>Figure I.6 : Exemple des deux structures de zones hybrides les plus documentées</i>	17
<i>Figure I.7 : Balance entre sélection et taux de recombinaison.</i>	18
<i>Figure I.8 : Couplage spatial</i>	19
<i>Figure I.9 : Relations cladogénétiques entre les Jaera sp.</i>	22
<i>Figure I.10 : Description des habitats étudiés</i>	23
<i>Figure I.11 : Photo d'un mâle J. albifrons.</i>	24
<i>Figure I.12 : Représentation d'une femelle et d'un mâle du complexe Jaera albifrons).</i>	24
<i>Figure I.13 : Dessins des différents péreiopodes dles espèces du complexe Jaera albifrons.</i>	25
<i>Figure I.14 : Arbre de type Neighbour-joining obtenu à partir de profils AFLP.</i>	27
<i>Figure I.15 : Arbre de phylogénie (16S)</i>	28
<i>Figure I.16 : Photos de péreiopodes</i>	30
<i>Figure I.17 : Processus pouvant être impliqués dans la spéciation.</i>	31
<i>Figure 2.1 : Sites d'échantillonnage toutes espèces confondues.</i>	77
<i>Figure 2.2 : Sites d'échantillonnage des espèces Jaera forsmanni et Jaera ischiosetosa.</i>	78
<i>Figure 2.3 : Analyses bio-informatiques effectuées.</i>	80
<i>Figure 2.4 : Distribution de la profondeur de couverture moyennée sur l'ensemble des individus</i>	81
<i>Figure 2.5 : Distribution de la profondeur de couverture moyenne obtenue sur l'ensemble des loci.</i>	81
<i>Figure 2.6 : Sites d'échantillonnage des espèces Jaera albifrons et Jaera praeirsuta.</i>	84
<i>Figure 2.7 : ACP obtenues sur l'ensemble des individus</i>	85
<i>Figure 2.8 : Arbre de relations entre les 4 espèces du complexe Jaera albifrons des côtes européennes</i>	86
<i>Figure 2.9 : ACP réalisée sur l'ensemble des individus.</i>	88

Figure 3.1 : Principe général des différents types de croisements réalisés en laboratoire.....	102
Figure 3.2 : Protocole d'élevage en laboratoire.....	103
Figure 3.3 : Actions possibles lors de la rencontre entre le mâle et la femelle.....	104
Figure 3.4 : Délai d'obtention des juvéniles.....	107
Figure 3.5 : Taille des portées.....	108
Figure 3.6 : Survie à 35 jours.....	110
Figure 3.7 : Survie à 70 jours.....	110
Figure 3.8 : Temps d'obtention des juvéniles par portées.....	111
Figure 3.9 : Nombre de descendants à maturité.....	114
Figure 4.1 : Sex-ratio au sein des quatre espèces européennes du complexe <i>Jaera albifrons</i> ...	132
Figure 4.2 : Distribution du sex-ratio chez les descendants des élevages.....	133
Figure 4.3 : Electrophorèse des produits d'amplification.....	135
Figure 4.4 : Nombre d'individus testés, séquencés ou pour lesquels <i>Wolbachia</i> a été détectée.	137
Figure 4.5 : Arbre construit par la méthode du neighbor-joining.....	139
Figure 5.1 : Sites d'échantillonnage en Bretagne et Normandie des individus utilisés pour les inférences démographiques.....	151
Figure 5.2 : Pipeline de traitement des données avant utilisation de <i>daði</i>	153
Figure 5.3 : Modèles démographiques simulés.....	157
Figure 5.4 : Distribution du polymorphisme utilisé pour inférer l'histoire de la divergence....	158
Figure 5.5 : Distribution de la diversité génétique totale.....	159
Figure 5.6 : Distribution de la diversité génétique intra-espèce.....	160
Figure 5.7 : Distribution de la différenciation génétique mesurée à chaque locus.....	161
Figure 5.8 : Distribution de la différenciation génétique mesurée aux loci en commun.....	162
Figure 5.9 : Comparaison des spectres joints des fréquences alléliques en Bretagne.....	164
Figure 5.10 : Comparaison des spectres joints des fréquences alléliques en Normandie.....	165
Figure 5.11 : Comparaison des résidus.....	166
Figure 5.12 : Représentation schématique de l'asymétrie de migration le long du génome....	167
Figure 6.1 : Etapes des traitements des données : cartes génétiques de liaison.....	184
Figure 6.2 : Exemple de carte de référence.....	187
Figure 6.3 : Bar plot représentant les cartes génétiques chez <i>J. albifrons</i>	190
Figure 6.4 : Bar plot représentant les cartes génétiques chez <i>J. praehirsuta</i>	190
Figure 6.5 : Bar plot représentant les cartes génétiques chez les femelles.....	191
Figure 6.6 : Bar plot représentant les cartes génétiques chez les mâles.....	191
Figure 6.7 : Circle plot : <i>J. albifrons</i>	192
Figure 6.8 : Circle plot : <i>J. praehirsuta</i>	193
Figure 6.9 : Circle plot : femelles.....	194
Figure 6.10 : Circle plot : mâles.....	195

Figure 6.11 : Distribution de la différenciation génétique de 200 000 SNPs simulés	197
Figure 6.12 : Distribution de la différenciation génétique observée à 31 474 SNPs	198
Figure 6.13 : Représentation de la différenciation génétique interspécifique.	199
Figure 6.14 : Scan génomique en Bretagne basé sur la carte génétique de la femelle <i>J. albifrons</i>	200
Figure 6.15 : Scan génomique en Normandie basé sur la carte génétique de la femelle <i>J. praehirsuta</i>	201
Figure D.1 : Schéma de l'introgession le long d'un gradient géographique	213
Figure D.2 : Croisement interspécifique en Bretagne dans l'hypothèse d'une infection unidirectionnelle par <i>Wolbachia</i>	215
Figure D.3 : Processus pouvant façonner le paysage génomique.....	218
Figure D.4 : Chaines d'association de traits impliqués dans des barrières à la reproduction.	220

LISTE DES TABLEAUX

<i>Tableau 2.1. Sites échantillonnés et individus prélevés.....</i>	<i>76</i>
<i>Tableau 2.2. Sites d'échantillonnage des populations et noms associés pour les analyses</i>	<i>82</i>
<i>Tableau 3.1. Quantification de l'isolement reproductif en Bretagne et Normandie.....</i>	<i>106</i>
<i>Tableau 4.1. Description des 6 loci testés dans la recherche de Wolbachia.....</i>	<i>127</i>
<i>Tableau 4.2. Individus testés pour Wolbachia.....</i>	<i>128</i>
<i>Tableau 4.3. Espèces d'isopodes marins et d'amphipodes marins testés pour Wolbachia.</i>	<i>129</i>
<i>Tableau 4.4. Composition du mix de PCR pour chacune des cinq paires d'amorces testées.....</i>	<i>130</i>
<i>Tableau 4.5. Détail des programmes de la PCR.....</i>	<i>130</i>
<i>Tableau 4.6. Sex-ratio des individus du complexe Jaera albifrons provenant de populations naturelles.....</i>	<i>131</i>
<i>Tableau 4.7 : Nombre d'individus utilisés dans le calcul du sex-ratio.....</i>	<i>132</i>
<i>Tableau 4.8. Individus séquencés et détectés pour Wolbachia est donné entre parenthèses.....</i>	<i>138</i>
<i>Tableau 5.1. Distribution de la diversité génétique.....</i>	<i>159</i>
<i>Tableau 5.2. Valeurs des paramètres des modèles ayant le meilleur AIC.....</i>	<i>169</i>
<i>Tableau 6.1. Description des familles obtenues lors de croisements intraspécifiques contrôlés</i>	<i>180</i>
<i>Tableau 6.2. Nombre de loci génotypés dans le cadre de la construction de cartes de liaison..</i>	<i>183</i>
<i>Tableau 6.3. Résumé des cartes génétiques de référence</i>	<i>187</i>

INTRODUCTION GENERALE

Au cours de cette thèse je me suis intéressée à l'isolement reproductif entre deux espèces d'isopodes marins, *Jaera albifrons* et *Jaera praeahirsuta*, par une analyse comparative de deux situations de sympatrie dans lesquelles ces deux espèces montrent une différenciation interspécifique contrastée : dans un cas hybridation introgressive, et dans l'autre isolement fort. De par leurs situations sur plusieurs plans du continuum de la spéciation, ces espèces et les zones étudiées au cours de ce manuscrit offrent une bonne opportunité pour l'étude des barrières impliquées dans l'isolement reproductif.

Je tacherai dans la partie introduction qui va suivre, de i) donner le contexte conceptuel du travail fourni durant cette thèse à travers un bref état de l'art des patrons, processus et mécanismes pouvant participer à la spéciation, et plus particulièrement ceux qui seront abordés au cours de cette thèse, puis de ii) présenter les modèles biologiques qui ont fait l'objet de ce travail de thèse ainsi que iii) d'introduire les grands thèmes développés au long de ce manuscrit.

1. De la spéciation à l'isolement reproductif complet

1.1 Spéciation : définition et mécanismes

Une espèce biologique est définie selon Mayr (1963) comme une population ou un ensemble de populations naturelles pouvant engendrer une descendance viable et féconde, génétiquement isolées d'autres groupes similaires. Elle est définie dans le temps par son origine et un évènement d'extinction ou une nouvelle spéciation (Thomas et al., 2010). La spéciation est un processus évolutif de séparation d'une population en deux populations filles, impliquant une diminution des flux de gènes conduisant à un isolement reproductif total de chacune des deux nouvelles populations (Coyne & Orr, 2004). Les mécanismes permettant cet isolement reproductif par la réduction de flux de gènes entre populations sont nommés barrières à la reproduction et sont nombreux et divers (*e.g.*, isolement physique, sélection de partenaire, ou encore isolement génétique). Cette définition trouve ses limites lorsque une interfécondité est possible (Abbott et al., 2013). Il est aujourd'hui admis que la spéciation peut avoir lieu dans tout contexte spatial, où les populations se retrouvent sur un continuum de la panmixie à l'isolement complet, dit continuum de spéciation (Seehausen et al., 2014). Trois types de spéciation ont été historiquement décrits :

- La spéciation avec absence de barrières physiques aux flux de gènes (*i.e.* sympatrique)

La spéciation sympatrique est la divergence de deux populations étant seulement due aux variants génétiques sans intervention d'obstacles extrinsèques, autrement dit d'obstacles géographiques ou physiques (Kondrashov & Mina, 1986). Elle est théoriquement liée à une sélection disruptive entre deux optimums distincts (Thomas et al., 2010). Cette sélection différentielle peut résulter d'une compétition pour les ressources ou d'un coût intrinsèque à l'adaptation ou par sélection sexuelle (Thomas et al., 2010). Les cichlidés *Amphilophus sp.* peuvent être pris pour exemple de spéciation sympatrique puisqu'après colonisation d'un lac par une espèce mère, deux espèces filles ont divergé jusqu'à devenir isolées l'une de l'autre par adaptation à des niches différentes et par homogamie (Barluenga et al., 2006).

- La spéciation avec flux de gènes partiels (*i.e.* parapatrique)

La spéciation parapatrique est difficile à démontrer par rapport à une spéciation sympatrique, car il y a une hétérogénéité de pressions de sélection entre des populations discrètes s'étant agrégées dans des micro-habitats différents ou ayant colonisé une niche adjacente et ayant un faible taux de migration. Il y a donc une variation adaptative discontinue avec isolement reproductif faible. On retrouve ce mode de spéciation chez les organismes planctoniques, ou encore les macroinvertébrés d'eau peu profonde du Pacifique Indo-Ouest (Briggs, 1999) et ce mode est possible dans le cas de flux de gènes limités et d'une contre sélection des migrants ou des génotypes hybrides (Bank et al., 2012).

- La spéciation avec absence complète de flux de gènes (*i.e.* allopatrique, ou péripatrique)

Il ne faut ici pas confondre allopatrie et isolement écologique. L'isolement écologique est intrinsèque, *i.e.* le génotype est adapté à l'habitat, tandis que la spéciation en allopatrie résulte d'un élément extrinsèque tel que des barrières géographiques *e.g.* rivière, montagnes ou encore vecteurs pour certains parasites (Giraud, 2006). Ces éléments vont séparer les populations en deux et des effets de dérive génétique (tirage aléatoire parmi tous les génotypes possibles ; plus la taille de la population est petite, plus la dérive est un phénomène important) ou de pressions de sélections différentielles (si environnements biotiques ou abiotiques différents) vont créer un isolement reproductif entre les deux populations, menant à une forte divergence.

La spéciation péripatrique est un cas particulier de spéciation sans flux de gènes. Il s'agit en général de la colonisation d'une nouvelle niche par une petite population, associée à forte dérive par effet fondateur conduisant à une forte homogamie et une sélection contre les migrants et hybrides.

Le contexte spatial participe à déterminer la mesure de l'isolement extrinsèque entre des populations pouvant diverger ou non. Dans ce contexte spatial, les forces évolutives sont les mutations, la dérive, la migration, la sélection naturelle et la sélection sexuelle, qui peuvent générer un isolement intrinsèque en interagissant ensemble (Butlin et al., 2008). Ces différents milieux dans lesquels évoluent les espèces sont donc sources d'adaptation pour celles-ci et peuvent conduire à l'apparition de différentes barrières selon les situations et le contexte. Parfois des périodes d'isolement (sans flux de gènes) peuvent être alternées avec des périodes où des flux de gènes sont possibles entre populations par un jeu de processus démographiques et migratoires, au cours desquels l'architecture du génome va être marquée. Certaines espèces peuvent également se retrouver sur plusieurs plans de ce continuum de spéciation. C'est le cas des gastéropodes marins *Littorina sp.* qui sont des espèces polymorphes, certaines vivant en sympatrie, d'autres étant localement parapatricues. Certaines espèces ont développé une divergence parallèle en réponse au gradient environnemental sur l'estran (notamment dans la forme des coquilles), et le taux d'avortement de l'embryon des formes intermédiaires est élevé (Butlin et al., 2008).

L'environnement marin est un milieu à forts contrastes tant au niveau des cycles de vie des différentes espèces (sessiles vs dispersives) qu'au niveau géographique (*e.g.* zone d'estran où les espèces sessiles doivent faire face à une émergence vs zone subtidale) (Bierne et al., 2003a, Zardi et al., 2011). C'est un milieu qui peut sembler ouvert avec un fort potentiel de dispersion, ce qui devrait favoriser le flux de gènes entre espèces. Cependant, on retrouve une grande diversité d'habitats avec souvent une adaptation des individus à ces microenvironnements. La diminution des flux de gènes permettant l'adaptation locale des espèces dans un milieu dispersif à large échelle représente un paradoxe nommé paradoxe de la spéciation marine et est une condition intéressante dans le cadre de la spéciation. Ce phénomène a par exemple été discuté par Bierne et ses collaborateurs (2003a) à travers une étude sur des barrières entre deux espèces de moules sympatriques. Une variation d'origine génétique explique dans cette étude la ségrégation écologique (patches à petite échelle définis par les larves en fonction de leur génotype) ainsi que l'asynchronisme des périodes de fraie/ponte. La multiplicité et l'intensité de ces mécanismes sont la clé du paradoxe de spéciation marine, mais parfois un contact secondaire peut mettre fin au processus de spéciation si les mécanismes d'isolement ne sont pas suffisamment forts.

1.2 Spéciation primaire vs contact secondaire

La spéciation avec contact secondaire implique un évènement de divergence unique, suivi par un ou plusieurs évènements de contact entre espèces divergentes avec des degrés variables de différenciation génétique dus à une introgression génomique différentielle et dont les résultantes sont des barrières semi-perméables aux flux de gènes (Barton & Hewitt, 1985, Harrison & Larson, 2016). Les périodes glaciaires ont notamment entraîné l'installation de beaucoup d'espèces dans des zones de refuge, et auraient ainsi permis la reprise de flux de gènes entre espèces lors de nouvelles rencontres durant les périodes interglaciaires (Bernatchez & Wilson, 1998, Hewitt, 2001, April et al., 2013).

Les mécanismes dits de «spéciation avec flux de gènes» englobent divers modes de spéciation de la divergence primaire en sympatrie à l'évolution de barrières post-contact secondaire (Smadja & Butlin, 2011, Gagnaire et al., 2013). Ces deux processus peuvent produire un même pattern de différenciation génomique interspécifique (mesures de type F_{ST}) et de réduction de la diversité intraspécifique dû à de la sélection d'arrière-plan ou de la sélection positive (Noor & Bennett, 2009, Cruickshank & Hahn, 2014). La distinction entre spéciation primaire vs secondaire peut donc être difficile, d'autant plus que plusieurs facteurs peuvent influencer sur les patterns de différenciation. Ainsi, la sélection naturelle affecte la différenciation à des mutations ciblées et à leurs variants neutres proches, induisant une hétérogénéité de la différenciation des loci le long du génome (Charlesworth et al., 1997), sans action d'un quelconque flux de gènes interspécifiques (Noor & Bennett, 2009, Cruickshank & Hahn, 2014, Burri et al., 2015). Le triage incomplet des lignées, c'est-à-dire le polymorphisme ancestral partagé maintenu sous l'effet de la sélection, ou non éliminé par la dérive génétique peut lui être confondu avec des signes d'introgression (Ravinet et al., 2017). Dans le cas relativement rare d'une évolution parallèle, un même allèle doit apparaître par mutation, présenter un caractère adaptatif et augmenter en fréquence dans plusieurs populations, jusqu'à être maintenu dans les populations contemporaines (Wood et al., 2005). Des études comparées d'architecture génétique sous-jacente à l'isolement reproductif peuvent donc fournir des informations importantes sur la cohérence de l'isolement reproductif au sein des espèces (Wagner & Mandeville, 2017, Mandeville et al., 2015).

1.3 Histoires démographiques

Le contexte biogéographique dans lequel deux espèces ont divergé est donc un aspect important (et très débattu) de la spéciation, qui peut être déterminé si l'on parvient à reconstituer l'histoire de la divergence entre espèces en retraçant les processus migratoires et démographiques qui ont pu avoir lieu, grâce aux empreintes laissées sur le polymorphisme du génome et pouvant être identifiées dans une certaine mesure (Sousa & Hey, 2013). Ces empreintes peuvent être comparées à des représentations simplifiées de différents scénarios évolutifs intégrant des effets ayant pu influencer les processus démographiques tels que la variation du taux de migration et de la taille efficace le long du génome (Gutenkunst et al., 2009). Ces scénarios dans leur forme la plus simple imitent un isolement strict, un isolement avec migration, un isolement avec migration ancestrale ou un contact secondaire.

1.4 Barrières à la reproduction

Les mécanismes permettant l'isolement reproductif par la réduction de flux de gènes entre populations sont nommés barrières à la reproduction et sont nombreux, divers (*e.g.*, isolement physique, sélection de partenaire, ou encore isolement génétique) et peuvent intervenir à différents stades du cycle de vie (Figure I.1). La compréhension de ces mécanismes d'isolement reproductif est nécessaire pour appréhender la divergence entre espèces.

Ainsi Coyne and Orr (2004) proposent de classer les barrières en trois grandes classes : i) les barrières à la reproduction pré-copulatoires, telles que l'isolement comportemental, écologique ou encore mécanique ; ii) les barrières prézygotiques post-copulatoires, en particulier de type gamétique ; et iii) les barrières post-zygotiques de type extrinsèques (mauvaise adaptation à la niche écologique, ou "stérilité comportementale") ou intrinsèques (hybrides souffrant d'une viabilité ou d'une fertilité réduite). L'isolement reproductif peut être composé de plusieurs de ces barrières. Ainsi, l'hybridation entre les espèces *Mytilus edulis* et *M. galloprovincialis* en Europe, est régulée par une fécondation préférentielle (Bierne et al., 2002), une sélection par l'habitat (Bierne et al., 2003a) et une contre-sélection des hybrides (Bierne et al., 2006).

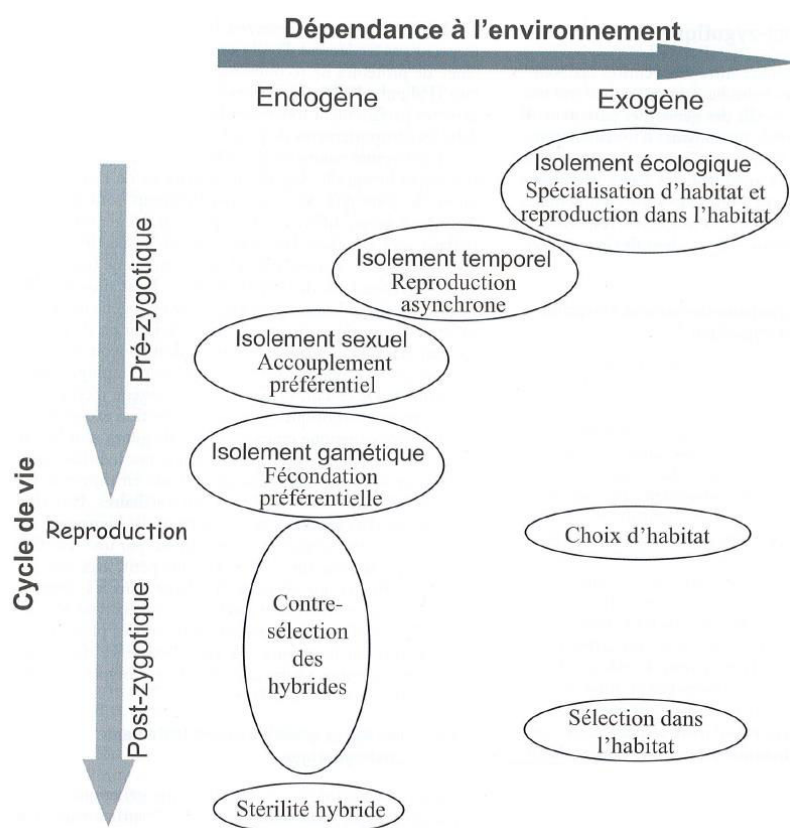


Figure I.1 : Barrières à la reproduction en fonction de la dépendance à l'habitat et du cycle de vie (Thomas et al., 2010).

Isolement pré-zygotique

On appelle barrière pré-zygotique tout obstacle à la reproduction ayant lieu avant la formation de zygotes. Ces barrières soutenues par la sélection naturelle sont nombreuses et diverses. On peut citer par exemple l'isolement comportemental (choix de l'habitat, choix du partenaire), temporel (fécondation décalée), mécanique (phénotypes non compatibles pour la copulation), ou encore gamétique. Elles sont néanmoins généralement classées dans 2 groupes : pré-copulatoire (*e.g.*, isolement écologique) ou post copulatoire (*e.g.*, isolement gamétique). Je n'aborderai ici à titre d'exemple que l'isolement écologique, temporel, comportemental et gamétique.

L'isolement écologique est une barrière qui a lieu *via* une sélection divergente de gènes ou traits particuliers entre populations conduisant à des adaptations aux ressources ou milieux différentes. Il s'agit donc d'un processus limitant le flux de migrants, l'adaptation d'hybrides ou encore le choix de partenaires se trouvant dans un même environnement. Un exemple est le stress thermique du au balancement des marées et la résistance à la dessiccation des organismes sessiles ou sédentaires vivant sur l'estran. Zardi et al. (2011) ont ainsi montré que

trois espèces d'algue (*Fucus sp.*) ont des morphotypes adaptés à une plus ou moins grande tolérance à l'émersion selon leur emplacement sur la zone de balancement des marées. Enfin, le nombre et la qualité des partenaires vont varier selon l'habitat choisi, contribuant donc à l'isolement reproductif. L'isolement temporel est défini comme un asynchronisme de la période de reproduction entre deux espèces, par exemple lors de l'émission des gamètes pour les espèces à fécondation externe. Par exemple, la zone intertidale pose des contraintes sur la synchronisation de la fertilisation externe liée au balancement des marées, car les gamètes sont rapidement dilués dans le milieu et leur mélange est bien sûr empêché en phase d'émersion. La fécondation externe chez les espèces intertidales est donc limitée par une fenêtre d'opportunités due aux cycles, marées, saisons (Monteiro et al., 2012).

L'isolement comportemental est basé sur le choix préférentiel pour un partenaire conspecific reconnu sur la base de critères phénotypiques. Par exemple chez les insectes, la communication chimique est primordiale, et chaque espèce de drosophile crée un bouquet hormonal qui lui est propre (Ferveur, 2005).

Ce processus évolutif peut être entraîné par la sélection sexuelle qui est alimentée par de rares mutations, ce qui va introduire de la variation dans la capacité de choix des femelles, et dans les caractères sexuels secondaires des mâles, favorisant l'évolution des préférences envers des ornements qui peuvent signaler entre autres une meilleure adaptation locale ou une bonne santé si le signal est honnête, et qui améliore ainsi la force de la sélection (van Doorn et al., 2009). D'autres mécanismes de sélection sexuelle (tels que le processus de Fisher-Lande, également appelé *runaway*) peuvent également entraîner une divergence entre espèces. Un cas particulier est le conflit sexuel qui peut mener à une rapide évolution des barrières à la reproduction. Généralement une reproduction pour être efficace nécessite que mâles et femelles aient des traits relativement proches pour se lier. Cependant, le conflit sexuel a lieu quand l'augmentation des chances de se reproduire pour l'un des deux sexes diminue la fitness, voire conduit à la mort des individus de l'autre sexe. Gavrillets (2000) a montré en modélisant les dynamiques de coévolution entre les traits liés à la reproduction chez mâles et femelles qu'une rapide évolution des mécanismes d'isolement reproductif conduite par le conflit sexuel peut avoir lieu lorsque les taux de spéciation augmentent après une colonisation de nouveaux habitats. Il y a une coévolution dans laquelle les femelles essaient de diminuer le taux de rencontre et les mâles tentent de l'augmenter. En général, l'accouplement avec un partenaire adapté au même habitat est bénéfique car augmente la probabilité pour la descendance d'avoir un phénotype optimal (Van Doorn *et al.*, 2009). La sélection naturelle et la sélection sexuelle associées permettent donc un isolement reproductif et une adaptation locale malgré un flux de gènes parfois important (van Doorn et al., 2009).

Enfin, une variation de compatibilité entre gamètes peut également être source d'isolement. Deux individus ou gamètes sont compatibles lorsque l'accouplement et la fertilisation ne sont pas empêchés par des mécanismes d'isolement (Gavrilets, 2000). Suzuki and Fukami (2012) ont montré un isolement génétique et reproductif intraspécifique entre deux morphotypes du corail *Acropora solitaryensis*, dû à une incompatibilité gamétique entre les deux formes. Il y a donc un fort isolement pré-zygotique et un isolement inter-espèce dus à une incompatibilité unidirectionnelle (diminution de fertilité) entre sperme et œufs. Cet isolement peut être renforcé par d'autres barrières, comme ici par une asynchronie des pontes et une ségrégation environnementale.

Isolement post-zygotique

Les barrières post-zygotiques interviennent après la fécondation en limitant l'homogénéisation du flux de gènes. Elles sont de l'ordre des mutations génétiques, causant une mauvaise expression ou l'absence d'expression des gènes. On parlera ici d'une part des incompatibilités génétiques en général (réarrangements chromosomiques, incompatibilités de Dobzhansky-Muller), et d'autre part de deux mécanismes pouvant influencer sur l'isolement global : le renforcement et l'introgression.

La dépression hybride intervient lorsque la valeur sélective des génotypes hybrides est inférieure à celles des génotypes parentaux (Thomas et al., 2010). Si cette valeur sélective est inférieure sur l'ensemble des environnements utilisés par les morphes parentaux alors il y a sélection exogène. Un isolement postzygotique extrinsèque basé sur une sélection divergente a par exemple été mis en évidence chez les épinoches à trois épines (Gow et al., 2007). Une inadéquation entre phénotype hybride et environnement a été notée, conduisant à une diminution du nombre d'hybrides tout au long des différentes phases du cycle de développement, et s'intensifiant au fil des générations.

L'isolement post-zygotique intrinsèque peut être classé en trois catégories temporelles selon (Dobzhansky, 1970). La létalité (les zygotes hybrides ne sont pas capables de survivre dans leur environnement ou n'atteignent pas la maturité sexuelle) et la stérilité (les hybrides ne produisent pas de gamètes fonctionnels ou ne sont pas en mesure de produire les signaux comportementaux nécessaires pour permettre leur reproduction) des hybrides de première génération ainsi que la dégénérescence hybride si la réduction de la valeur sélective apparaît dans les générations suivantes.

La règle de Haldane (1922) prédit que seul l'un des deux sexes, généralement celui qui est hétérogamétique, va s'avérer létal ou stérile. Cela peut être lié en particulier à la dominance des allèles (les allèles responsables de la stérilité et/ou de la contre-sélection sont récessifs).

Plusieurs phénomènes génétiques peuvent induire des incompatibilités et réductions de la fitness hybride. Les substitutions, inversions chromosomiques, ou encore les transpositions peuvent être des facteurs de la stérilité hybride, car des copies de gènes essentiels peuvent venir à manquer, ou il peut y avoir une divergence des gènes entre populations ou espèces due à une diminution de la recombinaison. Des interactions épistatiques délétères, spécifiques à l'hybride, entre les gènes qui ont divergé, peuvent être en cause dans les incompatibilités (Fierst & Hansen, 2010).

- Remaniements chromosomiques

Les remaniements chromosomiques modifient la position et donc les distances des locus sur le caryotype ce qui modifie le taux de recombinaison chez les hétérozygotes (Navarro & Barton, 2003). Chez ces derniers, lors de translocations et d'inversions, le crossing-over de certaines régions des chromosomes crée des lacunes dans le contenu de gènes et l'accumulation de tels remaniements entraîne l'isolement post-zygotique. La force d'un réarrangement délétère agissant comme barrière de reproduction est proportionnelle au niveau de sous-dominance, mais le taux de fixation d'un réarrangement est inversement proportionnel au niveau de sous-dominance. Les obstacles les plus forts sont donc les moins susceptibles d'être fixés dans une population, c'est ce qui est appelé paradoxe de la spéciation chromosomique (Navarro & Barton, 2003).

Les remaniements peuvent aussi agir comme barrière intrinsèque au flux génique par diminution de la recombinaison entraînant une divergence (Thomas et al., 2010).

Les réarrangements se divisent en trois grands groupes (Figure I.2):

- les inversions.

L'inversion d'un segment chromosomique fait suite à une cassure. Une inversion est péricentrique si le segment inversé comprend le centromère, ou bien paracentrique. Lors de l'appariement des régions homologues, une boucle d'inversion va se créer (Figure I.2) et 50% des gamètes produits montreront alors des délétions ou duplications de gènes. Une étude portant sur la drosophile a ainsi montré que la localisation de la majeure partie des traits impliqués dans l'isolement reproductif pouvait être due à une inversion (Noor et al., 2001). Il semblerait également que les espèces en sympatrie montrent de plus grandes différenciations chromosomiques qu'en allopatrie (Noor et al., 2001, Chang & Noor, 2007). Un autre exemple est celui de la mouche de la pomme *Rhagoletis pomonella*, dont les inversions ont permis un certain isolement reproductif et ainsi la formation de nouvelles espèces (Feder et al., 2005).

-les fusions/fissions.

La fusion/fission se fait à partir du centromère de deux chromosomes non homologues, et est due à une mauvaise ségrégation méiotique. La région entourant le centromère étant pauvre en gènes, l'impact des fusions devrait être négligeable (Navarro & Barton, 2003), et nécessiterait beaucoup de remaniements pour être délétère.

- les translocations.

Les translocations non-réciproques consistent en un transfert d'un segment d'un chromosome à l'autre. Elles sont dites réciproques si l'échange est symétrique entre les deux chromosomes.

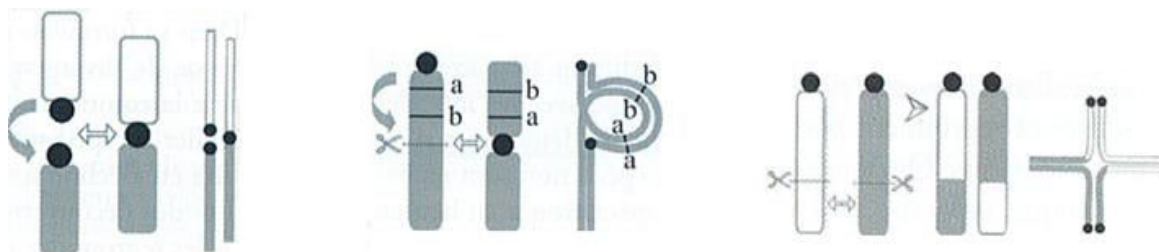
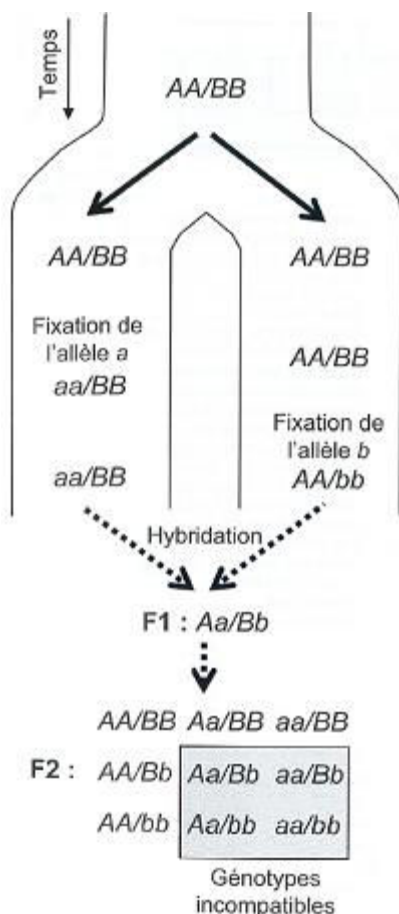


Figure I.2 : Représentation schématique de différents types de réarrangements chromosomiques. a) fusion/fission centrique, b) inversion, c) translocation. D'après King (1993), repris par Thomas et al. (2010).

- Incompatibilités génétiques de type Dobzhansky-Muller



Le modèle d'incompatibilités génétiques de type Dobzhansky-Muller (DMI, Orr, 1995) prédit que dans deux populations divergentes, l'accumulation des interactions épistatiques incompatibles entre allèles à des loci différents participe à l'isolement post-zygotique en causant une dépression hybride (Fierst & Hansen, 2010). L'explication du modèle théorique à « deux-locus » (Figure I.3) est que la divergence de deux populations à partir d'une population ancestrale (génotype aux deux locus AA/BB) va entraîner une fixation différentielle des allèles au cours du temps. Lors de leur remise en contact, l'hybridation produit un génotype de type hybride F1 (*i.e.* hétérozygote aux deux locus).

Figure I.3 : Représentation schématique de l'évolution de l'isolement post-zygotique de type Dobzhansky-Muller dans le cas d'un modèle à deux locus. D'après Thomas et al. (2010).

Or, en F2, la combinaison des deux allèles dérivés récessifs peut s'avérer délétère (soit 4 génotypes incompatibles sur 9 théoriquement possibles). Un exemple connu est le cas des F2 issus de croisements entre les moules *Mytilus edulis* et *M. galloprovincialis* (Bierne et al., 2006). Dans ce cas, ce sont les génotypes homozygotes, *i.e.* aa/bb et AA/BB qui sont contre-sélectionnés.

Le processus est continu, soumis à la sélection naturelle sur plusieurs générations et l'un des moteurs de la divergence entre populations due à des DMI est le facteur environnemental. Si deux populations occupent un environnement identique, et qu'il y a substitutions de mêmes allèles cela ne créera pas de DMI. Le scénario souvent proposé est que les populations allopatriques s'adaptent à leur propre environnement et que la dépression hybride serait un effet secondaire de la pléiotropie de la sélection divergente, les gènes entraînant ces adaptations amenant les incompatibilités. Les DMI se produisent donc plus facilement dans des populations adaptées à des environnements différents, et plus le nombre de loci impliqués dans l'incompatibilité et le nombre de générations augmentent, plus la probabilité d'évolution d'une incompatibilité augmente également (Fierst & Hansen, 2010). De plus, de nouvelles adaptations peuvent être couplées avec d'anciennes barrières intrinsèques, ce qui a pour conséquences de renforcer les barrières écologiques et l'accumulation de barrières intrinsèques entre les populations lors d'un contact secondaire (Abbott et al., 2013).

- Renforcement

Le renforcement est une phase particulière qui survient lorsqu'une fitness hybride réduite conduit à l'évolution d'un isolement pré-zygotique de type choix de partenaire plus fort (Servedio & Noor, 2003), réduisant ainsi le taux d'hybridation. Il peut avoir lieu en présence de flux de gènes pas trop importants (Matute, 2010), et dépend de la force et de la fréquence de la sélection contre les hybrides (Yukilevich, 2012). La signature typique du renforcement est une augmentation de l'isolement comportemental en sympatrie. Les mâles de populations de poissons du genre *Lucania* vivant en sympatrie ont une forte préférence de partenaire intraspécifique, potentiellement due à un mécanisme de renforcement. Cela accentue la spéciation chez *Lucania* et ce, même lorsque les réductions des fitness hybrides sont faibles. La force totale des barrières entraînant l'isolement reproductif est ici estimée aux alentours de 0.997 (1= isolation complète) (Gregorio et al., 2012).

- Rôle de bactéries endosymbiotiques de type *Wolbachia*

Les bactéries peuvent être responsables de phénomènes liés à la reproduction de leurs hôtes. Ainsi *Wolbachia* induit notamment une féminisation des mâles, des incompatibilités cytoplasmiques, une induction de la parthénogénèse ou encore une mortalité des mâles (Figure I.4). La féminisation peut se produire par transformation d'un mâle génétique en femelle phénotypique en bloquant le développement de la glande androgène (Werren et al., 2008). Les mâles génétiques infectés émettent alors des ovules qui doivent être fécondés par des spermatozoïdes de mâle non infecté pour que les descendants soient viables. Concernant les incompatibilités, un mâle infecté doit se croiser avec une femelle infectée par la même souche ou une souche compatible pour que la modification du spermatozoïde due à la bactérie soit levée dans l'ovule et que le zygote puisse se développer normalement. Cette bactérie se transmet principalement de manière verticale *via* les gamètes femelles (Hertig & Wolbach, 1924, Werren et al., 2008), bien que plusieurs études démontrent une transmission également horizontale entre espèces, permettant sa propagation aux différents clades d'invertébrés (Vavre et al., 1999, Cordaux et al., 2001).

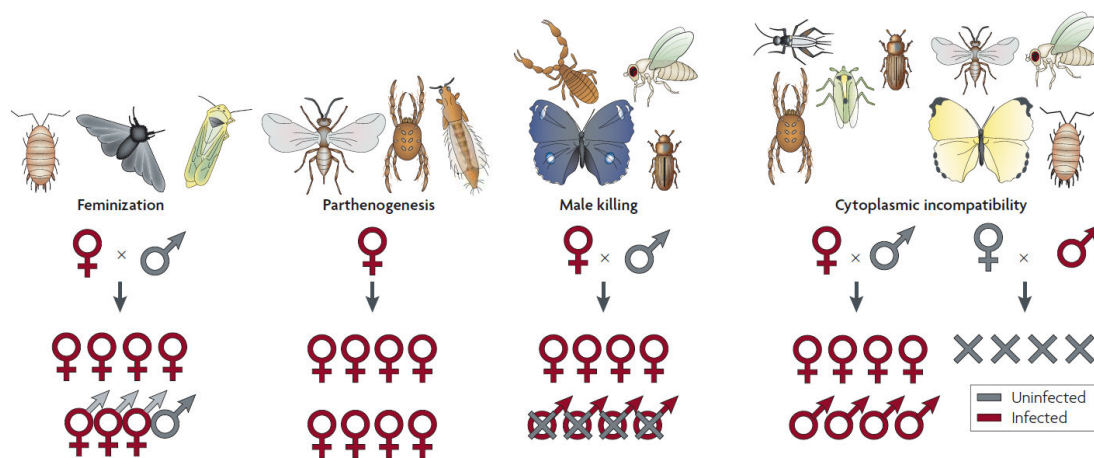


Figure I.4 : Représentation des différents phénotypes induits par *Wolbachia*. La féminisation est trouvée chez les hémiptères, isopodes et lépidoptères ; la parthénogénèse chez les espèces d'acariens, hyménoptères et thysanoptères ; l'élimination des mâles, chez les coléoptères, diptères, lépidoptères et ordre des pseudoscorpions ; les incompatibilités génétiques chez les acariens, coléoptères, diptères, hyménoptères, hémiptères, isopodes, lépidoptères et orthoptères (Werren et al., 2008).

2. Les zones hybrides : des laboratoires naturels

La spéciation est souvent étudiée en milieu naturel dans des populations admixées ou dans des zones hybrides, permettant ainsi l'étude des flux d'allèles entre populations partiellement isolées (Barton & Hewitt, 1985).

Lorsque ces zones hybrides font suite à une situation de contact secondaire, et qu'elles peuvent être comparées à des situations d'isolement reproductif entre les mêmes deux espèces mères, alors elles peuvent être particulièrement utiles à l'étude des mécanismes d'isolement et plus généralement à l'étude de la spéciation (Hewitt, 2001, Harrison & Larson, 2016). Harrison and Larson (2016) ont ainsi suggéré que la compréhension de l'origine et de la dynamique des zones hybrides ne se peut que par la comparaison des régions divergentes dans les populations parentales aux régions montrant une forte différenciation lorsque ces populations interagissent dans des zones hybrides.

Je commencerai par aborder ici les différents types de zones hybrides, que ce soit au niveau des distributions des espèces que des forces évolutives sous-jacentes à ces zones, puis je donnerai quelques pistes pour l'étude de la différenciation interspécifique.

L'hybridation (production de descendants entre deux espèces partiellement isolées) est un phénomène couramment trouvé en milieu naturel, et plus particulièrement entre espèces peu divergentes (Abbott et al., 2013, Mallet, 2005). Elle peut entraîner la formation de nouvelles espèces (spéciation hybride, Mallet, 2007) si l'espèce est plus compétitive (voir pour exemple Hegde et al., 2006), ou à l'opposé être moteur d'une introgression adaptative (Hedrick, 2013), *i.e.* sélection d'allèles conférant un avantage sélectif (sur un trait phénotypique par exemple) transmis par l'autre espèce (voir pour exemple Pardo-Diaz et al., 2012) (Figure 1.5). Dans certaines populations, des cas d'hybridation sans introgression peuvent avoir lieu, s'il y a contre-sélection des hybrides et de leurs descendants. Cela peut dans certaines conditions entraîner un renforcement de l'isolement reproductif sous forme de barrières pré-zygotiques accentuées, tel qu'observé sur des caractères sexuels (Servedio & Noor, 2003, Barton & de Cara, 2009).

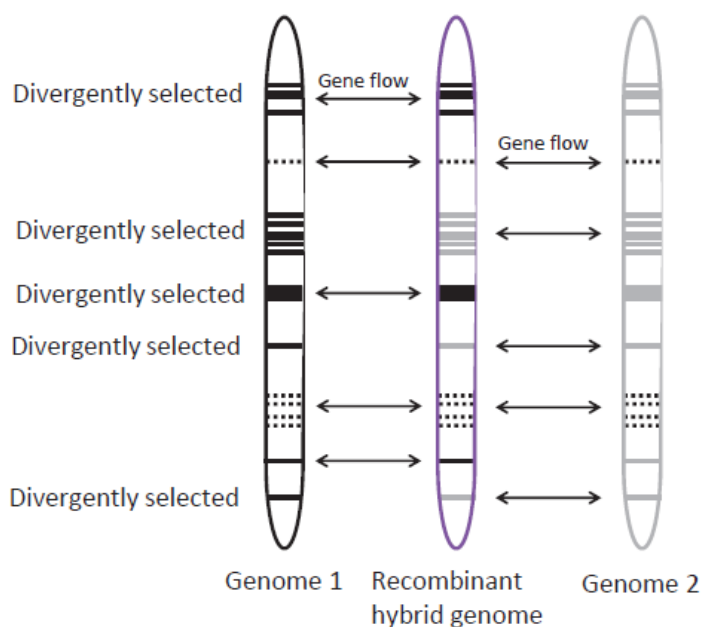


Figure I.5 : Représentation schématique de l’hybridation, pouvant mener à une introgression adaptative ou à une spéciation hybride. Le génotype hybride va accumuler des loci différemment sélectionnés chez les génotypes parentaux. Cette combinaison peut être avantageuse pour le génotype hybride et mener à une spéciation hybride (taxon indépendant) ou peut permettre à une des populations parentales d’évoluer par remplacement dans son génome de traits avantageux provenant de l’autre espèce. Un flux de gènes entre espèces peut avoir lieu par hybridation permettant d’échanger des variants neutres ou sélectionnés. Figure d’après Abbott et al. (2013).

L'hybridation des populations varie non seulement dans la nature de l'isolement et des mécanismes impliqués mais aussi en structure géographique.

Ce que l'on appelle traditionnellement "zone hybride" (exemple Figure I.6A), est une zone d'équilibre entre la migration de génotypes parentaux et la sélection contre les hybrides, aussi appelé « zone de tension », dont le jeu de force affecte l'évolution du taux de recombinaison (Figure I.7) (Ortiz-Barrientos & James, 2017, Noor & Bennett, 2009).

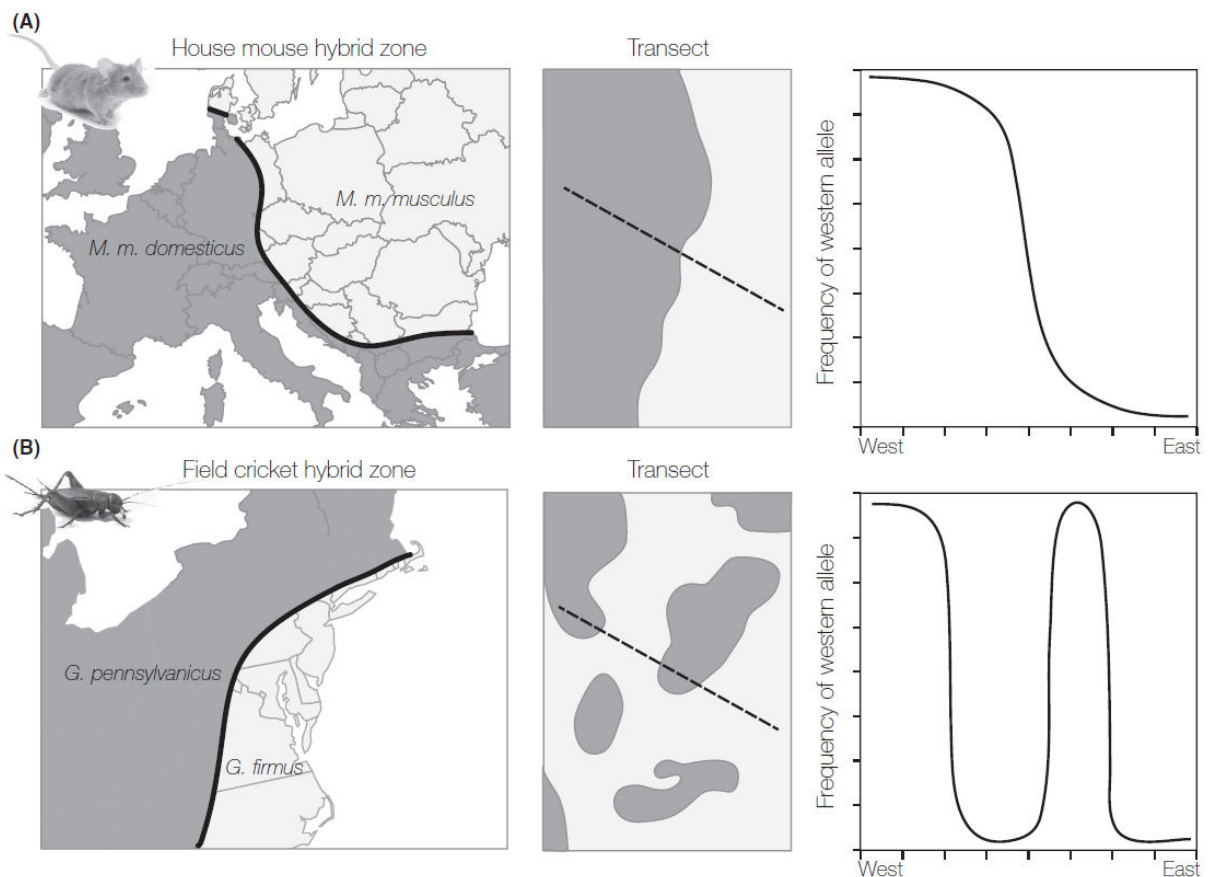


Figure I.6 : Exemple des deux structures de zones hybrides les plus documentées : (A) une zone hybride en cline formée par les deux sous-espèces *Mus musculus musculus* et *M. m. domesticus*. Ces populations interagissent dans une zone centrale à leur aires de distribution respectives, et le transect révèle un gradient continu entre les deux génotypes parentaux de chaque côté et les hybrides multigénérationnels au centre ; (B) une zone en mosaïque entre les espèces de criquets *Gryllus firmus* et *G. pennsylvanicus*. Chaque espèce occupe un habitat différent, distribué en patch, permettant la rencontre à chaque nouveau patch des deux espèces (Harrison & Larson, 2016).

La sélection naturelle peut être stable dans les habitats mais disruptive au niveau de la population entière : les individus à phénotype intermédiaire sont supprimés au fur et à mesure, pendant que d'autres s'installent à chaque migration et recombinaison (van Doorn et al., 2009).

Ces zones hybrides se caractérisent généralement par une structure en cline de fréquences alléliques, et peuvent être dynamiques (Barton & Hewitt, 1985, Buggs, 2007, Wielstra et al., 2017).

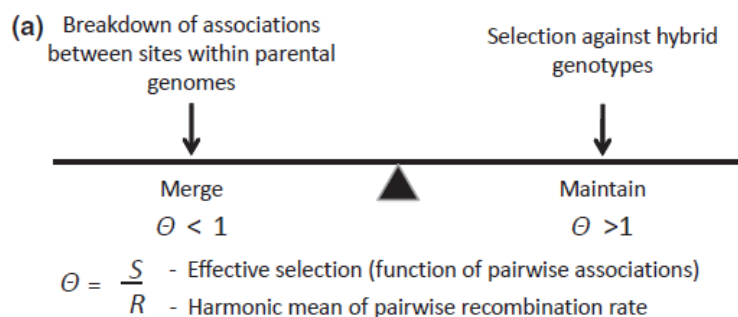


Figure I.7 : Balance entre sélection et taux de recombinaison, pouvant mener à une fusion des populations (à gauche du graphique) ou au maintien de populations sans échanges de flux de gènes. D’après Abbott et al. (2013).

Les zones dites en mosaïques (exemple Figure I.6B) sont décrites comme étant des zones de rencontre à la frontière de deux habitats préférentiellement utilisés par deux espèces distinctes (Harrison & Rand, 1989). La distribution souvent en patch de ces zones est due à la répartition des habitats discrets (par exemple, îles, lacs, plantes hôtes) conduisant à des zones hybrides fragmentaires.

D’autres zones (dite « mottled ») peuvent également résulter en zones en patch bien que non déterminées par l’environnement mais par l’histoire de la colonisation (Hauffe & Searle, 1993), ou par la stochasticité des populations (Gompert et al., 2010).

Ces zones hybrides en patch sont particulièrement intéressantes de par leur structure et la multiplicité des zones, potentiellement indépendantes (Butlin et al., 2014, Rougemont et al., 2016b), pour distinguer des divergences primaires de contact secondaires, notamment en l’absence de variations du contexte environnemental (Gompert et al., 2017, Harrison & Larson, 2014), ainsi qu’améliorer notre compréhension des mécanismes d’isolement et l’architecture génomique qui leur est associée (Harrison & Larson, 2016, Westram et al., 2016).

Les situations où l’on peut comparer des populations composées d’un mélange d’individus de deux espèces qui dans certains cas s’hybrident et dans d’autres cas restent fortement isolées sont particulièrement intrigantes. Un exemple illustratif est le lac Victoria où des populations mixtes de *Pundamilia pundamilia* et *P. nyererei* montrent plus ou moins d’hybridation en fonction de la variation des mécanismes de comportement de choix de partenaires eux-mêmes liés à variations de l’habitat (clarté de l’eau) (Seehausen, 2009). Cependant, il est parfois difficile

d'identifier des barrières à la reproduction, un couplage spatial entre clines exogènes (liés à une barrière environnementale) et endogènes (liés aux incompatibilités génétiques) pouvant avoir lieu (Bierne et al., 2011, voir Figure I.8).

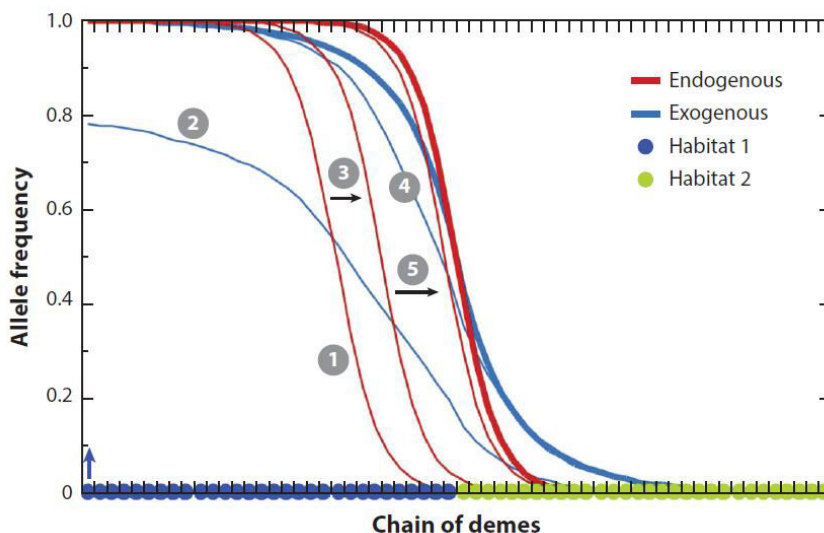


Figure I.8 : Couplage spatial entre un cline d'origine endogène (en rouge) et un cline d'origine exogène (en bleu) à la frontière de deux habitats, divisées en deux parcelles de 30 demeures chacune (Bierne et al., 2011). Le cline endogène est initialement positionné au dème 20 (1). Lors de l'apparition d'un nouvel allèle (cline exogène) dans l'habitat 1, celui-ci va se propager et chevaucher le cline endogène (2). Un déplacement du cline endogène va avoir lieu jusqu'à positionnement le long de la frontière environnementale. Les clines endogènes et exogènes se chevauchant, le couplage se maintient jusqu'à coïncider avec limite environnementale (3-5).
D'après (Gompert et al., 2017).

La comparaison de zones d'hybridation n'impliquant pas de différences environnementales entre les espèces parentales peut permettre d'interpréter les motifs de divergence, par exemple polymorphisme ancestral partagé ou flux de gènes actuel) ou évaluer si les patterns d'introgression différentiels sont dus à une recombinaison hétérogène, de la sélection ou des flux de gènes (Gagnaire et al., 2013; Powell et al., 2013; Rougemont et al., 2016), et de différencier les loci impliqués dans les barrières à la reproduction de ceux de l'adaptation locale.

Les processus d'hybridation peuvent mener à des patterns d'introgression variables et différentiels le long du génome, *i.e.* l'introgression est plus ou moins importante selon les régions du génome tel que montré dans la zone hybride entre *Mus musculus* et *M. domesticus* (Teeter et al., 2010, Janoušek et al., 2012), et laisser ainsi des traces sur le génome et son architecture (Harrison & Larson, 2014). L'introgression sera asymétrique si les allèles de l'une

des espèces parentales confère une plus grande fitness aux hybrides que ceux provenant de l'autre espèce (Hedrick, 2013), ou lorsque la taille de la population effective est plus grande chez une des deux espèces parentales (e.g., Lepais et al., 2009).

Enfin, les flux de gènes entre espèces ont comme conséquence de diminuer la divergence génétique. Cependant, à un niveau intermédiaire, cela peut au contraire augmenter la variation génétique en favorisant la diffusion des mutations bénéfiques et le potentiel adaptatif (Zardi et al., 2011).

Une manière d'observer la différenciation génétique entre espèces s'hybridant, et ainsi de pouvoir détecter des régions génomique faisant barrières aux flux de gènes et potentiellement impliquées dans l'isolement reproductif, est de faire des scans génomiques. La présence d'îlots de divergence (Turner et al., 2005) peut toutefois résulter de processus autres que la sélection entraînant une divergence. Ainsi des variations de taux de recombinaison peuvent être responsables de variations locales de divergence génétique le long du génome, et ce quel que soit le niveau de flux de gènes entre espèces. Une étude a par exemple montré que le génome hétérogène chez les *Ficedula flycatchers* serait une conséquence de la sélection d'arrière-plan et d'un balayage sélectif (Burri et al., 2015). Dans ce contexte, l'étude comparative de zones ayant évolué indépendamment (voir plus haut) se révèle être particulièrement intéressantes pour détecter l'isolement reproductif.

Note sur le vocabulaire

Tout au long de ce manuscrit, le terme « zone hybride » sera utilisé au sens large comme un site ou région où il y a de l'hybridation en cours, notamment en patch, et ne réfère pas nécessairement à une zone en cline. Toutes les populations mixtes, *i.e.* où les individus des deux espèces parentales peuvent être en contact, seront dites en sympatrie. "Syntopie" sera utilisé pour les populations mixtes au sein desquelles les deux espèces partagent le même habitat, et parfois "micro-allopatrie" sera utilisé pour les populations mixtes mais avec deux habitats (mais en aucun cas comme définition générale d'une zone hybride mosaïque pouvant confondre des concepts écologiques et géographiques (Fitzpatrick et al., 2008)).

3. Deux taxons du complexe d'espèce *Jaera albifrons* (Crustaceae) comme modèles d'étude

3.1 Taxonomie

Les individus du complexe *Jaera albifrons* sont des Isopodes Asellotes, de la famille des Janiridae. La taxonomie des *Jaera albifrons* a été revue plusieurs fois. Tout d'abord nommé *Oniscus marinus*, par Fabricius (1780), le genre *Jaera* a été défini par Leach (1813-1814). La description des variants sexuels des mâles a conduit à reconnaître plusieurs formes, d'abord comme races distinctes par Forsman (1949) avant d'être élevées au rang de sous-espèces puis d'espèces par Bocquet (1953). Le nombre d'espèces décrites a quelque peu varié selon les morphotypes décrits, pour finalement s'arrêter à 5 espèces : *Jaera albifrons albifrons*, *Jaera albifrons praehirsuta*, *Jaera albifrons ischiosetosa*, *Jaera albifrons forsmanni*, *Jaera albifrons posthirsuta*. **Pour une lecture simplifiée de mon rapport de thèse, je n'utiliserai comme nom d'espèces que : *Jaera albifrons*, *J. praehirsuta*, *J. ischiosetosa*, *J. forsmanni*, *J. posthirsuta*.** L'appellation *Jaera albifrons* fera donc toujours référence à l'espèce *stricto sensu*, sauf si explicitement spécifié qu'il s'agit du complexe d'espèces dans son ensemble.

Les 5 espèces du complexe *Jaera albifrons* sont des espèces d'isopodes marins, évoluant dans les eaux tempérées et froides des côtes de l'Atlantique Nord. *Jaera albifrons*, *J. praehirsuta* et *J. ischiosetosa* sont retrouvés du Portugal jusqu'au Groenland, ainsi que sur la côte Est du continent Nord-Américain. *J. forsmanni* est endémique de la Manche et de la mer d'Iroise (Naylor, 1972), alors que *J. posthirsuta* n'est retrouvé que sur la côte Est du continent Nord-Américain. Deux groupes d'affinité ont été établis par Cariou (1977) sur la base de distances génétiques (11 allozymes), l'un composé de *J. praehirsuta* et *J. forsmanni*, et le second de *J. ischiosetosa* et *J. albifrons*. Les différentes espèces occupent toutes une bande étroite d'habitat située sur l'estran (cf. ci-dessous) mais qui s'étend sur d'immenses aires de distribution tout autour de l'océan Atlantique Nord, de la mer du Nord, et de la Baltique. Les différentes espèces ont des aires de distribution très chevauchantes, et coexistent ainsi très souvent en populations mixtes d'au moins deux espèces (jusqu'à quatre espèces).

Veille (1979) a replacé le contexte de spéciation des *Jaera sp.* sous la dépendance des événements géologiques contemporains de leur radiation (Figure I.9). Ainsi ce serait à partir de l'individualisation de la mer Paratéthis (15 M.A.) suivi par différents remaniements des côtes avec déplacements de zones climatiques que le complexe *Jaera albifrons* se serait distingué et aurait colonisé la côte Atlantique.

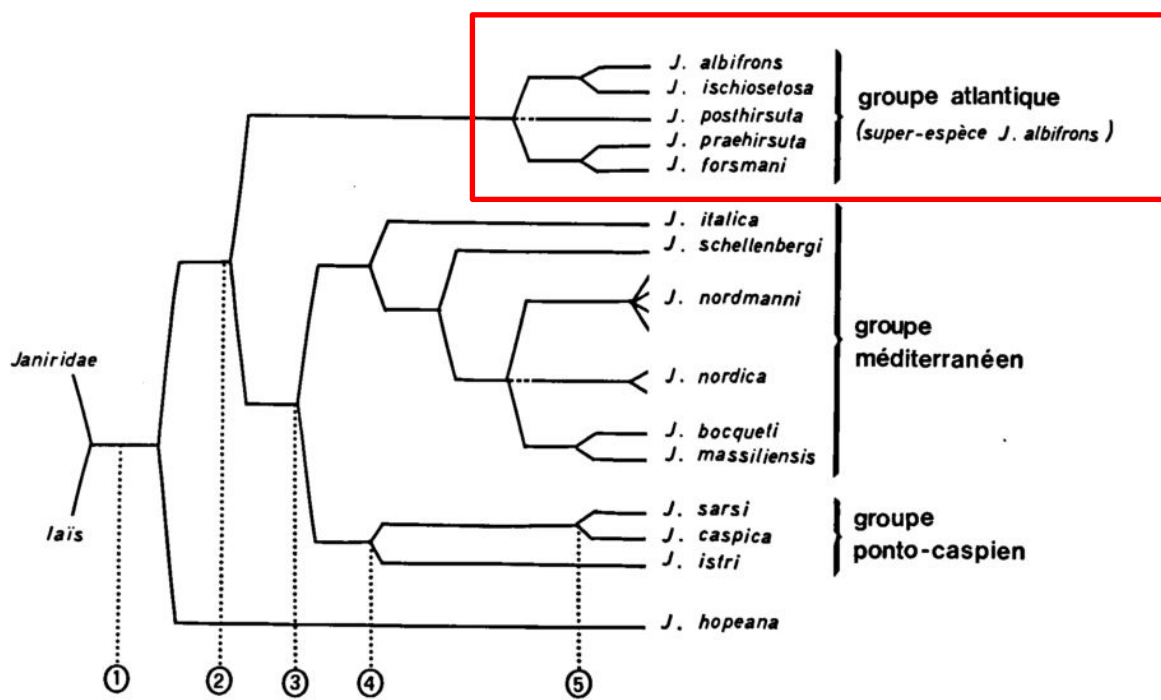


Fig. 10. Relations cladogénétiques vraisemblables entre les *Jaera*, selon la morphologie et la biologie sexuelle, et datations envisagées d'après la biogéographie. Les droites n'ont pas valeur de distance. 1, Isolement de la Méditerranée au Miocène (20 millions d'années A.P.); 2, isolement de la Paratéthys (15 M.A.); 3, formation du „Lago-mare” au Messinien (5,2 M.A.); 4, formation du lac des Carpathes (1,5 M.A.); 5, interglaciaire Riss-Würm. Les événements cladogénétiques situés à la droite du point 4 sont liés aux glaciations.

Figure I.9 : Relations cladogénétiques entre les *Jaera* sp.. D'après (Veuille, 1979).

3.2 Ecologie

Le complexe d'isopodes marins *Jaera albifrons* est composé de cinq espèces à développement direct, ne présentant aucune phase pélagique : les jeunes, dès leur sortie du marsupium maternel, sont comme les adultes, strictement marcheurs. La structure de leurs parties buccales est similaire entre espèces et indique qu'elles se nourrissent en broutant les films bactériens ou d'algues recouvrant les pierres ou le fucus (Jones, 1972b). Certaines espèces montrent un fort biais de sex-ratio en faveur des femelles, jusqu'à 12 femelles pour 1 mâle chez *J. ischiosetosa* à un site au sud des Pays de Galles (Piertney & Carvalho, 1996).

Les espèces du complexe *Jaera albifrons* sont des espèces intertidales, *i.e.* retrouvées sur la zone de balancement des marées, et exploitant les pierres situées le long de l'estran comme habitat, à l'exception de l'espèce *J. prae-hirsuta* qui a généralement une préférence pour les algues et est retrouvée sur la zone basse de l'estran (Bocquet, 1954, Naylor & Haahtela, 1966). Ces espèces semblent avoir des préférences écologiques légèrement différentes, notamment dans l'utilisation de l'étagement de la zone intertidale. Cependant, ces zones pourraient varier d'un site à l'autre en fonction des conditions écologiques locales (Bocquet, 1972, voir également

Figure I.10). La fréquence des hybrides naturels le long de ces zones de contiguïté reste très faible et n'excède guère un pour mille environ (Bocquet, 1972). Ces espèces supportent des variations de salinité importantes, bien que *J. ischioetosa* montre une plus forte résilience que *J. praehirsuta* ou *J. forsmanni* (Jones, 1972a). Par exemple, *J. albifrons* peut survivre 10 heures dans de l'eau à 2 ‰ (Fobes, 1974). Enfin, Bocquet (1972) suppose que ces espèces peuvent être entraînées par les courants côtiers et ainsi coloniser de nouveaux sites.



Figure I.10 : Description des habitats étudiés dans deux régions Française pendant ma thèse : (A) habitat typique de nos zones d'échantillonnage en Bretagne, où les algues et les pierres sont au même endroit ; (B) exemple d'habitat étudié en Normandie, où les *Fucus* forment une bande située plus en bas de l'estran que les pierres (et où ces algues ne sont habitées par aucune espèce du complexe *Jaera albifrons*).

3.3 Comportement sexuel



Figure I.11 : Photo d'un mâle *J. albifrons*

Les mâles (voir Figure I.11) sont reconnaissables des femelles par leur taille plus petite (environ 2 mm vs jusqu'à 5 mm pour la femelle, Figure I.12), par leur préoperculum en forme de T (vs rond pour la femelle) et leur caractères sexuels secondaires (Solignac, 1981, Jones & Fordy, 1971).

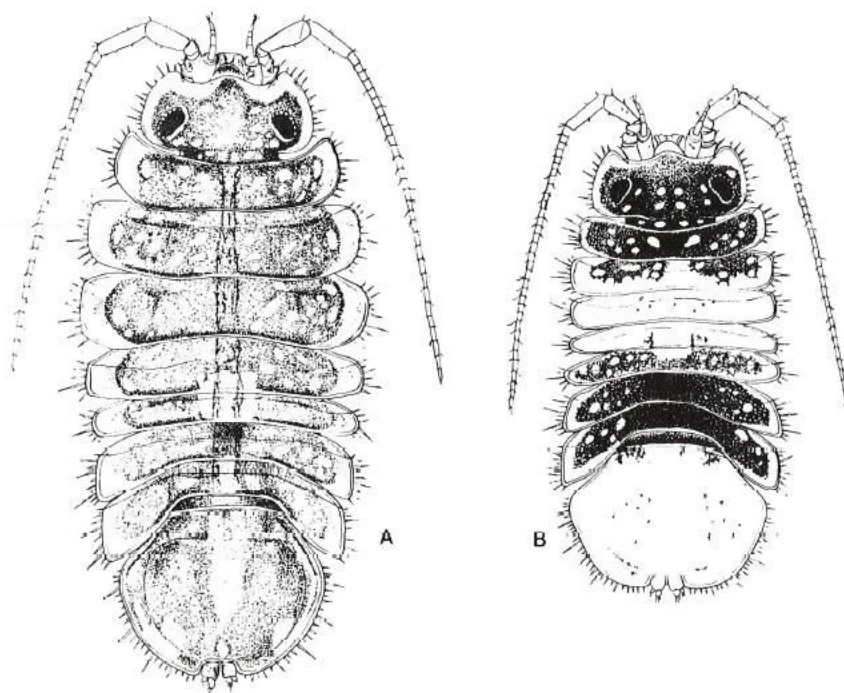


Figure I.12 : (A) femelle et (B) mâle du complexe *Jaera albifrons* (Solignac, 1981).

Ces caractères prennent la forme de soies et d'épines disposées sur les péréiopodes et sont spécifiques à chaque espèce. Les mâles les utilisent lors de la parade sexuelle, qui consiste à monter sur la femelle, se placer tête-bêche et brosser cette zone à l'aide des soies disposées sur leurs péréiopodes. Une exception est faite pour les mâles *J. forsmanni* qui n'ont pas de soies et semblent utiliser les épines pour « courber ». Ainsi, la femelle va pouvoir reconnaître ou non la parade du mâle et l'accepter ou bien se débattre à coup de pleotelson jusqu'à ce que le mâle parte. Sur la Figure I.13 les sept péréiopodes (P1 à P7, de gauche à droite) du côté gauche sont dessinés : A) représente une femelle, B, un mâle *J. forsmanni* ayant des épines (souvent 3/4) sur les P6 et P7, C) un mâle *J. praehirsuta* ayant des soies sur les carpus P1-P4 ainsi qu'une ou deux épines sur les P6/P7, D) un mâle *J. posthirsuta* ayant des soies sur les P6-P7, E) un mâle *J. ischiosetosa* ayant des soies sur les ischiums des péréiopodes P6-P7 et F) un mâle *J. albifrons* ayant de très courtes soies en « brosse » sur les carpus P6-P7. Ce nombre de soies est très variable en fonction des individus et notamment du site de prélèvement de l'individu. Ainsi les mâles *J. praehirsuta* peuvent avoir de <10 à >35 soies sur leurs péréiopodes (constatation personnelle).

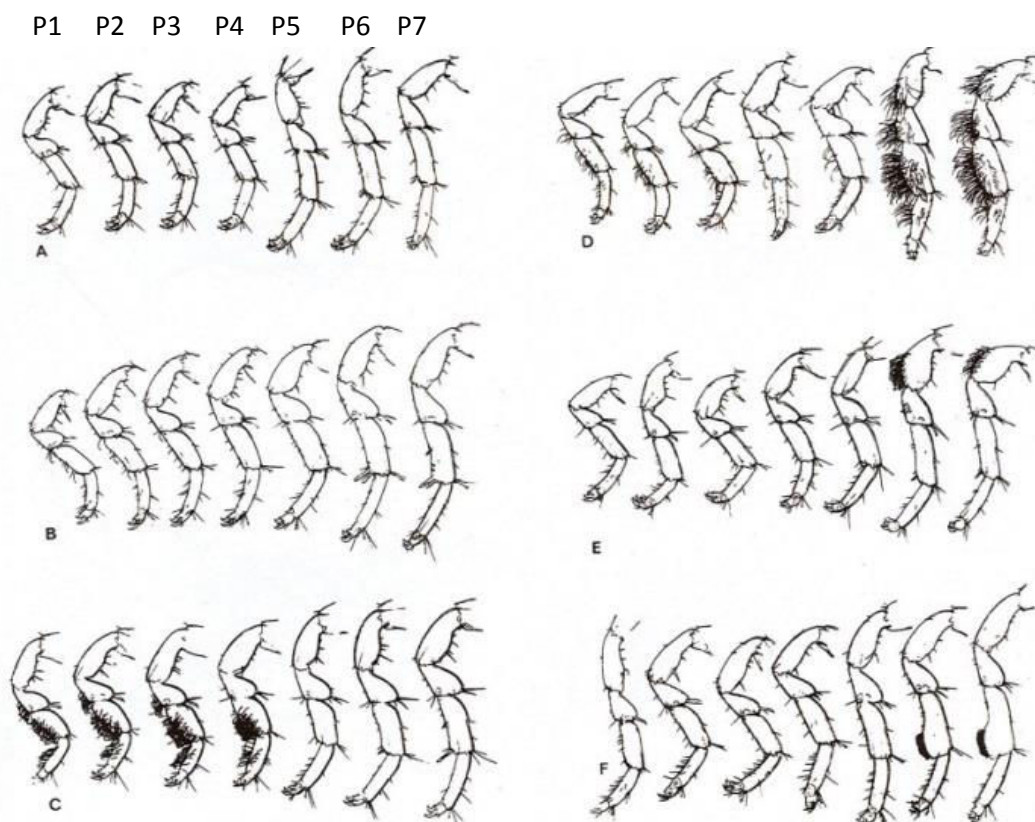


Figure I.13 : Dessins des différents péréiopodes chez les espèces du complexe *Jaera albifrons*. (A) femelle du complexe *Jaera albifrons*, (b) mâle *J. forsmanni*, (C) mâle *J. praehirsuta*, (D) mâle *J. posthirsuta*, (E) mâle *J. ischiosetosa*, (F) mâle *J. albifrons* (Solignac, 1981).

Les femelles des différentes espèces sont morphologiquement identiques et ne peuvent donc être discriminées par l'observation de leur phénotype. Cependant, une étude a montré que la taille du CZ (coupling zone) représentant la zone semblant être utilisée lors de la parade serait spécifique à chaque espèce et permettrait notamment de discriminer morphologiquement les femelles (Khaitov et al., 2006). Chez le mâle, cette zone est définie de la tête au premier segment abdominal portant l'organe copulateur, et chez la femelle il s'agit de la zone morphologique allant du pléotelson jusqu'à la limite entre le 4ème et le 5ème segment thoracique. C'est sur cette zone que les mâles vont venir faire leur parade. Actuellement, ce processus de choix de partenaire est considéré comme la barrière la plus forte.

3.4 Différenciation génétique

- nucléaire

Il semble que le complexe *Jaera albifrons* ait un système génétique de déterminisme du sexe. En effet d'après les caryotypes réalisés toutes les femelles possèdent un trivalent composé de 2 bras métacentrique et d'un élément acrocentrique qui n'est pas retrouvé chez les mâles (Staiger & Bocquet, 1954). Ce trivalent correspondrait donc à des chromosomes sexuels de type ZW_1W_2 (femelles hétérogamétiques).

Les premiers résultats cytogénétiques ont montré que les espèces ne peuvent être discriminées sur la base de leur caryotype, de part une forte variabilité cytogénétique interspécifique, mais également intraspécifique (Lécher & Prunus, 1971). Les espèces ne peuvent donc être discriminées que sur la base des variants sexuels des péréiopodes mâles (voir plus haut). Lécher (1968) a mis en évidence un polymorphisme chromosomique intraspécifique important. Un gradient chromosomique aurait ainsi lieu du Nord vers le Sud résultant de fusions centriques de plus en plus nombreuses sans que cela n'altère la reproduction (Bocquet, 1972), et qui pourrait être dus à des réarrangements Robertsonniens (Staiger & Bocquet, 1954). Une étude plus récente (Mifsud, 2011) montre que les espèces *J. ischiosetosa*, *J. albifrons* et *J. praehirsuta* sont génétiquement structurées et forment des clades monophylétiques (522 marqueurs AFLP polymorphes utilisés sur 32 individus, Figure 1.14), et qu'il y a des réarrangement chromosomiques de type fusions, semblant associées à une diminution de la fertilité hybride entre *J. ischiosetosa* et *J. albifrons*. Cette même étude a montré que la taille estimée du génome serait d'environ 1 Gb. Enfin, on peut également noter l'étude de (Siegismund, 2002) qui a montré une disparité et une différenciation liée au sexe pouvant limiter le flux de gènes entre ces espèces. Trois sur quatre locus décrits chez 22 populations de trois espèces de *Jaera* sont similaires, et le quatrième, le locus de l'enzyme Gpi, diffère selon les espèces et est très

différencié : il est autosomal ou lié au sexe selon les espèces et peut avoir un taux de recombinaison élevé entre locus et centromère. Grace aux nouvelles techniques de génotypage, il serait maintenant intéressant de regarder la structure génétique intraspécifique et interspécifique et d'explorer le patron de divergence d'une partie voire de la totalité de ce complexe d'espèces, afin de confirmer ou infirmer les précédents résultats d'hybridation pour ces espèces, et comprendre les patterns de différenciation chromosomiques interspécifiques.

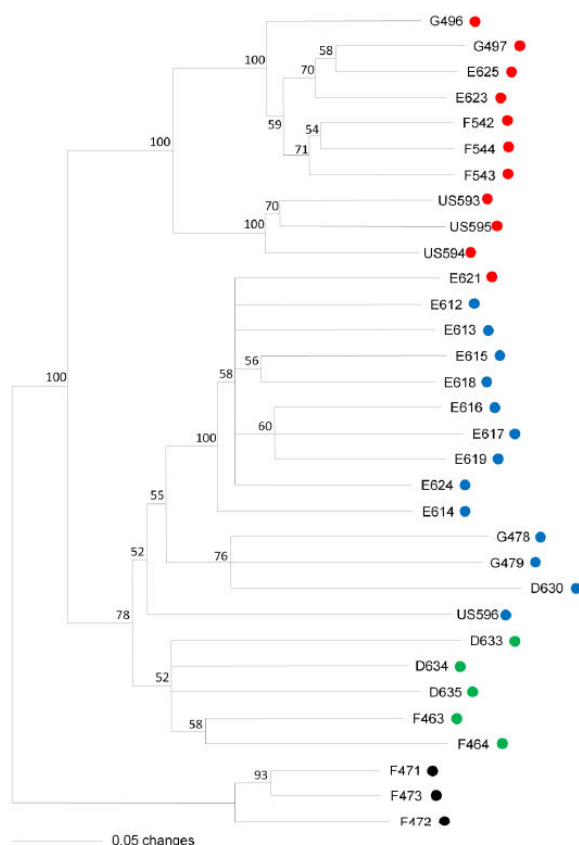
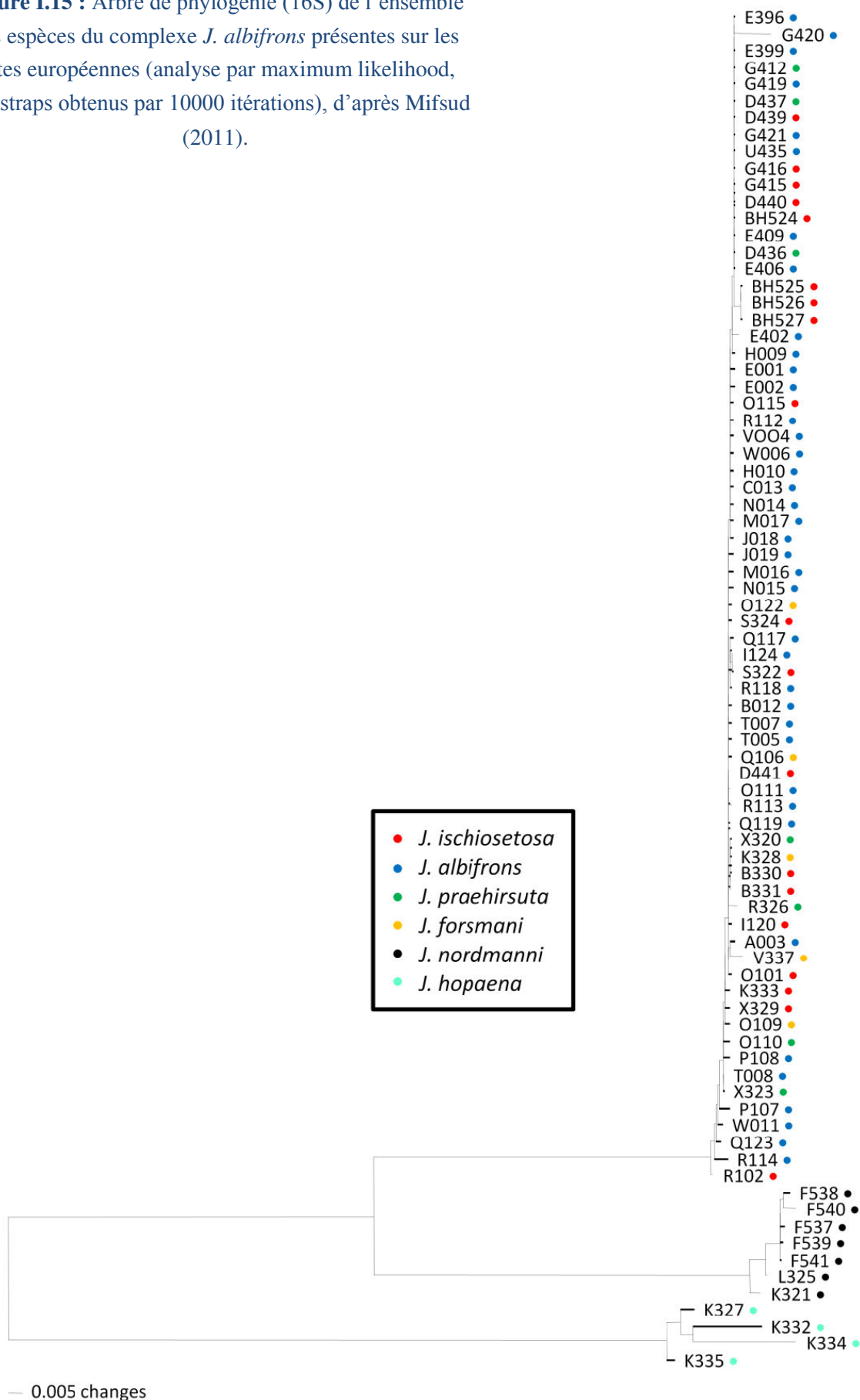


Figure I.14 : Arbre de type *Neighbour-joining* obtenu à partir de profils AFLP. En rouge sont représentés les *J. ischiosetosa*, en bleu les *J. albifrons*, en vert les *J. prae-hirsuta*, et en noir l'outgroup *J. nordmanni* (bootstraps obtenus par 10000 itérations). Le point rouge dans le clade *J. albifrons* serait une erreur d'identification. D'après Mifsud (2011).

-mitochondriale

Des analyses réalisées sur du CO1 (analyses effectuées durant ma thèse mais non présentées) et du 16S (Mifsud, 2011) montrent une très faible différenciation mitochondriale entre espèces, sous forme de clades polyphylétiques (Figure I.15). Mifsud (2011) a également détecté un haplotype qui serait largement partagé entre espèces. La discordance entre génome nucléaire fortement différencié et génome mitochondrial faiblement différencié ne fera pas l'objet d'une étude dans ce manuscrit mais en revanche sera utilisé comme élément de réflexion durant cette thèse.

Figure I.15 : Arbre de phylogénie (16S) de l'ensemble des espèces du complexe *J. albifrons* présentes sur les côtes européennes (analyse par maximum likelihood, bootstraps obtenus par 10000 itérations), d'après Mifsud (2011).



- *isolement post-zygotique*

Solignac (1976) a détecté que des hybrides de première génération entre *J. ischiosetosa* et *J. albifrons* sont viables, et ont une survie similaire à celle des morphotypes parentaux. En revanche une diminution de la fertilité semble atteindre ces hybrides et serait due à des réarrangements chromosomiques (Mifsud, 2011).

En revanche aucune autre étude de l'isolement post-zygotique entre *J. albifrons* et *J. praeheirsuta* n'a à ma connaissance été réalisée.

3.5 Isolement reproductif et hybridation

Trois barrières à la reproduction ont donc été décrites, limitant l'hybridation en milieu naturel (Solignac, 1981). Ces barrières sont de type écologique (différence d'habitat), sexuel (choix de partenaires) et génétique (incompatibilités génétiques menant à une dépression hybride). Cependant, en milieu contrôlé, des croisements entre toutes ces espèces peuvent être aisément réalisées à la seule condition de supprimer toute possibilité de choix aux partenaires d'espèces différentes (Bocquet, 1972), en dehors du croisement *J. albifrons* - *J. praeheirsuta* qui est plus difficile à obtenir (Bocquet & Solignac, 1969). Il semblerait que ces espèces vivant en sympatrie aient un taux d'hybridation assez faible en milieu naturel (en moyenne 0.6%) mais jusqu'à 15 à 30% de morphotypes intermédiaires entre *J. albifrons* et *J. praeheirsuta* auraient été trouvés dans une population unique de la côte Nord Atlantique de la France dans les années 60 (Solignac, 1969a).

En effet, Solignac (1969b) avait trouvé des individus morphologiquement intermédiaires sur une plage de Normandie, semblant être des hybrides entre *Jaera albifrons* et *J. praeheirsuta*. Il avait donc fait des croisements en laboratoire, afin de pouvoir comparer les morphotypes prélevés dans la nature de ceux obtenus par croisements expérimentaux, ce qui avait permis de confirmer sur une base strictement phénotypique qu'il semblait y avoir des hybrides en Normandie. Ils portaient des soies sur les P1-P4 à la manière des *J. praeheirsuta* (Figure I.16A), voir même parfois 1 ou 2 épine sur les P6/P7, ainsi que des soies en brosse, typique des *J. albifrons* (Figure I.16B).

Si cette situation d'hybridation en Normandie était confirmée, elle offrirait donc potentiellement une opportunité très intéressante de comparer des espèces en situation de sympatrie mais dont l'isolement reproductif, défini au moins par trois types de barrières à la reproduction, n'est pas toujours de même intensité.

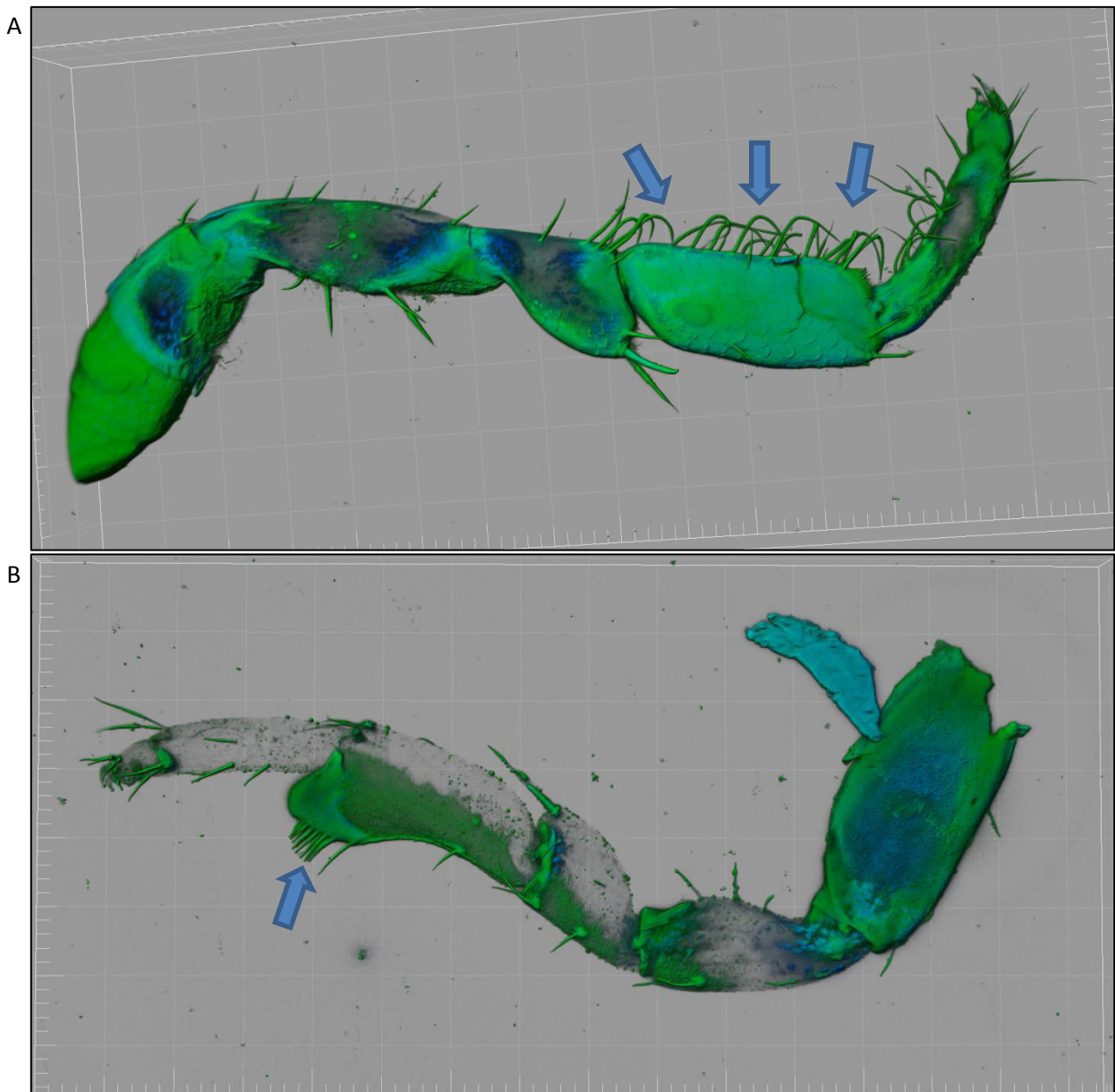


Figure I.16 : Photos d'un des deux péréiopodes P1 d'un individu *J. praehirsuta* (A) et d'un des deux péréiopodes P7 d'un individu *J. albifrons* (B). Les flèches montrent les caractères sexuels secondaires utilisés par les mâles pour la parade. Image obtenue par microscopie confocale à balayage laser. Une solution de 500 μ l contenant 10 μ M de Hoechst33342 (Sonde moléculaire, H21492) et 1,5 μ M de DiOC6 (3) (Sonde moléculaire D-273) a permis d'obtenir la fluorescence. DiO (vert) cible les composants hydrophobes du tégument tandis que Hoechst (bleu) cible l'ADN et peut marquer quelques polysaccharides (par exemple, la chitine). Le traitement d'image et la construction en 3D ont été réalisés avec les logiciels Fiji and IMARIS® (Bitplane). Crédit Sébastien Colin.

4. Objectifs de la thèse et structure du document

Un isolement reproductif complet peut être dû à un effet boule de neige impliquant plusieurs types de barrières à la reproduction qui entraînent une diminution des flux de gènes neutres ou sélectionnés, ce qui permet une accumulation des obstacles et augmente la différenciation entre espèces (Navarro & Barton, 2003). Cependant, au cours du temps, les barrières peuvent se renforcer ou bien, au contraire, être rompues partiellement ou totalement, comme à la suite d'une remise en contact de populations divergentes liée à des changements environnementaux, permettant la mise en place nouvelle d'un flux de gènes. Une manière d'apporter quelques réponses concernant l'évolution des barrières à la reproduction est donc d'analyser l'histoire démographique des populations étudiées. De plus, selon Wagner and Mandeville (2017) il est nécessaire de considérer à la fois l'ampleur de la divergence évolutive entre les taxons, et le niveau de flux génétique contemporain entre ceux-ci afin d'avoir la meilleure approche pour l'étude des barrières à la reproduction.

Comprendre l'histoire démographique, actuelle et ancienne, observer la répartition des espèces dans l'espace ainsi qu'analyser la distribution des loci participant à l'isolement reproductif le long du génome sont des objectifs nécessaires afin de comprendre les processus sous-jacents à la divergence entre taxons (Figure I.17).

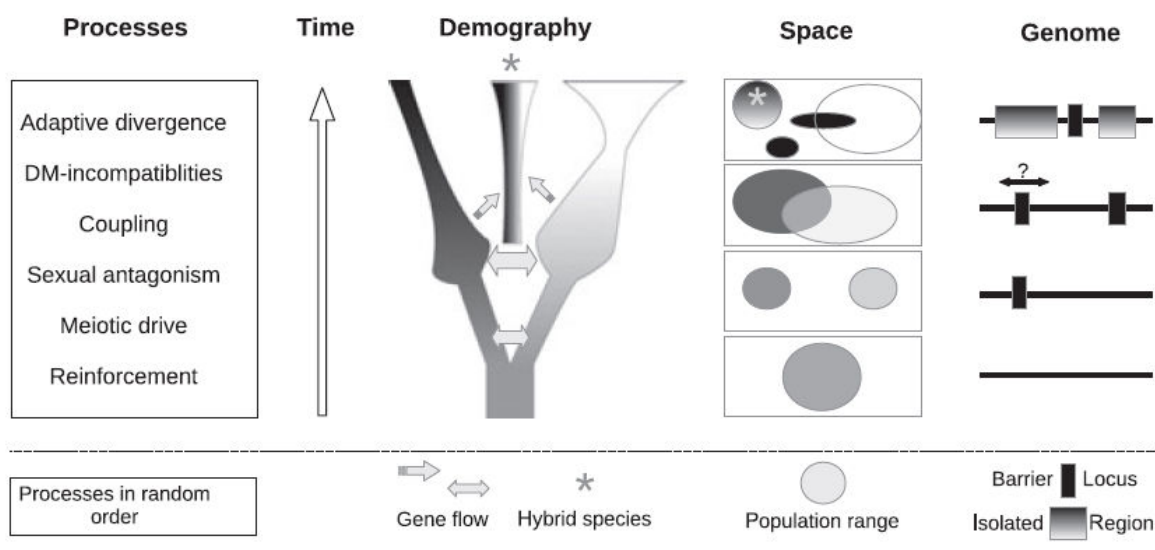


Figure I.17 : Processus pouvant être impliqués dans la spéciation, à travers le temps et l'espace, d'après Abbott et al. (2013) .

La taxonomie des espèces du complexe *Jaera albifrons* a longtemps été source de rectifications, mais on peut aujourd'hui distinguer 5 espèces qui semblent efficacement isolées dans la nature mais que l'on peut facilement croiser en laboratoire à l'exception de l'espèce *J. praehirsuta*. Or, dans les années 60 une hybridation potentiellement introgressive semble avoir été montrée sur

un site de Normandie entre *J. albifrons* et *J. praehirsuta* (Solignac, 1969b). Aujourd'hui, la compréhension d'une telle zone peut être améliorée avec l'émergence d'outils moléculaires puissants.

Durant cette thèse, mon objectif général est de détecter la force de l'isolement reproductif entre ces *Jaera albifrons* et *Jaera praehirsuta* face à deux situations contrastées : populations "classiques" où l'hybridation semble efficacement empêchée par l'action d'au moins trois types de barrières à la reproduction, et une population Normande où une hybridation forte entre ces deux espèces a été révélée entre 1965 et 1970 sur la base de descriptions morphologiques. Pour cela, j'ai cherché à répondre à quatre questions : où se situent ces espèces le long d'un continuum de spéciation ? Quelles sont les barrières à la reproduction impliquée ? Quelles sont les histoires démographiques associées à ces isolements ? Quelle est l'architecture génomique sous-tendant cet isolement ?

Pour répondre à ces questions, 6 grands thèmes ont été définis ou redéfinis au cours de ma thèse en fonction des résultats et forment la structure de ce manuscrit. Le chapitre 1 regroupe les deux articles publiés à ce jour, et un résumé (en français) associé à chacun de ces articles. Les autres chapitres sont écrits en français.

- **Chapitre 1 : Test de l'hypothèse d'hybridation entre *J. albifrons* et *J. praehirsuta*, différenciation génétique et comparaison de zones avec et sans hybridation entre ces deux taxons**

Contexte : Une zone d'hybridation exceptionnelle entre *J. albifrons* et *J. praehirsuta* a été décrite par M. Solignac en Normandie dans les années 60.

Questions : Cette zone d'hybridation est-elle toujours existante ? Si oui, détecte-t-on des niveaux variables de différenciation génétique interspécifique entre cette zone et d'autres où l'hybridation n'est pas suspectée ? Quelles sont les habitats de ces deux espèces entre ces différentes zones ?

Méthodes : Description des morphotypes et des niches exploitées et analyses génétiques à l'aide de marqueurs microsatellites dans une zone sans hybridation suspectée (située en Bretagne) et une zone dans laquelle l'hybridation avait été décrite (en Normandie).

- **Chapitre 2 : Structure génétique intraspécifique et interspécifique au sein du complexe *Jaera albifrons***

Contexte : Les deux espèces étudiées (*J. albifrons* et *J. prae-hirsuta*) font partie d'un complexe de 5 espèces dont 2 autres sont trouvées également sur les côtes nord-européennes.

Questions : Quelles sont les relations de ces deux espèces entre elles et avec les autres espèces du complexe ? Est-ce que la zone d'hybridation est plus étendue que ce que M. solignac avait suggéré (population exceptionnelle) ?

Méthodes : Elargissement des zones géographiques échantillonnées et analyses génétiques à l'aide de SNPs génotypés par RAD-seq.

- **Chapitre 3 : Isolement reproductif**

Contexte : Deux situations contrastées : populations de Bretagne où aucun cas d'hybridation n'a été rapporté, et populations Normandes où une hybridation introgressive est à l'oeuvre.

Questions : Quelles sont les barrières à la reproduction dans chaque région et leur force respective ?

Méthodes : Croisements expérimentaux intraspécifiques, interspécifiques et backcross en Bretagne et Normandie.

- **Chapitre 4 : La bactérie *Wolbachia* est-elle impliquée dans l'isolement reproductif ?**

Contexte : Un biais de sex-ratio a été détecté dans la littérature chez certaines espèces du complexe *Jaera albifrons*, et l'isolement reproductif démontré en chapitre 3 n'est pas incompatible avec l'action de *Wolbachia*.

Questions : Est-ce que *Wolbachia* est présente chez nos espèces ? Peut-elle y avoir un effet féminisant ? Cette bactérie participe-t-elle à l'isolement reproductif par la mise en place d'incompatibilités cytoplasmiques ?

Méthodes : Analyse du sex-ratio de plusieurs populations, et détection de *Wolbachia* à l'aide d'outils moléculaires.

- **Chapitre 5 : Histoire démographique sous-jacente à la divergence entre *J. albifrons* et *J. praeheirsuta***

Contexte : Deux zones avec un isolement reproductif et des barrières à la reproduction variables.

Questions : S'agit-il d'une divergence unique ou de spéciations parallèles ? Quelle pourrait être l'histoire démographique dans chacune des régions ?

Méthodes : Utilisation de spectres joints de fréquences alléliques obtenue par RAD-seq pour chacune des espèces dans les deux zones étudiées (avec et sans hybridation).

- **Chapitre 6 : Architecture génomique de l'isolement reproductif**

Contexte : Des réarrangements chromosomiques intra- et inter-spécifiques ont déjà été démontrés dans la littérature chez ce complexe d'espèce.

Questions : Quelle est la répartition des régions génomiques fortement différenciées entre espèces le long du génome ? Peut-on détecter des réarrangements chromosomiques ayant pu participer à l'isolement ?

Méthodes : Construction de deux cartes de liaison pour *J. albifrons* et *J. praeheirsuta* à partir de familles élevées au laboratoire, mesure de la différenciation génétique inter-spécifique en populations naturelles avec et sans hybridation, et analyse d'un scan génomique de cette différenciation dans ces deux situations.

Chapitre 1

**TEST DE L'HYPOTHESE
D'HYBRIDATION ENTRE *J.*
ALBIFRONS ET *J. PRAEHIRSUTA*,
DIFFERENCIATION GENETIQUE ET
COMPARAISON DE ZONES AVEC ET
SANS HYBRIDATION ENTRE CES
DEUX TAXONS**

Ce chapitre a pour objectif de tester l'hypothèse d'hybridation posée par M. Solignac dans les années 60-80 à partir de descriptions morphologiques (Bocquet & Solignac, 1969, Solignac, 1969b, Solignac, 1978b), en analysant la structure génétique interspécifique entre *J. albifrons* et *J. praehirsuta*, afin de vérifier si la différenciation génétique entre nos deux espèces est plus faible en Normandie (où l'on suspecte de l'hybridation), qu'en Bretagne (fort isolement reproductif). Ces analyses sont réalisées avec des marqueurs microsatellites que j'ai contribué à développer pendant mon stage de master et mon doctorat.

1. Développement de marqueurs microsatellites pour le complexe d'espèces *Jaera albifrons*

Article associé :

Ribardière, A., Broquet, T. & Daguin-Thiebaut, C. 2015. Microsatellite markers for the *Jaera albifrons* species complex (marine isopods). BMC Res. Notes 8: 632.

Le complexe d'espèces *Jaera albifrons* a été étudié dès les années 1950 en premier lieu par des études principalement systématiques et morphologiques (Bocquet, 1953, Harvey & Naylor, 1968, Prunus, 1968, Bocquet & Solignac, 1969, Khaitov et al., 2006), puis écologiques (Carvalho, 1989a, Haahtela, 1965, Naylor & Haahtela, 1966, Solignac, 1969a), comportementales (Veuille, 1980, Solignac, 1981) ou encore de type biochimique, portant sur des analyses de différenciation génétique à courte échelle géographique à l'aide d'allozymes (Carvalho, 1989b, Carvalho & Piertney, 1997).

Lorsque nous avons décidé de travailler à un niveau génétique sur le complexe d'espèces *Jaera albifrons*, nous avons dû au préalable développer des outils moléculaires plus performants, de type Simple Sequence Repeats, hautement polymorphes. 95 marqueurs microsatellites ont donc été développés à partir de 7 individus *J. albifrons* (étape en partie déléguée à la Société Genoscreen, Lille, France), avec détection des motifs microsatellites grâce au programme QDD v1 (Megléczy et al., 2010). Sur ces 95 marqueurs, 24 étaient amplifiables et polymorphes, avec des motifs répétés de 2 à 4 nucléotides. Ces 24 marqueurs ont été testés sur 49 mâles adultes *J. albifrons* échantillonnés dans 2 populations de Bretagne.

Le nombre moyen d'allèles par locus était de 4.73 ± 2.45 et l'hétérozygotie attendue moyenne était de 0.55 ± 0.23 . Tous les loci étaient polymorphes (2 à 13 allèles/locus) chez *J. albifrons* à l'exception d'un locus, cependant gardé car polymorphe chez *J. praehirsuta* et *J. forsmanni*. Sur les 24 marqueurs, 6 microsatellites déviaient de l'équilibre de Hardy-Weinberg, écart pouvant être dû à la présence d'allèles nuls, théorie supportée par les analyses réalisées avec le logiciel

Micro-checker (Van Oosterhout et al., 2004). On peut cependant noter qu'aucun déséquilibre de liaison entre paires de loci n'a été détecté.

La transférabilité de ces 24 marqueurs aux 3 autres espèces du complexe *J. albifrons* présentes sur la côte atlantique européenne (*J. praehirsuta*, *J. ischiosetosa* et *J. forsmanni*) a été testée. L'amplification a été considérée comme réussie lorsque 85% des individus testés montraient un génotype de bonne qualité. Chez 74 mâles *J. praehirsuta* provenant de 5 populations européennes, 23 des 24 marqueurs amplifiaient. Pour 18 mâles *J. ischiosetosa* de 2 populations nord-américaines, et 8 mâles *J. forsmanni* d'une population européenne 20 marqueurs amplifiaient. Dans la majorité des cas, ces loci étaient polymorphes.

Ce panel de marqueurs microsatellites est intéressant car nous avons obtenu un certain nombre de marqueurs polymorphes et utilisables pour au moins 4 espèces sur les 5 formant le complexe *Jaera albifrons*. Le développement de ces marqueurs microsatellites devrait ainsi permettre l'étude de plusieurs points intéressants tels que i) la confirmation de l'hypothèse d'hybridation entre les espèces *J. albifrons* et *J. praehirsuta* (chapitre 1) ii) la structure génétique intraspécifique et interspécifique entre les deux espèces précédemment citées (chapitre 2) iii) la comparaison de ces différenciations génétiques entre différents sites le long d'un gradient géographique allant de la Bretagne Nord à la Normandie, ou encore iv) des analyses de parenté nécessaires à des estimations de fitness.



Guillaume Evanno

Microsatellite markers for the *Jaera albifrons* species complex (marine isopods)

Ribardière *et al.*

TECHNICAL NOTE

Open Access



Microsatellite markers for the *Jaera albifrons* species complex (marine isopods)

Ambre Ribardière*, Thomas Broquet and Claire Daguin-Thiébaud

Abstract

Background: The *Jaera albifrons* complex contains five species of marine isopods (*J. albifrons*, *J. prae-hirsuta*, *J. ischiosetosa*, *J. forsmanni*, and *J. posthirsuta*). These species, occurring on the shores of the North-Atlantic Ocean, are partially reproductively isolated by barriers due to sexual isolation (mate choice), genetic incompatibilities, and ecological specialization. Microsatellite loci would be useful for parentage-based analyses of sexual selection and studies of genetic structure in the context of speciation.

Findings: Twenty-four microsatellite markers were developed for *J. albifrons* using pyrosequencing of enriched libraries. Patterns of polymorphisms were analyzed in 49 *J. albifrons* adult males sampled in two populations from Brittany (Western France). The average number of alleles per locus was 4.73 ± 2.45 and the average gene diversity was 0.55 ± 0.23 . Most markers also successfully amplified in the three sibling species *J. prae-hirsuta*, *J. ischiosetosa*, and *J. forsmanni*.

Conclusions: These polymorphic and cross-amplifiable markers will be useful for population genetics and parentage studies in the *J. albifrons* complex.

Keywords: SSR, Species complex, Cross-amplification, 454 pyrosequencing, Multiplex PCR

Findings

Background

The *Jaera albifrons* complex (Leach, 1814) is composed of five intertidal isopod species [1, 2]. Three species (*J. albifrons*, *J. prae-hirsuta*, and *J. ischiosetosa*) have a large distribution along the coasts of the North-Atlantic Ocean from South-Spain and South-USA up to Baltic and Arctic regions, while the two other species are restricted either to the North-American East coast (*J. posthirsuta*) or the European coasts (*J. forsmanni*). Interest in these species stems from the fact that gene flow is interrupted by several isolating barriers but hybridization can occur under particular circumstances. Isolating barriers include ecological isolation, sexual isolation, and genetic incompatibilities [3, 4]. Microsatellite loci will be useful for population genetic studies of the *J. albifrons* complex, and fine scale analyses requiring parentage assignment (e.g. for investigating mechanisms of sexual

selection). Microsatellite markers were developed using two French *J. albifrons* populations and cross-amplification was tested for the three other species that are found in Europe (*J. prae-hirsuta*, *J. ischiosetosa*, and *J. forsmanni*).

Methods

Total genomic DNA was isolated from seven *J. albifrons* individuals (three males and four females) using the Nucleospin® Tissue kit (Macherey–Nagel) and sent to Genoscreen (Lille, France) for microsatellite development. Libraries enriched for microsatellites were prepared according to Malausa et al. [5] and sequenced on a 454 GS-FLX Titanium pyrosequencer (Roche). Among 42,661 raw sequences, 2609 microsatellite motifs were detected using QDD v1 [6] with default parameters, yielding 168 potential primer pairs. Among them, 95 primer pairs maximizing the number of repeats were tested for amplification and polymorphism using a set of seven *J. albifrons* individuals. Nine loci were found to be monomorphic, 46 loci did not yield amplification products, and 16 gave uninterpretable amplification patterns. The remaining 24 promising loci included dinucleotide, trinucleotide and tetranucleotide repeat motifs (Table 1).

*Correspondence: ambre.ribardiere@sb-roscoff.fr
CNRS, Station Biologique de Roscoff, Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, Team Diversity and Connectivity of Coastal Marine Landscapes, 29680 Roscoff, France

Table 1 Microsatellite loci for *Jaera albifrons* and multiplex PCR conditions

Locus	F primer seq. (5' to 3')	R primer seq. (5' to 3')	Motif	Range size (pb)	GenBank Accession no.	PCR multi-plex	Dye	Primer (10 µM)		Init. temp.	Fin. temp.	No. final cycles	
								F** (µl)	R (µl)				
Ja30	CGTCATTATGGGTGGGTA	CGTCCATTCTCACCTGACTGC	(TC) ₉	173–185	KP749881	1	PET	0.1	0.2	60	50	20	
Ja37	TCGAGCATCATAGCAGACA	GCCATATAGTCTTTGGCACC	(GA) ₈	101–115	KP749883		FAM	0.4	0				
Ja41	GGTGTGGCGGAAAGATTCT	AGCTACTTACCGTCTGTT	(GA) ₈	172–180	KP749885		NED	0.1	0.2	0.3			
Ja91	ATCGGCATTCCACTAGAGAG	CCAGTCTTGAATGCCTTGGG	(CCT) ₆	073–082	KP749896		VIC	0.1	0.2	0.3			
Ja01	CAATCTTCGCTTGGCGTCT	CACGTGGGCTTAGGATTGT	(AG) ₈	226–247	KP749875	2	FAM	0.5	0	0.5	57	47	20
Ja02	TGCCACCAATTAACCAACA	CGACAGCCATAATGATCACC	(CT) ₁₀ N ₁₆ (CT) ₅	182–212	KP749876		NED	0.1	0.2	0.3			
Ja35	TGGATTACATTCCTCTGG	TCACGCATTCTAGTCTACGCT	(AT) ₈	173–179	KP749882		VIC	0.3	0.1	0.4			
Ja47	GGTGTCTAATGTAGTAGCGG	ATGGAACCAAAAAGGACG	(TG) ₁₀	116–124	KP749886	3	NED	0.1	0.2	0.3	60	50	20
Ja56	ACTAACACAACATGCACCTACA	ATTGTAGGTGCTGCCAAT	(ACA) ₆	261–276	KP749888		FAM	0.3	0.3	0.6			
Ja58	CCCACTGGACACTACTTGA	ACCCTATCGTTGTAGTTGAGGT	(CT) ₅	186–188	KP749889		VIC	0.3	0.1	0.4			
Ja99	ATCAACATGCAGCGGTGC	AGGCATCTCAGAGGAATCAGTT	(TC) ₅	152–154	KP749898		PET	0.1	0.2	0.3			
Ja21	TTCAAAATTAATGCAACGATG	CATTATCTGAAGCTGTGGTTC	(GA) ₁₀	272–286	KP749878	4	NED	0.25	0	0.25	60	50	25
Ja71	CTGCCCTATCAGTTGAGCTT	CACAGTCCATCAAAAAGC	(GA) ₇	263–269	KP749892		VIC	0.3	0	0.3			
Ja78	CACCTACAGCAGCATAAGAGT	AGAACAAAGGCAAGATGAGCTC	(AC) ₈	243–254	KP749893		PET	0.2	0.1	0.3			
Ja94	TGTTCCGATCGTGGTTTCA	CCTTCGGATGGTAATAGGCA	(AT) ₆	101–111	KP749897		PET	0.1	0.2	0.3			
Ja22	GCAAACTCGATCACCATTGGA	ACCACCGTCTCTATCAAC	(CT) ₉	106–128	KP749879	5	VIC	0.2	0.1	0.3	60	50	23
Ja64	TAAGCCGAGTCTCAACAGCA	CCCGTGCATAGCGGAACAG	(TG) ₃ N ₂₂ (TG) ₅	119–127	KP749890		PET	0.1	0.2	0.3			
Ja66	CGCAAGTACAAATTTCTATGCT	TCGAACGGTACTCAATGTGAAG	(TAGA) ₅	175–187	KP749891		FAM	0.4	0	0.4			
Ja80	TGATGGATGAGAGTGATCGT	TTGATTCAGATCCGATACCATGT	(GA) ₆	154–154	KP749894		NED	0.2	0.1	0.3			
Ja13	ATCCTTCATGAGTCCCGAGT	AAGTATGTCGAACTACCCG	(GA) ₁₂	211–233	KP749877	6	VIC	0.2	0.1	0.3	57	47	23
Ja23	CAGGCCATTTGCTCGAGAT	ACAGCCCTCTCCATCATGCGT	(GA) ₉	120–134	KP749880		VIC	0.1	0.2	0.3			
Ja39	CTGGAGTCAATCTCAGTCAAG	GCTGAGGAGCGGAGTAATC	(TC) ₈	128–144	KP749884		NED	0.15	0.15	0.3			
Ja55	ACAACAGGAACAACATCCGT	AGTTGTGATGTGGAGGACCA	(CAA) ₆	130–138	KP749887		PET	0.15	0.15	0.3			
Ja82	TCCGAATGCACCAAACTTGA	GGCTTGTATTCGAATTTACATGGA	(GAA) ₆	292–295	KP749895		FAM	0.4	0	0.4			

The volume of fluorescent (F**) and non-fluorescent (F) forward primer was adjusted for each marker. Columns Init. temp. and Fin. temp. give the initial and final temperatures (in °C) used for the touchdown PCR. Amplification products from multiplex 4–6 were diluted before electrophoresis (dilution level: 1/3)

Polymorphisms of these 24 loci were estimated in two populations of *J. albifrons* from Brittany: Lingoz (48°39'12.31"N, 3°57'0.43"W, $n = 24$ males) and Inizan (48°39'34.09"N, 3°56'25.66"W, $n = 25$ males), for which genomic DNA was extracted from entire individuals using NucleoSpin® 96 Tissue kit (Macherey–Nagel). Locus amplification was performed in six multiplex PCRs (three to five loci per PCR, Table 1), in 15 μ l solutions containing 13 μ l of reaction mixture and 2 μ l of template DNA. Reaction mixtures contained 0.5 U of Gotaq G2 Hotstart DNA polymerase (Promega), 1 \times PCR buffer, 0.25 mM of each dNTP, 2 mM of MgCl₂, 0.1 mg/ml of bovine serum albumin, and primers in locus-specific concentrations (Table 1). We used a touchdown PCR method, performed by a T100 Thermal Cycler (Bio-RAD) with the following conditions: initial denaturation at 95 °C for 4 min, followed by ten cycles of 95 °C for 30 s, annealing for 30 s with temperature step-downs (1 °C at each cycle) starting at an initial temperature specific to each multiplex (Init. temp. in Table 1), and 72 °C for 30 s.

This was followed by 20–25 final cycles of 95 °C for 30 s, final temperature (Fin. temp.) for 30 s, 72 °C for 30 s, and a final elongation at 72 °C for 10 min.

PCR products were electrophoresed in a ABI 3130XL capillary sequencer (Applied Biosystems) together with the SM594 size marker [7] and electropherograms were analyzed using Genemapper v4 (Applied Biosystems). The number of alleles per locus, allelic richness, and observed and expected heterozygosity were estimated in Fstat vs. 2.9.3.2 [8]. This software was also used to test for Hardy–Weinberg equilibrium (Global test, option “HW within samples”, 10,000 permutations, Bonferroni correction applied), population differentiation, and linkage disequilibrium (option “between all pairs of loci in each sample”, 11,040 permutations). The presence of null alleles was tested using Micro-Checker vs. 2.2.3 [9].

Finally, the transferability of these markers was tested on three other species from the *J. albifrons* complex: *J. praeirsuta* ($n = 74$ males from five European

Table 2 Polymorphism of *Jaera albifrons* microsatellite loci

Locus	Lingoz ($n = 24$)				Inizan ($n = 25$)				Cross-amplification (no. alleles)		
	No. alleles	Ho	He	F _{IS}	No. alleles	Ho	He	F _{IS}	<i>J. praeirsuta</i> ($n = 74$)	<i>J. ischiosetosa</i> ($n = 18$)	<i>J. forsmani</i> ($n = 8$)
Ja30	4	0.63	0.64	0.031	6	0.72	0.71	-0.013	7	5	6
Ja37	4	0.39	0.67	0.418	6	0.32	0.56	0.432	9 [#]	5 [#]	4
Ja41	4	0.46	0.45	-0.016	5	0.56	0.50	-0.113	10	8	5
Ja91	4	0.50	0.53	0.053	3	0.60	0.56	-0.137	6	3	4
Ja01	9	0.43	0.85	0.496*	11	0.59	0.84	0.302	22	3 [#]	4 [#]
Ja02	7	0.71	0.73	0.035	6	0.76	0.68	-0.118	11	4	5
Ja35	5	0.54	0.58	0.074	4	0.64	0.66	0.035	7	2	4
Ja47	4	0.67	0.61	-0.087	4	0.40	0.55	0.278	5	1	4
Ja56	6	0.67	0.62	-0.071	4	0.28	0.35	0.200	8	4	3
Ja58	2	0.00	0.08	1.000	2	0.08	0.27	0.713	6	2	1
Ja99	2	0.38	0.31	-0.211	2	0.36	0.35	-0.029	3	2	2
Ja21	5	0.57	0.69	0.183	6	0.52	0.70	0.257	12	6	3
Ja71	4	0.68	0.76	0.101	5	0.72	0.71	-0.013	12	9	- [#]
Ja78	5	0.61	0.75	0.197	5	0.71	0.70	-0.009	8	6 [#]	4 [#]
Ja94	5	0.63	0.68	0.078	5	0.56	0.70	0.198	9	5	5
Ja22	6	0.79	0.73	-0.094	6	0.52	0.58	0.102	10	6	2
Ja64	2	0.04	0.31	0.869*	2	0.12	0.35	0.662	6	2	3
Ja66	3	0.25	0.29	0.140	2	0.16	0.15	-0.067	5	1 [#]	1
Ja80	1	NA	NA	NA	1	NA	NA	NA	4	1	2
Ja13	9	0.54	0.71	0.240	8	0.40	0.60	0.332	16	8	10
Ja23	4	0.50	0.69	0.275	7	0.76	0.71	-0.078	8	5	4
Ja39	8	0.92	0.85	-0.077	12	0.88	0.89	0.008	15	2	5 [#]
Ja55	3	0.17	0.59	0.720*	5	0.48	0.67	0.289	9	3	4
Ja82	2	0.29	0.25	-0.150	2	0.12	0.12	-0.043	2	2	2

The number of alleles, expected heterozygosity (He), observed heterozygosity (Ho) and F_{IS} are given for each locus in two populations of the *J. albifrons* species (* indicates a significant deviation from HWE, $p < 0.05$). The number of alleles obtained through cross-amplification is given for samples of *J. praeirsuta*, *J. ischiosetosa*, and *J. forsmani* ([#] indicates amplification success below 85 %, see text). Loci presented in the same order as in Table 1

populations), *J. ischiosetosa* ($n = 18$ males from two North-American populations), and *J. forsmani* ($n = 8$ males from one European population).

Results and discussion

The average number of alleles per locus for the two pooled *J. albifrons* populations ($n = 49$) was 4.73 ± 2.45 and the average gene diversity was 0.55 ± 0.23 (details per locus and population in Table 2). All loci were polymorphic (2–13 alleles per locus) except Ja80, which is nonetheless reported here because it was polymorphic in *J. prae-hirsuta* and *J. forsmani* (Table 2) and could thus be useful at least for these species. Microsatellites Ja01, Ja55, and Ja64 deviated significantly from HWE ($p < 0.001$ in one of the two populations). Micro-checker results suggested that null alleles might be segregating at these loci as well as three additional markers associated with large F_{IS} values (Ja13, Ja27, and Ja58). Null alleles are often unavoidable in highly polymorphic species such as many marine invertebrates [10] and relevant microsatellite loci should be used only in analyses where their effect can be detected and corrected (e.g. parentage assignment). Moreover, the occurrence of null alleles is expected to be variable across geographic regions and species, so that the results reported here for two populations might not apply to other areas or species (our two samples came from nearby, albeit differentiated populations, $F_{ST} = 0.01$, $p < 0.0001$). Markers used in empirical studies should be chosen accordingly, and the multiplex design proposed here could be adapted. There was no linkage disequilibrium for any pair of loci.

Cross-species amplification was considered successful if more than 85 % of the individuals tested produced a good quality genotype at the first attempt and without optimization. With this criterion, 23 out of 24 microsatellite markers successfully amplified in *J. prae-hirsuta* ($n = 74$), 20 in *J. ischiosetosa* ($n = 18$) and 20 in *J. forsmani* ($n = 8$). These loci appeared to be polymorphic in nearly all cases (Table 2). These markers seem readily transferable to other species for European populations of *J. prae-hirsuta* and *J. forsmani*, and even to North-American populations in the case of *J. ischiosetosa*. The panel of microsatellites reported here thus provides a useful set of markers for parentage analyses and studies of the interspecific genetic structure within the *J. albifrons* complex.

Availability of supporting data

The sequences containing microsatellite motifs are available through the National Centre for Biotechnology Information under accession numbers KP749875 to KP749898 (<http://www.ncbi.nlm.nih.gov/>).

Authors' contributions

AR performed field sampling, final multiplex optimization, genotyping and data analysis. TB contributed to field sampling, designed the primers, and supervised the analysis and description of polymorphism for the final set of loci. CDT performed initial tests for amplification and polymorphism, and supervised final multiplex optimization and genotyping. All authors contributed to writing the manuscript. The authors read and approved the final manuscript.

Acknowledgements

We thank J. Coudret and C. Houbin for their help during sampling and species determination, and J. Jaquiéry, F. Viard and G. Yannic for discussions and comments. This work benefited from access to the Biogenouest genomic platform at Station Biologique de Roscoff, and was supported by the French Agence Nationale de la Recherche (grant ANR-13-JSV7-0001-01 to T.B.).

Competing interests

The authors declare that they have no competing interests.

Received: 11 May 2015 Accepted: 19 October 2015

Published online: 02 November 2015

References

- Bocquet C. Evolution of a superspecies of marine isopods. *Syst Zool.* 1954;3(4):148–62.
- Prunus G. Etude de systématique des populations chez l'isopode *Jaera (albifrons) albifrons* Forsman. *Arch Zool Exp Gen.* 1968;109:643–702.
- Mifsud DV. The genetic basis of speciation in the *Jaera albifrons* species group of intertidal isopods: PhD dissertation, University of Aberdeen; 2011.
- Solignac M. Isolating mechanisms and modalities of speciation in the *Jaera albifrons* species complex (*Crustacea, Isopoda*). *Syst Zool.* 1981;30(4):387–405.
- Malaua T, Gilles A, Meglecz E, Blanquart H, Duthoy S, Costedoat C, et al. High-throughput microsatellite isolation through 454 GS-FLX titanium pyrosequencing of enriched DNA libraries. *Mol Ecol Resour.* 2011;11(4):638–44.
- Meglecz E, Costedoat C, Dubut V, Gilles A, Malaua T, Pech N, et al. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics.* 2010;26(3):403–4.
- Mauger S, Couceiro L, Valero M. A simple and cost-effective method to synthesize an internal size standard amenable to use with a 5-dye system. *Prim Res Biotech.* 2012;2:40–6.
- Goudet J. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3), updated from Goudet 1995. Available from <http://www.unil.ch/izea/software/fstat.html>. 2001.
- van Oosterhout C, Hutchinson WF, Wills DP, Shipley P. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes.* 2004;4:535–8.
- Hare MP, Karl SA, Avise JC. Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Mol Biol Evol.* 1996;13(2):334–45.

2. Analyse de la différenciation génétique entre deux espèces du complexe *Jaera albifrons*, en zone d'isolement reproductrice et en zone d'hybridation

Article associé :

Ribardière, A., Daguin-Thiébaud, C., Houbin, C., Coudret, J., Broudin, C., Timsit, O., & Broquet, T. (2017). Geographically distinct patterns of reproductive isolation and hybridization in two sympatric species of the *Jaera albifrons* complex (marine isopods). *Ecology and Evolution*.

Cette étude a démarré grâce au postulat de Solignac (1969b) selon lequel en Normandie, sur le site de Luc-sur-Mer (49°18'55.5"N 0°20'29.9"W), il y aurait de l'hybridation entre *J. albifrons* et *J. prae-hirsuta*. Ces espèces ne partageant généralement pas les mêmes habitats, il semblait s'agir d'un cas isolé. En effet, jusque-là, l'habitat de prédilection de *J. albifrons* semblait sous les pierres sur la partie haute de l'estran et celui de *J. prae-hirsuta* sur le fucus plus bas sur l'estran (Bocquet, 1954).

En 2013, malgré une recherche intensive sur une large portion de côte (décrite dans le papier publié) seules 5 femelles ont été trouvées au niveau du site de Luc-sur-Mer. En revanche, un échantillonnage préliminaire sur un site proche de Luc-sur-Mer, mentionné par Solignac, a permis de prélever 1 mâle à morphotype intermédiaire, *i.e.* qui porte à la fois des caractères sexuels secondaires typiques des *J. albifrons* et des *J. prae-hirsuta* sur le site de Sainte-Honorine-des-Pertes (49°21'15.0"N 0°47'49.6"W, cf. Figure 1 dans l'article associé). Sur un autre site normand, Grandcamp (49°23'24.3"N 1°02'09.1"W), une des femelles gravides échantillonnées et élevées en laboratoire a donné naissance à 6 mâles morphologiquement intermédiaires.

C'est à la suite de cette découverte que j'ai commencé à travailler sur ce sujet (stage de master 2, printemps 2014). Les objectifs étaient de i) tester l'hypothèse d'hybridation entre *J. albifrons* et *J. prae-hirsuta* ii) examiner la structure géographique des populations montrant une hybridation et s'interroger sur la nature des mécanismes d'isolement dans ces populations, et iii) comparer les patterns génétiques entre populations s'hybridant et populations ne semblant pas s'hybrider.

Pour cela, un premier échantillonnage a été réalisé début d'année 2014 en Normandie, sur les sites de Grandcamp, Sainte-Honorine-des-perthes, et Longues avec l'aide de Thomas Broquet, Céline Houbin et Olivier Timsit. A Grandcamp, 1 individu à phénotype intermédiaire a été trouvé tandis qu'à Longues aucun morphotype intermédiaire n'a été détecté mais la présence de *J. prae-hirsuta* était faible dans cette zone (1 *J. prae-hirsuta* pour 15 *J. albifrons*). Dans ces deux cas,

tous les individus détectés se situaient sous les pierres. Un large échantillonnage a ensuite été réalisé sur le site de Sainte-Honorine-des-pertes où l'individu à morphotype intermédiaire avait été découvert l'année précédente, et nous avons effectué la recherche le long de la plage sur environ 250 mètres, en soulevant chaque caillou et en nettoyant 10 échantillons de fucus (d'environ 50 cm³) afin de connaître la distribution géographique des individus. 199 individus ont été échantillonnés, dont sept d'entre eux montraient un phénotype intermédiaire, mais aucun individu n'a été trouvé sur les algues. Les morphotypes intermédiaires se situaient dans une zone moins dense en termes d'individus, mais aucune structuration géographique entre les différents morphes n'a été détectée. Il ne semble donc pas y avoir à fine échelle de zone hybride classique, *i.e.* présence des morphotypes intermédiaires sur le milieu de plage, avec d'un côté de la plage tous les individus *J. albifrons* et de l'autre côté, tous les individus *J. prae-hirsuta*.

Lors de ce premier échantillonnage en Normandie, nous avons donc pu remarquer qu'aucun individu ne semble être sur des algues, tous les individus étant trouvés en sympatrie sous les pierres, avec un nombre relativement faible de morphotypes intermédiaires sur l'ensemble des 3 sites (3.5%).

En Bretagne en revanche, nous avons échantillonné 18 sites, et la majorité des *J. albifrons* ont été trouvés sous les pierres et la majorité des *J. prae-hirsuta* sur des algues. Sur les 6 sites où les deux espèces ont été trouvées en sympatrie (cf. Figure 1 dans l'article associé), les micro-habitats étaient les mêmes que pour les autres sites de Bretagne, avec peu de chevauchement (2.25% sur l'ensemble des sites) et aucun morphotype intermédiaire n'a été détecté. Les zones de sympatrie en Bretagne sont donc définies comme des assemblages de deux micro-habitats. Ces assemblages sont suffisamment étroits (algues et substrats rocheux séparés de quelques centimètres à quelques mètres) pour que les individus occupant chacune des deux niches écologiques ne soient pas totalement physiquement isolés (le chevauchement moyen de 2.5% témoignant de l'association imparfaite des espèces à chacun des habitats au sein d'un site d'échantillonnage).

Afin de répondre aux questions précédemment posées, nous avons analysé la différenciation interspécifique au sein de chaque région à l'aide de 23 marqueurs microsatellites. En Bretagne ces marqueurs montrent une plus forte différenciation génétique interspécifique (analyse de variance génétique hiérarchique, $F_{CT} = 0.132$) qu'en Normandie, où 21 des 23 marqueurs montrent une plus faible différenciation entre espèces ($F_{CT} = 0.017$) alors que les 2 autres marqueurs restent fortement différenciés ($F_{CT} = 0.428$). Ces résultats suggèrent donc i) une hybridation entre *J. albifrons* et *J. prae-hirsuta* au sein des sites Normands ii) un flux de gènes hétérogène (barrières semi-perméables) en Normandie et un isolement reproductif fort en


Bretagne iii) une introgression mutuelle mais *a priori* asymétrique, les hybrides et individus introgressés semblant génétiquement plus proches des *J. praehirsuta* que des *J. albifrons* (voir Figure 5 dans l'article associé).

Cette étude apporte plusieurs autres éléments intéressants : i) nous avons comparé deux zones dont une avec hybridation et une à fort isolement reproductif, ce qui a révélé des patterns génétiques semblant être différents chez ces deux espèces selon les régions ii) nous avons détecté des loci outliers susceptibles d'être dans des zones sous sélection directe ou indirecte iii) nous avons détecté des zones d'hybridation éloignées, sans populations pures apparentes dans un milieu proche, ce qui pose la question des processus permettant le maintien des morphotypes parentaux. Ce travail a permis de confirmer l'intérêt de ce complexe d'espèces pour l'étude plus globale de la spéciation, notamment par l'analyse comparative de la différenciation génétique interspécifique entre une zone hybride et une zone d'isolement reproductif. Ce travail devra donc continuer avec une analyse à l'échelle du génome (loci obtenus par séquençage de sites de restriction liés à l'ADN (RAD-sequencing, Baird et al., 2008) afin de comprendre l'histoire démographique de ces deux espèces dans chaque région, ou encore de pouvoir placer les loci à fort F_{ST} le long du génome afin de détecter l'architecture génomique de la différenciation.

Les sites en Normandie étant éloignés de plusieurs kilomètres, il semblait intéressant de rechercher jusqu'où cette zone de sympatrie avec hybridation pouvait s'étendre. Suite à cette étude nous sommes donc retournés faire un échantillonnage supplémentaire qui a permis de montrer que la zone abritant des populations avec hybridation interspécifique est beaucoup plus étendue que ce qui était supposé (Solignac, 1969b, Solignac, 1969a, Solignac, 1978b) mais n'a pas permis de détecter une quelconque limite de zone hybride. Une description plus détaillée de cet échantillonnage ainsi que les analyses génétiques résultantes sont l'objet du chapitre 2.

ORIGINAL RESEARCH

Geographically distinct patterns of reproductive isolation and hybridization in two sympatric species of the *Jaera albifrons* complex (marine isopods)

Ambre Ribardière^{1,2}  | Claire Daguin-Thiébaud^{1,2} | Céline Houbin^{2,3} | Jérôme Coudret^{1,2} | Caroline Broudin^{2,3} | Olivier Timsit⁴ | Thomas Broquet^{1,2}

¹Centre National de la Recherche Scientifique, Team Diversity and Connectivity of Coastal Marine Landscapes, Roscoff, France

²Sorbonne Universités, UPMC Univ Paris 06, Roscoff, France

³Centre National de la Recherche Scientifique, Roscoff, France

⁴Groupe d'Etude des Milieux Estuariens et Littoraux de Normandie, Centre Régional d'Etudes Côtières, Luc-sur-Mer, France

Correspondence

Ambre Ribardière, Centre National de la Recherche Scientifique, Team Diversity and Connectivity of Coastal Marine Landscapes, Roscoff, France.

Email: ambre.ribardiere@sb-roscoff.fr

Funding information

French Agence Nationale de la Recherche, Grant/Award Number: ANR-13-JSV7-0001-01

Abstract

Sympatric species that in some populations hybridize and in other populations remain reproductively isolated open interesting research possibilities for the study of hybridization and speciation. Here, we test for such a situation in two littoral isopods (*Jaera albifrons* and *J. prae-hirsuta*) that occur in mixed populations and where past morphological descriptions suggested that the two species are generally reproductively isolated except in rare populations where hybridization may be happening. Using field surveys and microsatellite genetic structure analyses in two regions from France (Normandy and Brittany), we confirmed that introgressive hybridization occurs in a subset of mixed *J. albifrons*/*J. prae-hirsuta* populations (region Normandy) where the two species are found in the same habitat (pebbles on the shore). Moreover, we found that introgression in these populations is differential, 21 of 23 microsatellite markers showing little genetic divergence between species (hierarchical analysis of molecular variance $F_{CT} = 0.017$) while the remaining two loci were strongly differentiated ($F_{CT} = 0.428$). By contrast, *J. albifrons* and *J. prae-hirsuta* in mixed populations from region Brittany occupied distinct habitats (pebbles and seaweeds, respectively) with little overlap and showed stronger genetic divergence ($F_{CT} = 0.132$). In hybridizing populations, the majority of individuals show morphological traits that are characteristic of one or the other species. This raises the question of the forces that act to maintain this polymorphism, noting that hybridizing populations seem to be geographically isolated from potential source parental populations and show no detectable habitat divergence between species.

KEYWORDS

crustaceans, genetic incompatibilities, introgression, isolating barriers, Mosaic hybrid zone, sexual isolation

1 | INTRODUCTION

Natural hybridization events inform our understanding of isolating barriers between species, the conditions of species coexistence despite

hybridization, and the mechanisms of speciation. The archetypal hybrid zone structure is a region of contact between two otherwise allopatric species. In such hybrid zones, flanked on one side by populations of one species and on the other side by the other species, the dynamics

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

of the system is most often driven by a balance between immigration from pure parental populations and selection against hybrids (the tension zone model, Barton & Hewitt, 1985). In such systems, individuals freely hybridize in the contact zone and hybrids have reduced fitness due to the segregation of genetic incompatibilities. Studies of naturally hybridizing populations have also increasingly highlighted the role of other isolating barriers, including environmentally induced selection (Arnold, 1997; Endler, 1977; Moore, 1977) and sexual isolation (Poelstra et al., 2014; Seehausen, vanAlphen, & Witte, 1997).

Hybridizing populations vary not only in the nature of isolating mechanisms that are involved but also in geographic structure. Hybrid zones are typically characterized by a clinal structure (gradients of allelic frequencies between pure parental populations). More complex spatial structures are found when the environment induces differential selection on hybridizing species and the distribution of habitats is discrete (e.g., islands, lakes, host plants) or otherwise heterogeneous, leading to patchy hybrid zones (mosaic hybrid zones, Harrison & Rand, 1989; and other types of replicated hybridizing populations, reviewed, e.g., in Harrison & Larson, 2016). Variable degrees of patchiness can also be induced by colonization history or population stochasticity (Gompert, Lucas, Fordyce, Forister, & Nice, 2010). Whatever causes patchiness, patchy systems allow us to compare multiple, potentially independent contact zones (Bierne et al., 2003; Butlin et al., 2014; McKinnon & Rundle, 2002). Such comparisons are also possible in hybrid zones that have a simpler spatial structure but that can be sampled along replicated transects (e.g., Teeter et al., 2010), and, notably, in experimental populations (Pritchard & Edmands, 2013). These comparative analyses may increase our understanding of isolation mechanisms, their associated genomic architecture, and, promisingly, speciation (Harrison & Larson, 2016; Westram, Panova, Galindo, & Butlin, 2016).

Particularly intriguing are the situations where one can compare populations composed by a mixture of individuals of two species that in some instances hybridize and in other instances remain strongly reproductively isolated. That is, sympatric or mixed populations that may or may not be reproductively isolated; hereafter, we will use the term “mixed populations,” defined as populations where individuals of two species are close enough so that they can meet and interact frequently. An illustrative case in point is the lake Victoria cichlid “speciation transect” (Seehausen, 2009) where mixed populations of *Pundamilia pundamilia* and *P. nyererei* show more or less hybridization depending on variations in premating behavioral mechanisms themselves linked with variations in habitat (water clarity). Fish studies have provided a few other related examples where a pair of species shows contrasted levels of reproductive isolation when in sympatry (benthic and limnetic threespined sticklebacks, Taylor et al., 2006; swordtail fish, Culumber et al., 2011; and lake whitefish, Gagnaire, Pavey, Normandeau, & Bernatchez, 2013; river and blueback herring, Hasselman et al., 2014; and river and brook lampreys, Rougemont et al., 2015). Comparing sympatric non-hybridizing/hybridizing populations provides power to interpret admixture patterns (e.g., shared ancestral polymorphism vs. current gene flow) or assess whether differential introgression patterns are due to heterogeneous recombination, selection, or gene flow (Gagnaire et al., 2013; Powell et al., 2013; Rougemont et al., 2016).

Here we focus on the *Jaera albifrons* group, a complex of five marine isopod species that live on the shores of the temperate and cold waters of the North-Atlantic Ocean. It includes *J. albifrons*, *J. prae-hirsuta*, *J. ischiosetosa*, *J. forsmani*, and *J. posthirsuta* (Bocquet, 1953, 1972; Naylor & Haahtela, 1966). Note that *Jaera albifrons* designates one of the five species of the *Jaera albifrons* group (the distinction will be noted using the words “complex” or “group” throughout). All five species occupy a narrow but geographically extended belt in the intertidal zone and they have largely overlapping distribution ranges. In short, individuals from one species frequently coexist with individuals from at least one other species throughout their distribution range, and mixed populations are the rule rather than the exception. In this context, the five species of the *Jaera albifrons* group were shown to be reproductively isolated by at least three types of barriers: (1) ecological isolation (variations in local habitat preferences along the seashore), (2) sexual isolation (differences in male secondary sexual traits used in tactile courtship, and strong female-driven mate choice), and (3) genetic incompatibilities (reviewed in Solignac, 1978, 1981; Mifsud, 2011). The reproductive isolation resulting from the combination of these pre- and postzygotic barriers is thought to be very strong in nature.

However, intermediate male sexual traits have been reported in a few populations, suggesting that hybridization may happen in some rare places (Solignac, 1978). One such potentially hybridizing population has been intensively studied by Charles Bocquet and Michel Solignac between 1965 and 1970. They described a *Jaera albifrons/J. prae-hirsuta* mixed population located in Luc-sur-Mer, Normandy (France) where the analysis of male secondary sexual traits and experimental crosses led them to conclude that this population contained an exceptional proportion of hybrids (15%–32% depending on sampling event and classification thresholds, Bocquet & Solignac, 1969; Solignac, 1969a,b, 1978). Based on morphological descriptions for a large number of individuals sampled or raised in the laboratory from this population (nearly 2000 ind., Solignac, 1978) and comparison with experimental crosses (Bocquet & Solignac, 1969), their conclusion on hybridization between species seems very strong. This past work convincingly suggests that hybridization was occurring in at least one particular site in 1965–1970 while all other known *J. albifrons/J. prae-hirsuta* sympatric populations were reproductively isolated. As discussed above, such a situation seems interesting for the study of isolating barriers and speciation.

With this study, our objectives were (1) to test for hybridization between *J. albifrons* and *J. prae-hirsuta* using genetic tools, (2) to investigate the geographic structure of hybridizing populations and the nature of the isolating mechanisms, and (3) to compare genetic patterns within hybridizing vs. nonhybridizing mixed populations. For this purpose, we searched for mixed populations and morphologically intermediate individuals as described in Normandy ca. 50 years ago, analyzed the genetic structure of local populations using a panel of 23 microsatellite loci, and compared it with mixed populations from another French region (Brittany) where the two species had been described as reproductively isolated.

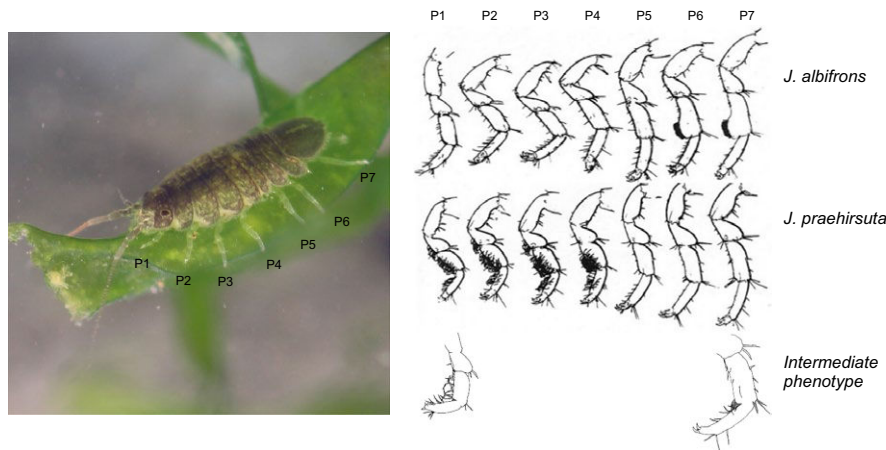


FIGURE 1 Morphological differentiation at peraeopods (numbered P1-P7) between males *Jaera albifrons* (carpus of peraeopods P6 and P7 extended as a lobe with a number of straight setae) and *J. praeahirsuta* (peraeopods P1-4 with many curved setae) as found in region Brittany (western France, drawings reproduced from Solignac, 1981 with authorization). A few individuals with secondary sexual traits intermediate to *J. albifrons* and *J. praeahirsuta* were found in region Normandy. The length of the individual (a female) represented on the picture is 4 mm. Photography credit to Guillaume Evanno & Thomas Broquet

2 | MATERIALS AND METHODS

2.1 | Model species

Members of the *Jaera albifrons* complex are small marine crustaceans (2–5 mm total adult length, Figure 1). Of the five species comprising the *Jaera albifrons* complex, two are restricted to the temperate waters of the North-American east coast (*J. posthirsuta*) or the European coast (*J. forsmanni*), while the three others are more widely distributed on both sides of the Northern Atlantic (Bocquet, 1972). These five species are found in abundance in the intertidal zone, where they can show local habitat preferences involving variations in micro-habitat (under rocks or on seaweeds), level on the intertidal zone, and salinity (Jones, 1972; Naylor & Haahtela, 1966). However, these habitat preferences vary widely, meaning that ecological isolation is also very variable (Solignac, 1981; Veuille, 1976).

The identification of species within the *Jaera albifrons* complex is based on male secondary sexual traits (Figure 1). Mating is preceded by a courtship behavior whereby males mount females in a head-to-tail position and used different parts of their peraeopods to brush or press the female's back. Males of the five species differ in the distribution of setae and spines borne by the peraeopods used to court females (Jones & Fordy, 1971; Solignac, 1978), and a female's acceptance or rejection is a major driver of reproductive isolation between species.

2.2 | Species survey and sampling

We sampled *J. albifrons* and *J. praeahirsuta* in two regions. First we focused on the area where Michel Solignac had described hybridization between these two species in 1965 and 1970 (Solignac, 1978). For that, we surveyed strictly all potential habitats on a 25-km portion of the coast around this original site, finding the population studied by M. Solignac to be extinct (Luc-sur-Mer, Figure 2), possibly due to the regular removal of pebbles from the beach for touristic activities. We extended

this survey 35 km East and 35 km West by visiting a large number of (but not strictly all) potential habitats, where we found three sites with a mixture of *J. albifrons* and *J. praeahirsuta* (sites 7–9, see Results, Table 1, and Figure 2). This gave us a 95-km continuous portion of the coastline where we have a precise, although not strictly exhaustive, view of the distribution of species (from Grandcamp-Maisy to Honfleur, Second World War landing beaches, highlighted in yellow in Figure 2).

Less intensive surveys were more recently conducted further West and East in order to check for additional mixed or pure *J. albifrons*/*J. praeahirsuta* populations (such populations were found but not analyzed in this study, see Results).

Second, we searched for similarly mixed populations of the same pair of species in a region where no hybridization had been found despite extensive field studies (area around Roscoff biology station, Brittany, France, Figure 2, Bocquet & Solignac, 1969; Solignac, 1969b). In this region, we selected six sampling sites where the two species were found at the same location (Figure 2 and Table 1).

In both regions (Brittany and Normandy), we looked for individuals of the *Jaera albifrons* complex under rocks and on seaweeds (*Ascophyllum nodosum*, *Fucus vesiculosus*, and *Fucus serratus* essentially) in the intertidal zone. Animals found on rocks were collected in the field using a small brush. By contrast, samples of seaweeds were brought back to the laboratory where we checked for the presence of *Jaera* individuals by shaking algae repeatedly in freshwater (Solignac, 1978). All individuals where kept alive until identification based upon observation of male secondary sexual traits. Because females of all species are morphologically identical, this study is based on male individuals only. As adult females are larger than males, many females could be left alive in the field. All collected individuals were fixed in ethanol after species identification.

Finally, one sampling site (Ste-Honorine-des-Pertes, site 8 in Figure 2) was selected for a detailed analysis of the micro-distribution of individuals. In this site, we performed an exhaustive survey on a large portion of the beach, recording the precise localization of each individual with a Trimble GeoExplorer 6000 GPS (average horizontal accuracy

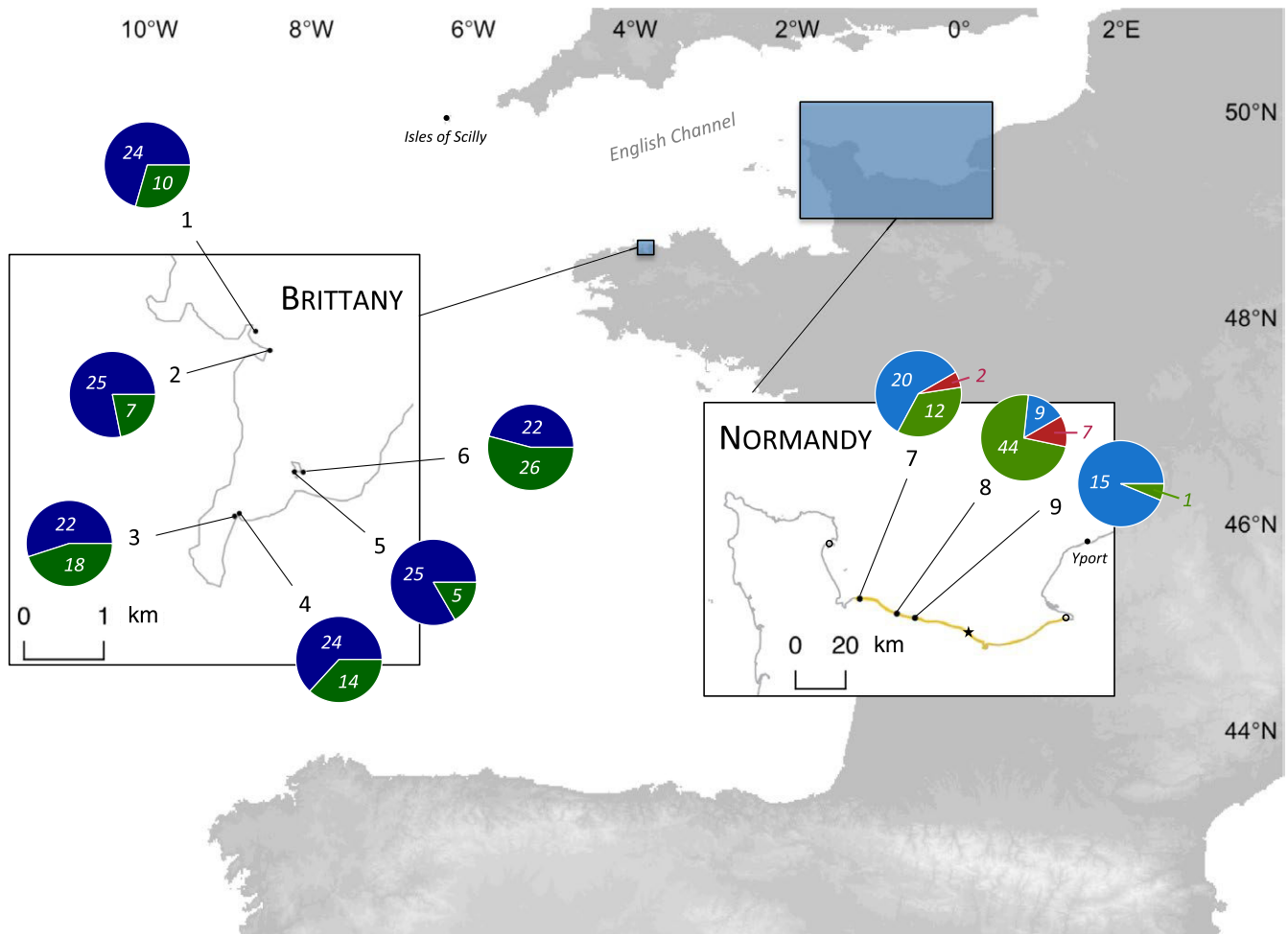


FIGURE 2 Sampling sites in two regions of Western France (“Brittany” and “Normandy”). Pie charts give the number of males showing secondary sexual traits typical of *Jaera albifrons* (in blue), *Jaera praeahirsuta* (in green), and intermediate phenotypes (in red) sampled at each location. Note that these numbers reflect the relative proportions of each species at each site in Normandy, where the two species occupy the same microhabitat and cannot be distinguished in the field. By contrast, in Brittany, pie charts are representative of sample sizes but not necessary of the relative density of each species (because there the two species occupy two different habitats with little overlap, see text). Mixed populations from Normandy were found following an intensive survey (geographic extent shown in yellow, details in text), revealing that the hybridizing population originally studied by Michel Solignac in 1965–1970 (Luc-sur-Mer, indicated by a star) is now extinct. The nearest nonmixed populations that we could find comprised only *Jaera albifrons* (empty circles). Additional mixed populations with some individuals showing an intermediate phenotype were further found in Yport and the Isles of Scilly (black dots, see text)

55 mm). There we also collected females, which were kept in the laboratory until they produced offspring (sperm storage allows females to produce offspring in the absence of males). These offspring were reared in the laboratory for at least 6 weeks, until each individual could be sexed and each male could be identified based on secondary sexual traits. This identification of series of male sibs gave a presumptive species identification for their mother (e.g., a female could be classified as *J. albifrons*, *J. praeahirsuta*, or “hybrid” if it gave birth to a series of males bearing *J. albifrons*, *J. praeahirsuta* or intermediate morphological traits, respectively).

2.3 | Genetic analyses

All genetic analyses are based on individual genotypes obtained at 23 microsatellite loci (all loci described in Ribardière, Broquet, & Daguin-Thiebaut, 2015). DNA extraction and amplification followed

the protocols described in Ribardière et al. (2015). Pairwise linkage disequilibrium between loci was tested in *FSTAT* version 2.9.3.2 (Goudet, 2001) in order to avoid redundant information. Departure from Hardy–Weinberg equilibrium was also tested in *FSTAT* in order to detect technical artifacts (null alleles or allelic dropouts) or departure from random mating within sampling sites. The occurrence of null alleles, already detected in the *Jaera albifrons* complex with these microsatellites (Ribardière et al., 2015), was specifically investigated with the software *MICROCHECKER* (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004). About 10% of the genotypes were randomly replicated in order to evaluate the frequency of genotyping errors and for each locus genotyping error rate was calculated as error rate = (number of false genotypes)/(total number of repeated genotypes). The level of polymorphism was estimated by measuring observed and expected heterozygosity in *FSTAT*.

TABLE 1 *Jaera albifrons* and *J. praeirsuta* sampling locations, sample sizes, and genetic diversity (observed and expected heterozygosity H_o and H_e , and F_{IS}). Individuals with an intermediate phenotype ($n = 9$) from sites 7 and 8 are not included. The statistical significance of F_{IS} is indicated (*: p -value < .05)

Sampled site	Coordinates	<i>J. albifrons</i>				<i>J. praeirsuta</i>			
		N	H_o	H_e	F_{IS}	N	H_o	H_e	F_{IS}
Brittany									
1	48°40'27.6"N, 3°57'11.3"W	24	0.48	0.55	0.137*	10	0.53	0.61	0.145
2	48°40'20.9"N, 3°57'00.3"W	25	0.47	0.57	0.188*	7	0.55	0.64	0.142
3	48°39'10.8"N, 3°57'03.0"W	22	0.51	0.54	0.070	18	0.54	0.62	0.135*
4	48°39'12.3"N, 3°57'00.4"W	24	0.5	0.57	0.131*	14	0.52	0.62	0.168*
5	48°39'34.1"N, 3°56'25.7"W	25	0.48	0.55	0.115	5	0.63	0.69	0.097
6	48°39'33.5"N, 3°56'31.2"W	22	0.45	0.52	0.132*	26	0.5	0.58	0.153*
Normandy									
7	49°23'30.4"N, 1°02'09.6"W	20	0.49	0.52	0.059	12	0.45	0.53	0.163*
8	49°21'15.7"N, 0°47'54.2"W	9	0.42	0.48	0.141	44	0.52	0.58	0.107*
9	49°20'53.5"N, 0°41'03.2"W	15	0.43	0.47	0.094	1	-	-	-

Our test of reproductive isolation or hybridization between *J. albifrons* and *J. praeirsuta* within our two sampling regions (Brittany and Normandy) is based upon estimates of genetic differentiation. We estimated the distribution of genetic variance among sampling sites within a species (F_{SC}) and between species (F_{CT}) in a hierarchical analysis of molecular variance (AMOVA, Excoffier, Smouse, & Quattro, 1992) implemented in ARLEQUIN version 3.5 (Excoffier & Lischer, 2010). We performed this analysis independently within each region. The between-species component F_{CT} will thus inform us on the strength of genetic differentiation between *J. albifrons* and *J. praeirsuta* within each region. Small samples were not included in these analyses (i.e., individuals with an intermediate phenotype, and individuals from site 9, where one of the two species was represented by only one individual). These analyses were performed using both allelic frequencies (F_{ST} -like) and the distance between alleles (R_{ST} -like), and significance was assessed using 10,100 permutations as implemented in ARLEQUIN. We ran these analyses first using all markers (global AMOVA) and then for each microsatellite locus independently (locus-by-locus AMOVA).

Because some loci showed strongly contrasted levels of between-species genetic structure (F_{CT}) in Normandy vs. Brittany (see results), the significance of the difference between $F_{CT-Normandy}$ against $F_{CT-Brittany}$ was tested by bootstrapping individuals 1,000 times in R version 3.2.3 (R Core Team, 2016). This allowed us to estimate how often the two F_{CT} values obtained from a given resampled dataset overlapped, and thus, whether $F_{CT-Normandy}$ differed significantly from $F_{CT-Brittany}$ at the locus tested.

Pairwise estimates of genetic differentiation between samples were also obtained in a nonhierarchical model in ARLEQUIN (i.e., F_{ST} between all pairs of populations, where a population is defined by a given species in a given sampling site, Table 1). This is useful (1) to evaluate whether the global differentiation between species is consistent across sampling sites (i.e., using F_{ST} between species within each site separately), and (2) to investigate genetic structure within each species separately, in particular by testing for isolation by distance between

sampling sites with SPAGEDI 1.4 (Hardy & Vekemans, 2002) and GENEPOP 4.2.2 (Rousset, 2008).

Finally, the distribution of genetic variance was also investigated using individual analyses without a priori grouping of samples. We first ran a clustering analysis within each region using STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) with an admixture model (10 independent repetitions, burn-in period = 50,000, MCMC = 300,000). The most likely number of clusters (K) was determined via HARVESTER v.0.6.1 (Earl, 2012) using ΔK as described by Evanno, Regnaut, and Goudet (2005). The AncestryDist option was used in STRUCTURE to calculate 95% probability intervals for an individual's membership q to each cluster. The width of such intervals (difference between upper and lower bounds) gave us an estimate of the precision of membership values. Second, we performed a principal component analysis (PCA) based on all individual genotypes using the R package ade4 (Dray & Dufour, 2007). Because the locus-by-locus AMOVA showed that two loci had a striking behavior (see Results), we ran both analyses (PCA and STRUCTURE) with and without these two loci.

3 | RESULTS

3.1 | Population survey

In Brittany, we sampled 142 *J. albifrons* and 80 *J. praeirsuta* in six sites where the two species were co-occurring (Figure 2 and Table 1, only males are considered throughout the paper, unless stated otherwise). In these sites, all *J. albifrons* but two were found under rocks, while all *J. praeirsuta* except three were found on seaweeds (located directly above rocks or within a radius of a few meters). In this region, all individuals could be morphologically identified at the species level without any overlap of traits (i.e., no individuals with intermediate morphology).

In Normandy, we sampled 44 *J. albifrons* and 57 *J. praeirsuta* from the three mixed populations that we found (sites 7–9, Figure 2).

TABLE 2 Distribution of the genetic variation estimated through hierarchical analyses of molecular variance in regions Brittany and Normandy (as defined in Figure 1). We present the results obtained with and without loci Ja41 and Ja64, which show a very strong differentiation in Normandy (see F_{CT} at these two loci)

Source of variation	Brittany			Normandy		
	% of total variation	F -stat	p -Value	% of total variation	F -stat	p -Value ^a
23 loci						
Among sampling sites						
Between species	13.2	$F_{CT} = 0.132$.001	7.4	$F_{CT} = 0.074$.336
Within species	1.3	$F_{SC} = 0.015$	<.001	1.5	$F_{SC} = 0.016$.003
Within sampling sites						
Among individuals	9.7	$F_{IS} = 0.114$	<.001	8.8	$F_{IS} = 0.097$	<.001
Within individuals	75.8		<.001	82.3		<.001
21 loci (without Ja41 & Ja64)						
Among sampling sites						
Between species	12.5	$F_{CT} = 0.125$.002	1.7	$F_{CT} = 0.017$.328
Within species	1.2	$F_{SC} = 0.013$	<.001	1.8	$F_{SC} = 0.019$.001
Within sampling sites						
AMONG individuals	10.1	$F_{IS} = 0.117$	<.001	8.4	$F_{IS} = 0.087$	<.001
Within individuals	76.3		<.001	88.1		<.001
2 loci (Ja41 and Ja64)						
Among sampling sites						
Between species	20.2	$F_{CT} = 0.202$.003	42.8	$F_{CT} = 0.428$.332
Within species	2.5	$F_{SC} = 0.032$	<.001	-0.7	$F_{SC} = -0.012$.65
Within sampling sites						
Among individuals	8.9	$F_{IS} = 0.115$	<.001	11.6	$F_{IS} = 0.2$	<.001
Within individuals	68.4		<.001	46.3		<.001

^aThe statistical significance of between-species variation was tested using permutations of sites between species, which is essentially powerless in region Normandy where only two sites harboring mixed populations were found. The relevant p -values (indicated in italic) are thus meaningless, and the differentiation between species was better tested in this case using nonhierarchical F -statistics within each site (see text).

Contrary to the Brittany situation, all individuals were collected under rocks, while no individuals could be found on seaweeds. In addition, we found nine individuals clearly showing intermediate morphological traits as described by Solignac (Figure 1, "hybrid types" 5–13 in Solignac, 1978; p.172–177). These individuals will be referred to as "intermediate phenotypes" hereafter. The portion of the coast that was intensively surveyed revealed a single pure *J. albifrons* population located >70 km East of mixed populations (empty circle in Figure 2).

Later, less intensive surveys revealed another pure *J. albifrons* population 30 km West (Figure 2), and interestingly, one additional mixed *J. albifrons*/*J. praeahirsuta* population (with intermediate phenotypes) further East (location Yport, Figure 2). No *J. praeahirsuta* individuals were found outside of mixed populations anywhere in Normandy.

The fine-scale distribution of individuals at site 8 (Ste-Honorine-des-Pertes, Normandy) showed that individuals (61 males and 138 females) of the two species and intermediate phenotypes were largely intermingled, with *J. praeahirsuta* being distributed all along the shore while the distribution of *J. albifrons* was more irregular (Figure S1).

3.2 | Genetic diversity

All microsatellite loci could be amplified in individuals of the two species (consistent with Ribardièrre et al., 2015) as well as in individuals with intermediate phenotypes. There was no linkage disequilibrium after Bonferroni's correction between all pairs of loci in each population, so that all 23 markers were kept for further analyses. The level of polymorphism was globally consistent across species (Table 1). Unless stated otherwise, individuals with intermediate phenotypes were removed from the following analyses for we had too few of them (2 in Grandcamp, site 7, and 7 in Ste-Honorine-des-Pertes, site 8, Figure 2). Observed heterozygosity H_o (0.45 in *J. albifrons* and 0.51 in *J. praeahirsuta*) was on average lower than within population gene diversity H_e (0.51 and 0.59, respectively), resulting in a significant departure from Hardy-Weinberg equilibrium (F_{IS} values in Table 1). Departure from HWE was driven in Brittany by loci Ja37, Ja39, Ja55, and Ja94, and in Normandy by locus Ja55 (loci with significant positive F_{IS} in one to six samples, data not shown). These loci, except Ja94, showed signs of a null allele in more than half of populations, as tested using

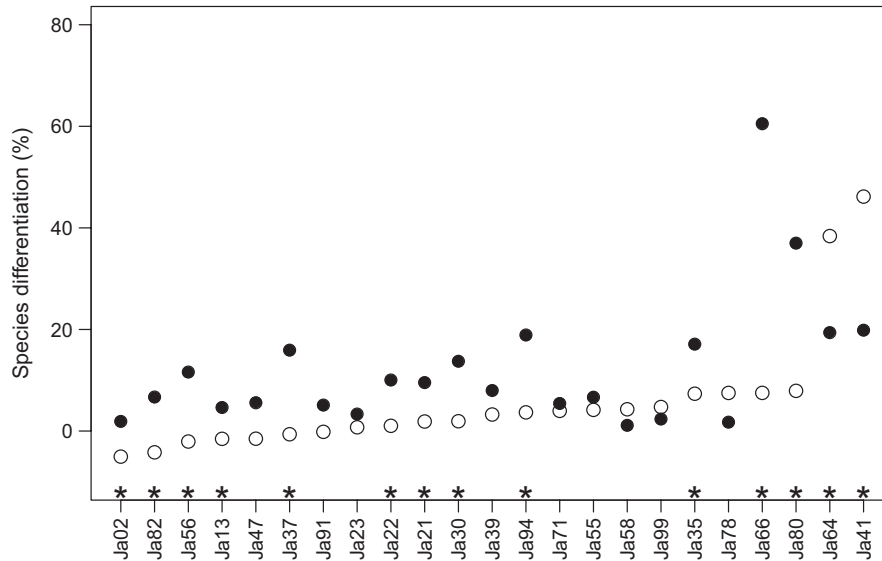


FIGURE 3 Locus-by-locus genetic differentiation between *J. albifrons* and *J. praeheirsuta* (F_{CT} , expressed in this figure as the percentage of genetic variation due to differences between species) in Brittany (black dots) and in Normandy (white dots). Significant differentiation between $F_{CT-Brittany}$ and $F_{CT-Normandy}$ is represented by a star above the locus name. Here the loci are arranged by increasing order of $F_{CT-Normandy}$. We see that all white dots but two lie near zero (no differentiation between species in Normandy) while the two remaining loci (Ja41 and Ja64) show a strong differentiation. These two microsatellites are the only loci showing a significantly stronger differentiation between species in Normandy than in Brittany (white vs. black dots). In Brittany, the average level of differentiation between species is near 13% and there is also heterogeneity across loci

MICROCHECKER. While we did not detect locus-specific HW disequilibrium patterns that were consistent across populations, the main downstream quantitative analysis (analysis of molecular variance) was run with and without the four loci cited above (and we report locus-specific results as well). Genotyping error rates estimated from replicated individuals ranged from 0% to 5.88% per locus (average 1.7%) and were due roughly equally to allelic dropouts and false alleles.

3.3 | Genetic differentiation between *J. albifrons* and *J. praeheirsuta*

The hierarchical analysis of molecular variance based on 23 microsatellite loci (Table 2) showed that the between-species differentiation was higher in Brittany ($F_{CT-Brittany} = 0.132$, $p < .005$) than in Normandy ($F_{CT-Normandy} = 0.074$, $p = .34$). Accordingly, the locus-by-locus AMOVA showed that most loci were less differentiated between species in Normandy than in Brittany (Figure 3). However, two loci (Ja41 and Ja64) revealed a strikingly different pattern: These two loci showed a very strong level of differentiation (Ja41, $F_{CT-Normandy} = 0.462$; Ja64, $F_{CT-Normandy} = 0.384$) in Normandy region while the remaining 21 loci showed no or little differentiation (locus specific $F_{CT-Normandy}$ ranged from -0.05 to 0.079), Figure 3). Without these two peculiar loci, there is thus no genetic differentiation between species in Normandy ($F_{CT-Normandy} = 0.017$, $F_{CT-Brittany} = 0.125$, Table 2). Note that this result is unchanged when removing the four loci that showed a departure from HW equilibrium in some populations (not shown).

Moreover, the differentiation between species at loci Ja41 and Ja64 in Normandy was not only much stronger than at other loci but it was also stronger than the differentiation observed at the same two loci in

Brittany (Ja41, $F_{CT-Brittany} = 0.199$, Ja64, $F_{CT-Brittany} = 0.194$, Figure 3), and this difference was significantly different from 0 (bootstrap p -value $< .001$ for Ja41 and $p = .001$ for Ja64, Figures S2 and S3). The distribution of allelic frequencies at these two loci is presented in Figures S4 and S5.

Although the heterogeneity across loci appeared somewhat lower in Brittany (Figure 3), there was nonetheless some variation, with two other loci showing a particularly strong level of differentiation (Ja66, $F_{CT-Brittany} = 0.605$; Ja80, $F_{CT-Brittany} = 0.37$). The difference in F_{CT} between regions was significant at these two loci (bootstrap p -value $< .001$ for Ja66 and Ja80, Figure S3).

Because we studied only three mixed populations in Normandy, and one of them contained nearly only *J. albifrons* (site 9, Longues-sur-Mer, Figure 2, Table 1), the permutation procedure implemented in ARLEQUIN to test for the significance of $F_{CT-Normandy}$ is essentially powerless (the between-species component of genetic variation is tested by permuting populations within species). We therefore checked whether the patterns found in the AMOVA (which considers all sites simultaneously) were consistent across sites. Tables 3 and 4 show pairwise F_{ST} values calculated in a simple nonhierarchical framework. Most importantly, it shows that the between-species differentiation was consistent across sites, both in Normandy (F_{ST} between species equal to 0.104 at Grandcamp, site 7, and 0.101 at Ste-Honorine-des-Pertes, site 8) and in Brittany (F_{ST} between species within sites in $[0.1; 0.19]$). These results consider all loci, but the same geographical consistency is observed when considering the locus-specific patterns described above (data not shown). That is, the global AMOVA results are repeatable across sites (e.g., F_{ST} at the two sites from Normandy = 0.546 (site 7) and 0.396 (site 8) when considering only loci Ja41 and Ja64, and $F_{ST} = 0.05$ (site 7) and 0.056 (site 8) with all other loci).

TABLE 3 Pairwise genetic differentiation between populations in Brittany. Above diagonal: pairwise F_{ST} . Below diagonal: exact test of population differentiation p -value (in bold when significant). Values in the gray area correspond to inter-specific F_{ST}

	<i>J. albifrons</i>						<i>J. prae-hirsuta</i>					
	1	2	3	4	5	6	1	2	3	4	5	6
<i>J. albifrons</i>												
1	–	0.02	0.061	0.06	0.035	0.034	0.178	0.161	0.14	0.14	0.062	0.167
2	0.003	–	0.066	0.055	0.036	0.05	0.175	0.151	0.137	0.141	0.059	0.161
3	0	0	–	0.005	0.021	0.017	0.182	0.177	0.155	0.149	0.099	0.187
4	0	0	0.190	–	0.015	0.024	0.174	0.167	0.141	0.14	0.092	0.179
5	0	0	0.003	0.014	–	0.009	0.207	0.195	0.17	0.166	0.1	0.196
6	0	0	0.008	0.001	0.116	–	0.201	0.184	0.157	0.152	0.103	0.19
<i>J. prae-hirsuta</i>												
1	0	0	0	0	0	0	–	–0.007	0.01	–0.005	–0.048	0.017
2	0	0	0	0	0	0	0.671	–	–0.005	–0.004	–0.068	0.008
3	0	0	0	0	0	0	0.160	0.715	–	–0.008	–0.061	0.004
4	0	0	0	0	0	0	0.737	0.657	0.919	–	–0.074	0.005
5	0.004	0.003	0	0	0	0	0.979	0.988	0.999	0.999	–	–0.075
6	0	0	0	0	0	0	0.042	0.278	0.295	0.268	1	–

TABLE 4 Pairwise genetic differentiation between populations in Normandy. Above diagonal: pairwise F_{ST} . Below diagonal: exact test of population differentiation p -value (in bold when significant). Values in the gray area correspond to inter-specific F_{ST}

	<i>J. albifrons</i>		<i>J. prae-hirsuta</i>	
	7	8	7	8
<i>J. albifrons</i>				
7	–	0.025	0.104	0.073
8	0.037	–	0.153	0.101
<i>J. prae-hirsuta</i>				
7	0	0	–	0.018
8	0	0	0.015	–

Individual analyses bring some complementary information, in particular because the individuals with intermediate morphology could be included in spite of their low abundance (as well as individuals from site 9). Running STRUCTURE with $K = 2$, we found that *J. albifrons* and *J. prae-hirsuta* cluster into two clearly identified groups both in Brittany and Normandy using a panel of 23 loci (Figure 4). However, while an individual's membership q to its assigned cluster was similar for both species in Brittany and Normandy (membership averaged over all individuals and 10 STRUCTURE runs, Brittany: $\bar{q}_{albifrons} = 0.99$, $\bar{q}_{prae-hirsuta} = 0.97$, and Normandy: $\bar{q}_{albifrons} = 0.97$, $\bar{q}_{prae-hirsuta} = 0.95$), the uncertainty associated with q was larger in Normandy (average width of 95% probability interval $w_{albifrons}^{95} = 0.24$, $w_{prae-hirsuta}^{95} = 0.32$) than Brittany ($w_{albifrons}^{95} = 0.12$, $w_{prae-hirsuta}^{95} = 0.19$). As it turned out, the apparent genetic clustering in Normandy was almost entirely due to the effect of two loci only (Ja41 and Ja64), while a STRUCTURE analysis using the remaining 21 loci showed that the two species were genetically homogeneous (Figure 4b).

The results from molecular analyses of variance and clustering analyses could be well visualized using PCA performed with all 23 loci (Figure 5). The two isopod species were clearly differentiated in one region (Brittany) and less so in the other one (Normandy). We see also that individuals with intermediate phenotypes were genetically indistinguishable from individuals with *J. prae-hirsuta* traits.

The particular geographical distribution of individuals (replicates of mixed populations comprising intermediate phenotypes and absence of clinal structure, see Discussion) precluded the use of genetic tools dedicated to the analysis of hybridization in clinal hybrid zones.

3.4 | Genetic structure within species

Because there is ongoing hybridization and introgression between the two species in Normandy (see Discussion), the within-species genetic structure is best investigated using samples from Brittany. Brittany is also the region where we have more sampling locations. The genetic structure among *J. albifrons* samples appeared slightly higher than within *J. prae-hirsuta*. This is visible from pairwise F_{ST} estimates (Table 3), ranging from 0.005 to 0.066 (13 of 15 pairs significantly different) when considering only *J. albifrons* within Brittany, and –0.075 to 0.017 (one of 15 pairs significant) when considering only *J. prae-hirsuta*.

In addition, a significant pattern of isolation-by-distance (Figure S6) was observed in Brittany both in *J. albifrons* ($R^2 = 0.73$, Mantel test p -value < .01) and in *J. prae-hirsuta* ($R^2 = 0.11$, p -value = .02). In line with the pairwise F_{ST} results, genetic differentiation increased more rapidly with distance in *J. albifrons* than in *J. prae-hirsuta*, although 95% confidence intervals calculated in GENETOP overlapped (10,000 permutations, *J. albifrons* [0.013, 0.037], *J. prae-hirsuta* [–0.0005, 0.018], Figure S6).

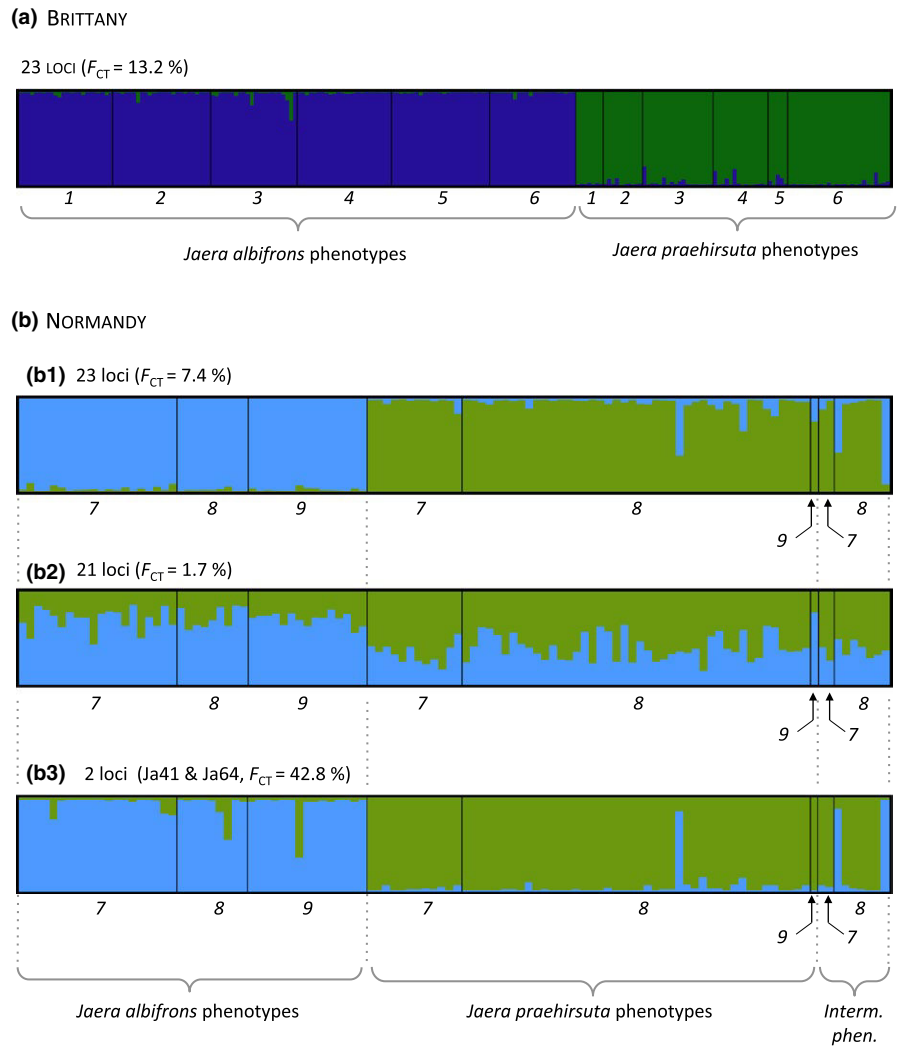


FIGURE 4 STRUCTURE clusters ($K = 2$) defined in Brittany (a) and Normandy (b). Because loci Ja41 and Ja64 are outliers in Normandy (see text and Figure 3), STRUCTURE results for this region are presented with all loci (23 loci, panel b1), without the two outliers (21 loci, panel b2), and considering only Ja41 and Ja64 (2 loci, panel b3). Numbers refer to sampling sites (Figure 2). We see that individuals morphologically identified as *J. albifrons* or *J. prae-hirsuta* cluster into two distinct groups in Brittany (regardless of sampling location, panel a), while this will remain true in Normandy only due to the effect of two markers of twenty-three (panels B1, B2, and B3)

4 | DISCUSSION

The first result of this study is the clear-cut confirmation that introgressive hybridization is happening between *J. albifrons* and *J. prae-hirsuta* in at least two mixed populations from Normandy, France. As developed below, this opens interesting questions regarding the conditions of coexistence of the two parental morphs in hybridizing populations that seem to receive no influx from pure parental populations and shows no detectable ecological heterogeneity.

4.1 | Hybridization between *Jaera albifrons* and *J. prae-hirsuta*

Analyses of molecular variance (Table 2 and Figure 3) and admixture analyses (Figures 4 and 5) both showed that mixed populations from Normandy have a homogeneous genetic structure at 21 of 23 multi-allelic loci. Critically, this genetic homogeneity contrasts with the differentiation observed in mixed populations from Brittany, where individuals bearing sexual traits specific to *J. albifrons* or *J. prae-hirsuta* cluster into clearly marked genetic groups (Figures 4a and 5). Hence, shared ancestral polymorphism cannot explain the lack of

differentiation between species in Normandy, which therefore supports the hypothesis of ongoing hybridization.

These findings agree with the conclusions reached by C. Bocquet and M. Solignac nearly 50 years ago, who studied the morphological variation of secondary sexual traits in a population from the same region (Luc-sur-Mer, Figure 2, Bocquet & Solignac, 1969; Solignac, 1969a,b, 1978 chapter 6). Similarly to the results reported by these authors, we found that in Normandy, several males have intermediate phenotypes and the two species occupy the same habitat (under stones and pebbles on the shore) while in Brittany we did not detect any intermediate phenotypes and the two species occupy two different habitats, with some overlap; *J. albifrons* lives primarily under stones, while *J. prae-hirsuta* is found primarily on seaweeds.

We conclude from these observations and the contrast in species divergence in the two regions that the two species are currently hybridizing in populations from Normandy, but not in Brittany.

4.2 | A semi-permeable barrier to gene flow

The genetic homogeneity observed across species in Normandy further shows that hybridization has been introgressive, as correctly

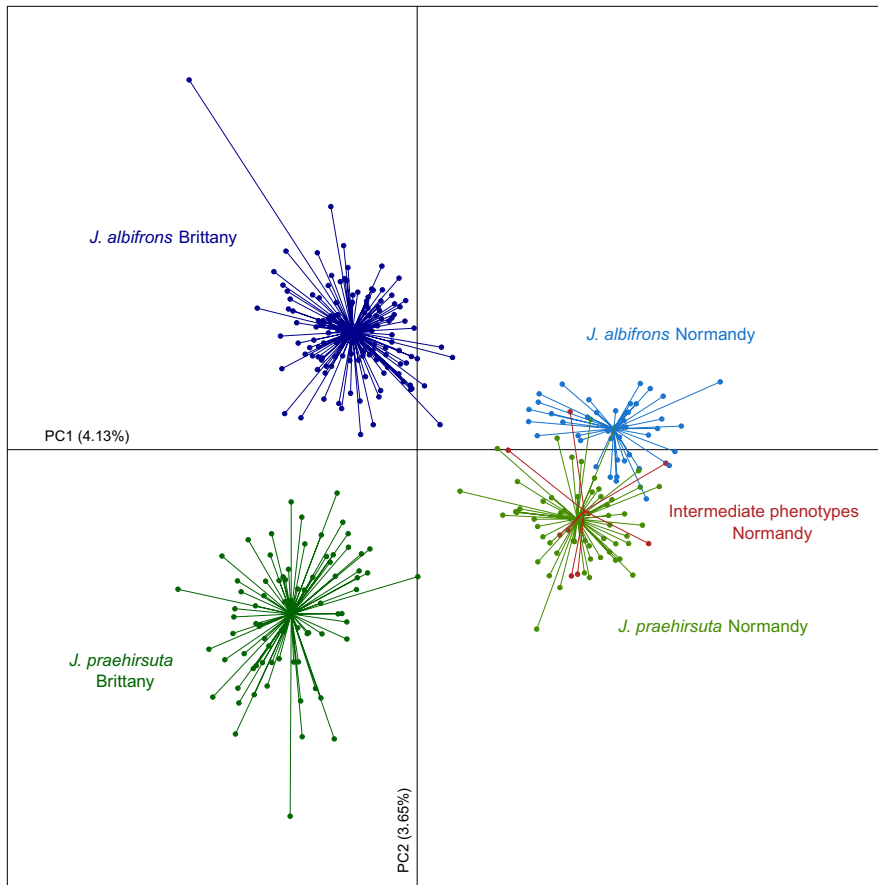


FIGURE 5 Principal component analysis based on individual multilocus genotypes at 23 microsatellite loci. The first axis separates individuals from Brittany (left) and Normandy (right). The second axis shows the genetic divergence between males bearing sexual traits typical of *J. albifrons* vs. *J. praeahirsuta*. We see that these two types of males are less genetically differentiated in region Normandy (in agreement with results from the analysis of molecular variance and STRUCTURE results from Figure 4) and males with an intermediate phenotype are undifferentiated from *J. praeahirsuta*

concluded by Solignac (1969b) from the continuous range of morphological characteristics observed in natural populations and by comparison with experimental crosses (Bocquet & Solignac, 1969). While it is now clearly established that introgression proceeds differentially across loci in hybrid zones (with no known exceptions, Harrison & Larson, 2014), investigating this variation was not part of our original plan with this study given that we were using a panel of only 23 loci. However, the locus-by-locus AMOVA analyses revealed a surprisingly trenchant pattern, whereby 21 loci showed no differentiation at all between species in populations from Normandy (F_{CT} in $[-0.05; 0.08]$) and the two remaining loci were strongly differentiated ($F_{CT} = 0.384$ and 0.462). Moreover, these two loci were also significantly more differentiated in the hybridizing populations (Normandy) than in reproductively isolated ones (Brittany) while showing no differentiation within each species (see F_{SC} values in Table 2). This strongly suggests that there is a semi-permeable barrier to gene flow between *J. albifrons* and *J. praeahirsuta* in hybridizing populations from Normandy. Hypotheses other than a reduction in gene flow at these loci seem impossible to reconcile with the fact that the same two loci are significantly less differentiated in nonhybridizing sympatric populations. Alternate hypotheses such as the differential sorting of ancestral polymorphism or reduced variability at these loci due to a locally low recombination rate would require a history of differentiation between species whereby ancestral polymorphism or recombination have evolved differentially in the two regions studied (ca 250 km apart).

While this is theoretically possible, a more parsimonious hypothesis is that the two loci are encompassed in one or two genomic regions where interspecific gene flow is hampered because these regions are linked with one or several isolating barriers.

Additional indicators of a semi-permeable barrier to gene flow are two other markers (Ja66 and Ja80) also showing a heterogeneous pattern. As most markers, they are more differentiated in Brittany than in Normandy, but, interestingly, they show a stronger differentiation than the other loci in Brittany (Figure 3). We cannot currently make assumptions based only on these results, which emphasize the necessity to study the heterogeneity of genome with an extended set of genetic markers.

4.3 | Is introgression symmetrical?

Interspecific crosses are generally not equally likely in both directions, especially when behavioral isolation is involved (e.g., Coyne & Orr, 2004, p. 226). Such asymmetries leave specific signatures in the genome that are most easily detected by comparing uni- and bi-parentally inherited genetic variation (typically, markers from the mitochondrial and nuclear DNA, e.g., Toews & Brelsford, 2012). It would be interesting to test for asymmetric introgression of mtDNA in our system given that Bocquet and Solignac (1969) have suggested that interspecific crosses may occur more easily in one direction (female *J. praeahirsuta* × male *J. albifrons*) than the other. This result was

obtained from experimental crosses with individuals from a hybridizing population (Luc-sur-Mer, Normandy), and the asymmetry was further confirmed by Solignac (1981) using individuals from other origins (nonhybridizing populations). Yet we could not use this approach here because mtDNA analyses so far have indicated that most of the genetic variation is shared by all species of the *Jaera albifrons* complex (perhaps excluding the American species *J. posthirsuta*, which has not been included in these analyses). That is, the four European species form a polyphyletic clade (16S rDNA, Mifsud, 2011; and COI, A. Ribardière and T. Broquet, unpublished), in strong contrast with the patterns obtained with nuclear data (AFLP, Mifsud, 2011; and microsatellites, this study). There are no mitochondrial haplotypes or clades that are specific to *J. albifrons* or *J. prae-hirsuta* (not shown), and the cyto-nuclear discordance, also certainly interesting in its own right, is not informative of recent introgression directionality. The symmetry of introgression can in some cases be evaluated by taking advantage of differences between nuclear loci (differential introgression), but our microsatellite dataset is too limited for this approach. Nevertheless, results of admixture analyses show that individuals showing an intermediate phenotype share more genetic background with *J. prae-hirsuta* (Figures 4 and 5) which suggests that introgression is asymmetric (genetic variation from *J. albifrons* introgressing into the genetic background of *J. prae-hirsuta*). This can also be seen by looking at allelic frequencies at loci Ja41 and Ja64 for the phenotypically intermediate individuals, which are similar to that of individuals bearing *J. prae-hirsuta* traits and different from the frequencies observed in *J. albifrons* (Figures S4 and S5).

4.4 | Geographical structure and persistence of hybridizing populations

In a 1969 paper, C. Bocquet and M. Solignac reported that many “morphological hybrids” have been observed during the preceding 15 years in their study area of Luc-Sur-Mer (Bocquet & Solignac, 1969). There is no more suitable habitat at this site, but Solignac (1978, p. 171) mentioned that “hybrids” had been found in Ste-Honorine-des-Pertes, which is one of the two sites sampled in the present study (site 8). This means that hybridizing populations have persisted in Normandy for at least several decades. Moreover, during a recent additional survey aiming at extending the 95-km coastline region studied here, we found intermediate phenotypes in a population located roughly 100 km East of the mixed populations studied here (location “Yport”, Figure 2). Even more unexpectedly, we detected a *J. albifrons*–*J. prae-hirsuta* mixed population with some intermediate male phenotypes in the Isles of Scilly, UK, an archipelago that is located more than 400 km away across the English Channel (Figure 2). This means that hybridization between these two species is probably much more widespread than previously thought (Solignac, 1969a,b, 1978). Perhaps more importantly, this also strongly suggests that hybridizing populations have been persisting for a long time.

In this study, we identified nine individuals (of 110 males found in Normandy) showing intermediate morphological traits. While this figure depends on what one recognizes as morphologically pure or

intermediate individuals, the majority of males clearly show strict *J. albifrons* or *J. prae-hirsuta* sexual traits despite the extensive genetic introgression demonstrated here (bimodal hybrid zone, Jiggins & Mallet, 2000). We concur with Solignac's observation (1978, p. 188) that the coexistence of the two morphs in spite of introgressive hybridization is of great interest, and we discuss below the mechanisms that may allow this coexistence in the long term and in repeated areas. This part of the discussion will focus on hybridizing populations only (i.e., results from Normandy).

The closest population containing only *J. albifrons* that we found was located at more than 30 km from the hybridizing populations. Moreover, the mixed populations that we found in this region are geographically isolated from one another (there is most likely a discontinuity at least between sites 7 and 8, Figure 2). All species of the *Jaera albifrons* group have a direct development without a dispersive larval phase, and in Normandy they do not live on seaweeds, which could potentially drift across populations. Gravid females caught in the water column could occasionally be moved over a great distance, but we did not find any *J. prae-hirsuta* outside of the mixed *J. albifrons*/*J. prae-hirsuta* populations in Normandy. It is difficult to conduct an exhaustive survey over large areas for such small species and we could have missed pure *J. albifrons* (and perhaps even *J. prae-hirsuta*) populations at dispersal distance from our hybridizing populations, but we feel that it is unlikely. Given that these crustaceans do not have a dispersive larval phase, and given the patchy distribution of habitats, we infer from our surveys that hybridizing populations are replicated and patchily distributed.

The hybridizing populations analyzed in this study do not seem to be flanked by—or otherwise functionally connected to—pure parental populations. They seem to be independent replicates of hybridizing populations potentially distributed on a much larger geographical area than the one studied here (e.g., on the U.K. coast). The influx of individuals from pure parental populations of *J. prae-hirsuta* (and probably *J. albifrons*) is thus most likely not one of the forces acting to stabilize the system. This interpretation needs further testing (e.g., from additional surveys and analyses of spatial genetic structure), as if this hypothesis is confirmed, it would exclude dispersal-dependent models of hybrid zones, chief among them the tension zone model (Barton & Hewitt, 1985), which relies on a balance between immigration of parental genotypes and selection against hybrids. An alternative hybrid zone model without immigration from parental populations involves ecological variation and an advantage of hybrids in intermediate habitats (Moore, 1977). This hypothesis seems also to be excluded in our case because we were unable to detect any variation in habitat within hybridizing populations (there was no identifiable variation in the distribution of *J. albifrons*-like, *J. prae-hirsuta*-like, and morphologically intermediate individuals within a site). Other classical models are also inappropriate, for they combine the tension zone balance with ecological variation, either through environmentally induced selection against hybrids (Endler, 1977) or a patchy distribution of habitats favoring one or the other species (mosaic hybrid zones, Harrison & Rand, 1989). The literature is also rich in empirical studies of hybridizing populations that do not fit one of these classical

models, but it seems that situations where species coexist in spite of extensive introgression (i.e., bimodal hybrid zones) most often involve either an income of individuals from pure parental populations or ecological variation and habitat specialization within hybrid zones (or both). When incompletely isolated species occupy different ecological niches, comparative analyses of replicate hybrid zones have made quite clear that the maintenance of bimodality is correlated with the opportunity for ecological specialization (e.g., Culumber et al., 2011; Gagnaire et al., 2013). In the *Jaera albifrons*/*J. praeheirsuta* system, the two species are more differentiated in our populations from Brittany where they specialize in two different habitats (rocks vs. seaweeds) but, interestingly, parental forms coexist despite introgression in hybridizing populations in Normandy where there is probably no habitat specialization.

Assuming that there is no immigration from pure parental populations and no ecological variability, what evolutionary forces would allow pure *J. albifrons* and *J. praeheirsuta* phenotypes to coexist in hybridizing populations? Past work suggests two strong candidates. First, the strongest isolating barrier between species of the *Jaera albifrons* complex is sexual isolation (Solignac, 1981). The courtship behavior of males (plus perhaps unknown male characteristics such as pheromone production or other unnoticed phenotypic variation) and female preference may still partially isolate *J. albifrons* from *J. praeheirsuta* in hybridizing populations. One compelling hypothesis in this regard is that females of one species accept heterospecific mating more readily than females of the alternate species. This is nearly the rule in case of behavioral isolation (e.g., Coyne & Orr, 2004) and empirical tests suggest that this happens in the hybridizing populations studied here (although with a limited sample size, Bocquet & Solignac, 1969). The hypothesis that sexual isolation is a strong component is also in line with other examples where hybrid zones remain bimodal (reviewed in Jiggins & Mallet, 2000).

The second candidate is selection against hybrids. This hypothesis is supported by Solignac (1978, p. 186), who reported a strong hybrid breakdown in experimental F2 and backcrosses using individuals from the hybridizing population of Luc-sur-Mer, Normandy. This suggests that the *J. albifrons*/*J. praeheirsuta* hybridizing populations documented in this study persist through a balance between hybridization versus partial (perhaps asymmetrical) sexual isolation and selection against certain recombined genotypes. This situation seems to be infrequent and the conditions of persistence for such a system deserve further inquiry. In particular, it is a potential empirical example of the model proposed by M'Gonigle, Mazzucco, Otto, and Dieckmann (2012), in which sexual selection, spatial variation in local carrying capacity, and female mate-search costs allow partially divergent species to persist despite ecological equivalence. More generally, it may provide some insight into the debated role of sexual selection and sexual isolation in species divergence and coexistence. Yet it is remarkable that *J. albifrons* and *J. praeheirsuta* coexist in the long term despite their seemingly ecological equivalence. It is possible that some unknown frequency-dependent mechanism (e.g., via an action of pathogens or parasites) is acting to lower the likelihood of extinction of one or the two morphs.

4.5 | Hybridizing and nonhybridizing mixed populations

Why do *J. praeheirsuta* and *J. albifrons* hybridize when in contact in some populations and not in others? There are two obvious differences between our studied populations from Brittany and Normandy. First, where the two species were found to be nonhybridizing, they live on clearly different habitats (under rocks vs. on brown algae). The two habitats are located centimeters to meters away, and a few individuals of each species were found on the habitat favored by the other species, meaning that *J. albifrons* and *J. praeheirsuta* meet each other frequently in these sites (i.e., why we call them mixed, or sympatric, populations: The two habitats are not separated by a large geographical distance and can be reached by dispersing individuals of either species). Yet these two habitats are radically different and are bound to impose a serious barrier to gene flow between species (ecological isolation). While neither species is restricted to one of these habitats in other parts of their range, the availability of distinct habitats may play an important role facilitating the coexistence of the two species and reducing hybridization opportunities. Appropriate seaweeds can also be found in hybridizing populations from Normandy, but they may not represent a suitable habitat there (e.g., because of wave exposure). Whatever the reason, there are no *Jaera* on algae in the hybridizing populations reported here. One plausible hypothesis is thus that ecological diversification facilitates the coexistence and divergence of *J. albifrons* and *J. praeheirsuta* wherever this is possible. Another interesting idea, suggested by an anonymous reviewer, is that hybridization allows the hybrid and subsequently *J. praeheirsuta* to acquire adaptations to the array of challenges presented in switching from a seaweed to a rocky habitat. One avenue of research to tackle these questions will be to search for pure *J. praeheirsuta* populations occupying rocky habitats (and *J. albifrons* populations on seaweeds).

The second difference between hybridizing and reproductively isolated populations is their geographic location. Contrary to the idea that natural hybridization is the exception and not the rule in this system (Solignac, 1969a,b, 1978), we suggest that hybridizing populations may be found in a large geographic area. It is therefore possible that populations from two large geographic zones (one encompassing Normandy and the other one encompassing Brittany) have a distinct demographic and evolutionary history (e.g., with variations in the conditions of secondary contacts between species after periods of isolation). The intra-specific genetic differentiation between regions appeared to be similar in *J. albifrons* and *J. praeheirsuta* ($F_{ST} = 0.14$ and 0.15 , respectively) and similar in intensity to the genetic differentiation between species observed in Brittany. This is also apparent on the PCA where the first axis partitions the genotypes in function of their geographic origin (Brittany on the left, Normandy on the right, Figure 5).

Testing these ideas will require analyzing the genetic structure of the two species over a large geographical scale and surveying mixed populations for morphological or genetic signs of hybridization in different habitat conditions.

ACKNOWLEDGMENTS

This study stands on a tremendous amount of past work, notably due to Michel Solignac, who passed away during the writing of this article. While the authors did not know him personally, they wish to acknowledge his achievements. We also thank Franck Gentil for initiating T.B. to the *Jaera albifrons* complex. We thank Nicolas Bierne, Pierre-Alexandre Gagnaire, Julie Jaquière, Didier Jollivet, Thomas Lenormand, Frédérique Viard, and Glenn Yannic for discussions or comments on a previous draft. We are grateful to Denis Roze for bringing back useful observations from the Isles of Scilly. We thank the "Centre de Recherches en Environnement Côtier" (CREC, university of Caen) and the "Groupe d'Etude des Milieux Estuariens et Littoraux de Normandie" (GEMEL-Normandie) for giving us access to accommodation and laboratory space during field surveys in Normandy. This work benefited from access to the Biogenouest genomic platform at Station Biologique de Roscoff and was funded by the French Agence Nationale de la Recherche (grant ANR-13-JSV7-0001-01 to T.B.).

DATA ACCESSIBILITY

Multilocus genotypes at each sampling location are available in Dryad, <https://doi.org/10.5061/dryad.st557>.

CONFLICT OF INTEREST

None declared.

REFERENCES

- Arnold, M. L. (1997). *Natural hybridization and evolution*. Oxford University Press.
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, 16, 113–148.
- Bierne, N., Borsa, P., Daguin, C., Jollivet, D., Viard, F., Bonhomme, F., & David, P. (2003). Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology*, 12, 447–461.
- Bocquet, C. (1953). *Recherches sur le polymorphisme naturel des Jaera Marina (Fabr.) (Isopodes Asellotes): Essai de systématique évolutive* (pp. 436). Archives de zoologie expérimentale et générale Tome: Centre national de la recherche scientifique. 90 pp. 187–450.
- Bocquet, C. (1972). *La spéciation des Jaera albifrons. Etat présent des connaissances et des problèmes*. pp. 131–137.
- Bocquet, C., & Solignac, M. (1969). Etude morphologique des hybrides expérimentaux entre *Jaera (albifrons) albifrons* et *Jaera (albifrons) prae-hirsuta* (isopodes asellotes). *Archives de Zoologie Expérimentale et Générale*, 110, 435–452.
- Butlin, R. K., Saura, M., Charrier, G., Jackson, B., Andre, C., Caballero, A., ... Rolan-Alvarez, E. (2014). Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution*, 68, 935–949.
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland: Sinauer associates.
- Culumber, Z. W., Fisher, H. S., Tobler, M., Mateos, M., Barber, P. H., Sorenson, M. D., & Rosenthal, G. G. (2011). Replicated hybrid zones of *Xiphophorus swordtails* along an elevational gradient. *Molecular Ecology*, 20, 342–356.
- Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, 1–20.
- Earl, D. A. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.
- Endler, J. A. (1977). *Geographic variation, speciation, and clines*. New-Jersey, USA: Princeton University Press.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620.
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567.
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491.
- Gagnaire, P. A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, 67, 2483–2497.
- Gompert, Z., Lucas, L. K., Fordyce, J. A., Forister, M. L., & Nice, C. C. (2010). Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: Extensive admixture and a patchy hybrid zone. *Molecular Ecology*, 19, 3171–3192.
- Goudet, J. (2001). *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3), updated from Goudet 1995*. Retrieved from <http://www.unil.ch/izea/software/fstat.html>
- Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, 2, 618–620.
- Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105, 795–809.
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, 25, 2454–2466.
- Harrison, R. G., & Rand, D. M. (1989). Mosaic hybrid zones and the nature of species boundaries. In D. Otte & J. A. Endler (Eds.), *Speciation and its consequences* (pp. 111–133). Sunderland, Massachusetts, USA: Sinauer Associates, Inc.
- Hasselman, D. J., Argo, E. E., McBride, M. C., Bentzen, P., Schultz, T. F., Perez-Umphrey, A. A., & Palkovacs, E. P. (2014). Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. *Molecular Ecology*, 23, 1137–1152.
- Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, 15, 250–255.
- Jones, M. (1972). Effects of salinity on the survival of the *Jaera albifrons* Leach group of species (Crustacea: Isopoda). *Journal of Experimental Marine Biology and Ecology*, 9, 231–237.
- Jones, M. B., & Fordy, M. R. (1971). A Stereoscan electron microscope study of male reproductive characters in the *Jaera albifrons* group of species (Crustacea: Isopoda). *Marine Biology*, 10, 265–271.
- McKinnon, J. S., & Rundle, H. D. (2002). Speciation in nature: The threespine stickleback model systems. *Trends in Ecology & Evolution*, 17, 480–488.
- M'Gonigle, L. K., Mazzucco, R., Otto, S. P., & Dieckmann, U. (2012). Sexual selection enables long-term coexistence despite ecological equivalence. *Nature*, 484, 506–509.
- Mifsud, D. V. (2011). *The genetic basis of speciation in the Jaera albifrons species group of intertidal isopods* (Vol. PhD. pp. 201). University of Aberdeen.
- Moore, W. S. (1977). Evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology*, 52, 263–277.
- Naylor, E., & Hahtela, I. (1966). Habitat preferences and interspersions of species within the superspecies *Jaera albifrons* Leach (Crustacea: Isopoda). *Journal of Animal Ecology*, 35, 209–216.
- Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Muller, I., ... Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, 344, 1410–1414.

- Powell, T. H. Q., Hood, G. R., Murphy, M. O., Heilveil, J. S., Berlocher, S. H., Nosil, P., & Feder, J. L. (2013). Genetic divergence along the speciation continuum: The transition from host race to species in *Rhagoletis* (Diptera: Tephritidae). *Evolution*, *67*, 2561–2576.
- Pritchard, V. L., & Edmands, S. (2013). The genomic trajectory of hybrid swarms: Outcomes of repeated crosses between populations of *Tigriopus californicus*. *Evolution*, *67*, 774–791.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.
- Ribardière, A., Broquet, T., & Daguin-Thiebaut, C. (2015). Microsatellite markers for the *Jaera albifrons* species complex (marine isopods). *BMC Research Notes*, *8*, 632.
- Rougemont, Q., Gaigher, A., Lasne, E., Cote, J., Coke, M., Besnard, A. L., ... Evanno, G. (2015). Low reproductive isolation and highly variable levels of gene flow reveal limited progress towards speciation between European river and brook lampreys. *Journal of Evolutionary Biology*, *28*, 2248–2263.
- Rougemont, Q., Roux, C., Neuenschwander, S., Goudet, J., Launey, S., & Evanno, G. (2016). Reconstructing the demographic history of divergence between European river and brook lampreys using approximate Bayesian computations. *PeerJ*, *4*, doi: 10.7717/peerj.1910.
- Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106.
- Seehausen, O. (2009). Progressive levels of trait divergence along a 'speciation transect' in the lake Victoria cichlid fish *Pundamilia*. In R. Butlin, J. Bridle & D. Schluter (Eds.), *Speciation and patterns of diversity* (pp. 155–176). Ecological Reviews. Cambridge, UK: Cambridge University Press.
- Seehausen, O., vanAlphen, J. J. M., & Witte, F. (1997). Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science*, *277*, 1808–1811.
- Solignac, M. (1969a). Contiguous distributions sympatry and natural hybridization in a super-species *Jaera Albifrons* Leach (Isopods Asellidae). *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences Serie D*, *268*, 1610–1612.
- Solignac, M. (1969b). Hybridation introgressive dans la population complexe des *Jaera albifrons* de Luc-sur-Mer. *Archives de Zoologie Expérimentale et Générale*, *110*, 629.
- Solignac, M. (1978). *Nature, déterminisme et origine des mécanismes d'isolement dans le complexe Jaera albifrons (Isopodes, Asellotes)*. Phd Thesis. pp. 232. Université Paris 6.
- Solignac, M. (1981). Isolating mechanisms and modalities of speciation in the *Jaera albifrons* species complex (Crustacea, Isopoda). *Systematic Zoology*, *30*, 387–405.
- Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, *15*, 343–355.
- Team, R. C. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: Team, R. C.
- Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W., & Tucker, P. K. (2010). The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, *64*, 472–485.
- Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*, 3907–3930.
- Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*, 535–538.
- Veuille, M. (1976). Biogeography of the *Jaera albifrons* superspecies (Isopoda, Asellota) on the Atlantic coast of Canada. *Canadian Journal of Zoology*, *54*, 1235–1241.
- Westram, A. M., Panova, M., Galindo, J., & Butlin, R. K. (2016). Targeted resequencing reveals geographical patterns of differentiation for loci implicated in parallel evolution. *Molecular Ecology*, *25*, 3169–3186.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Ribardière A, Daguin-Thiébaud C, Houbin C, et al. Geographically distinct patterns of reproductive isolation and hybridization in two sympatric species of the *Jaera albifrons* complex (marine isopods). *Ecol Evol*. 2017;00:1–14. <https://doi.org/10.1002/ece3.3106>

Supplementary material

Figure S1: Distribution of individuals along the coast at Ste-Honorine-des-Pertes (site 8, Normandy) during a survey realized in summer 2014. Using a precise localization system (5.5 cm horizontal accuracy), we found that male individuals bearing sexual traits typical of *J. albifrons* or *J. praeheirsuta* occupied the same microhabitats and show no particular distribution patterns (both species are intermixed with no apparent distributional gradient). Females were included in this analysis, and some of them were identified as *J. albifrons* or *J. praeheirsuta* based on the sexual characters born by their sons (raised in the lab). For clarity purpose we did not distinguish males and females in this figure. Note also that many individuals are not visible here due to the near superposition of their locations.

Figure S2: Bootstrap distribution of F_{CT} (genetic differentiation between species) in Brittany (black) and Normandy (white) at loci Ja41 (Panel A) and Ja64 (B). Here we see that 1000 resampling iterations rarely produced a situation where $F_{CT-Normandy}$ was greater or equal to $F_{CT-Brittany}$ at these two loci (p -value ≤ 0.001). Observed F_{CT} are indicated by triangles (Brittany: black, Normandy: white).

Figure S3: Bootstrap distribution of F_{CT} (genetic differentiation between species) in Brittany (black) and Normandy (white) at 23 microsatellite loci. Observed F_{CT} in each region are indicated by triangles.

Figure S4: Allelic frequencies at locus Ja41 in *Jaera albifrons* (blue) and *Jaera praeheirsuta* (green) from Brittany (panel A) and Normandy (panel B). Note that the F_{CT} values reported here were calculated in hierarchical analyses of variance without the individuals showing an intermediate phenotype (see main text).

Figure S5: Allelic frequencies at locus Ja64 in *Jaera albifrons* (blue) and *Jaera praeheirsuta* (green) from Brittany (panel A) and Normandy (panel B). Note that the F_{CT} values reported here were calculated in

hierarchical analyses of variance without the individuals showing an intermediate phenotype (see main text).

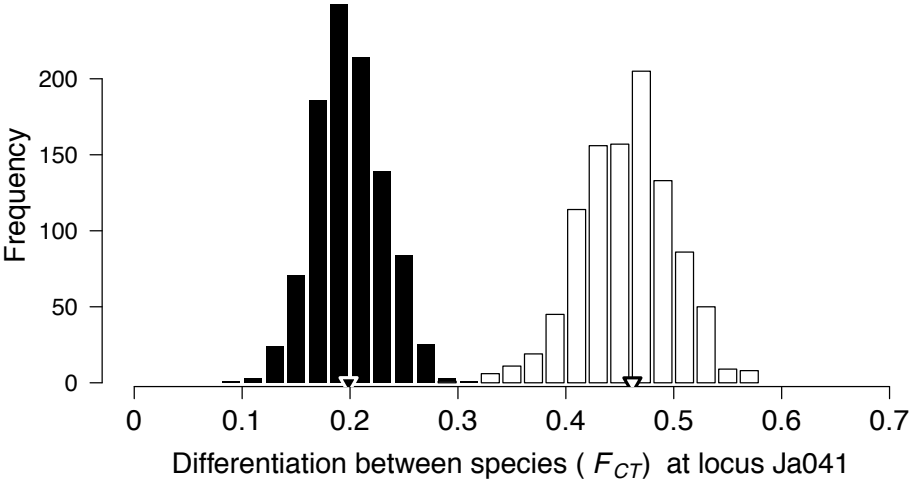
Figure S6: Patterns of isolation-by distance observed in Brittany for *Jaera albifrons* (blue circles; $y=0.0209x + 0.0011$, $R^2 = 0.73$, Mantel test p -value < 0.01) and *J. prae-hirsuta* (green circles; $y=0.0065x - 0.0133$, $R^2=0.11$, Mantel test p -value = 0.02).

Figure S1



Figure S2

A



B

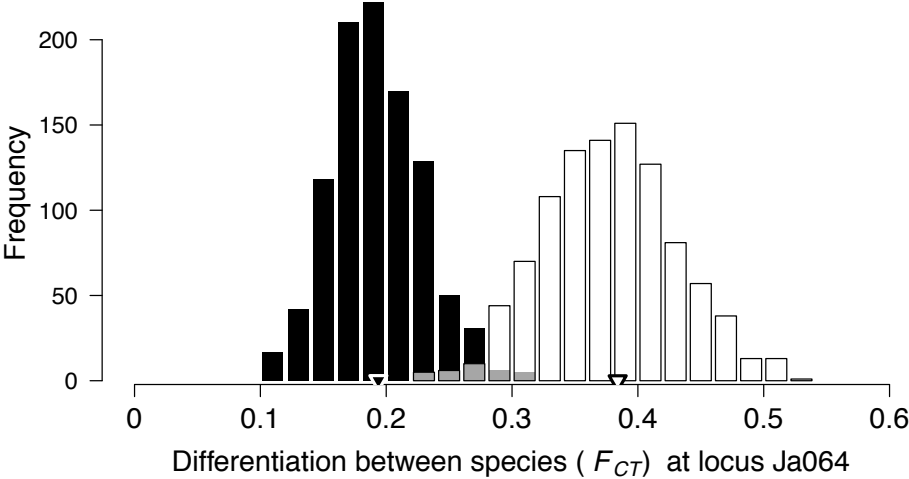


Figure S3

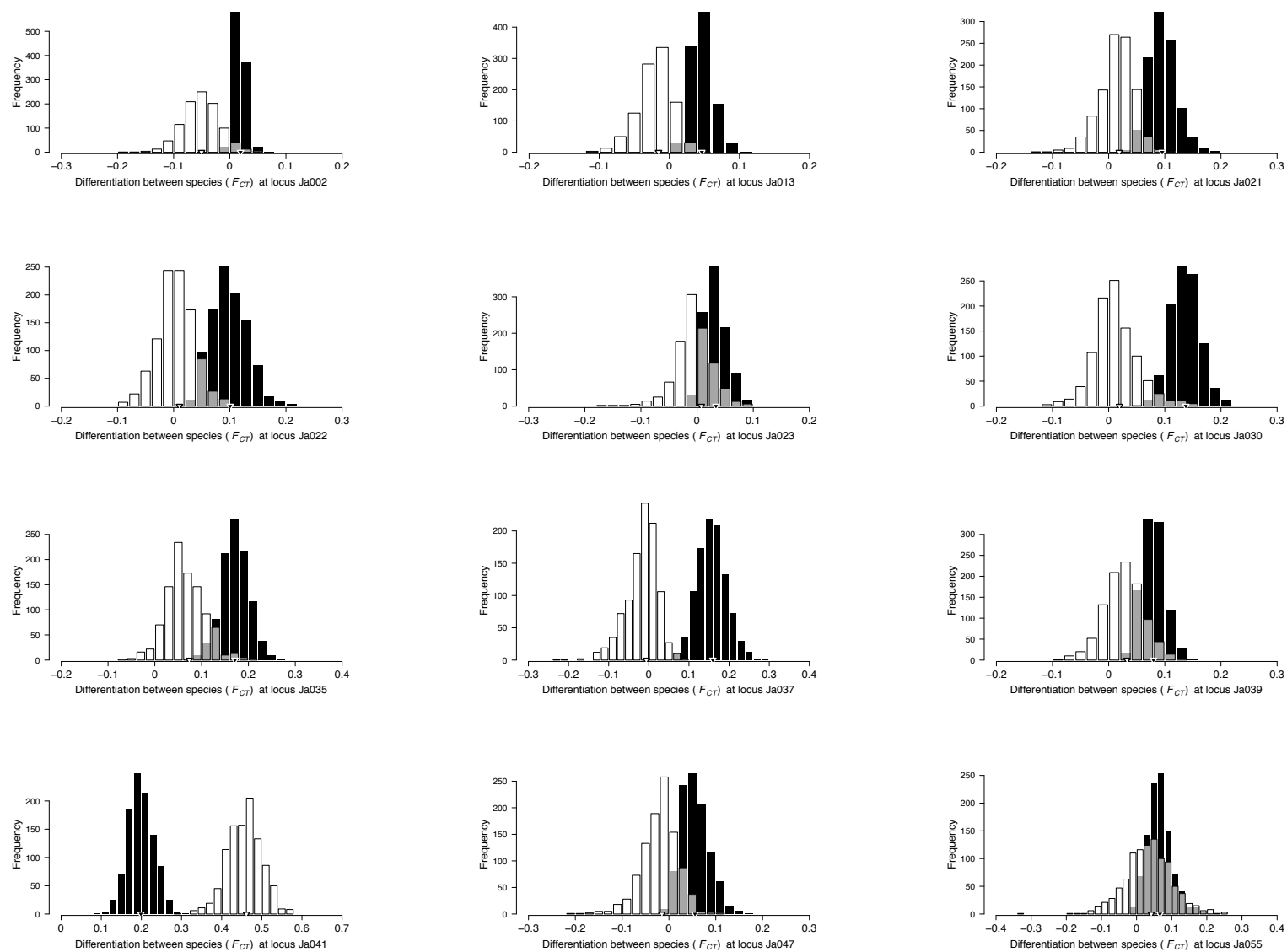


Figure S3 (continued)

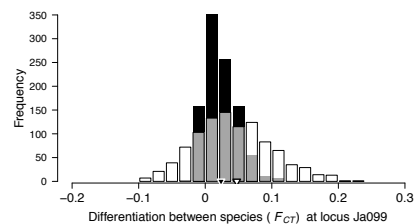
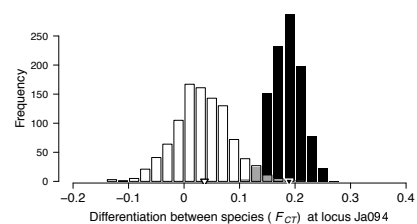
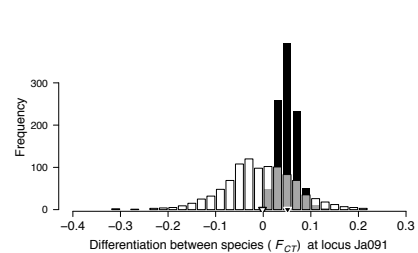
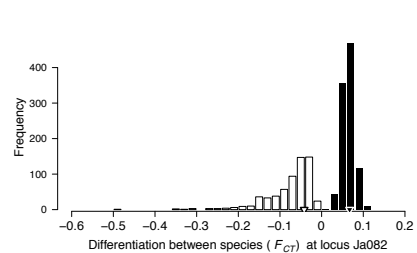
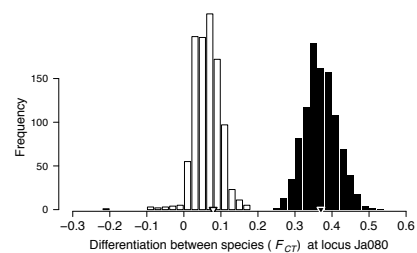
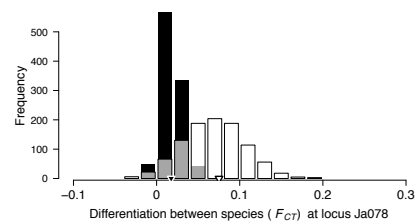
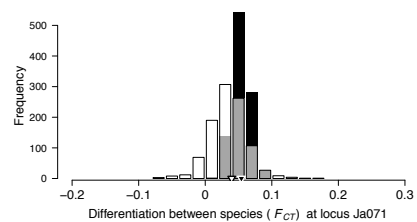
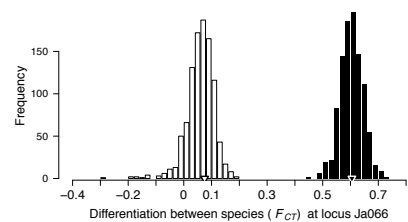
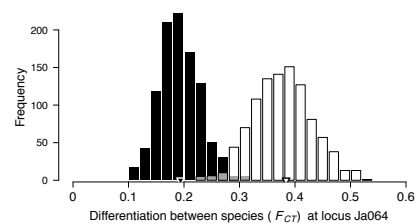
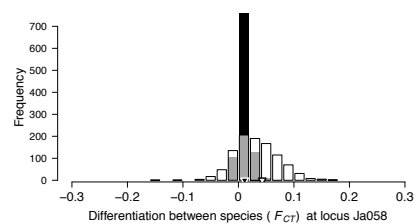
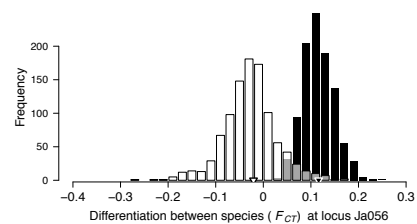


Figure S4

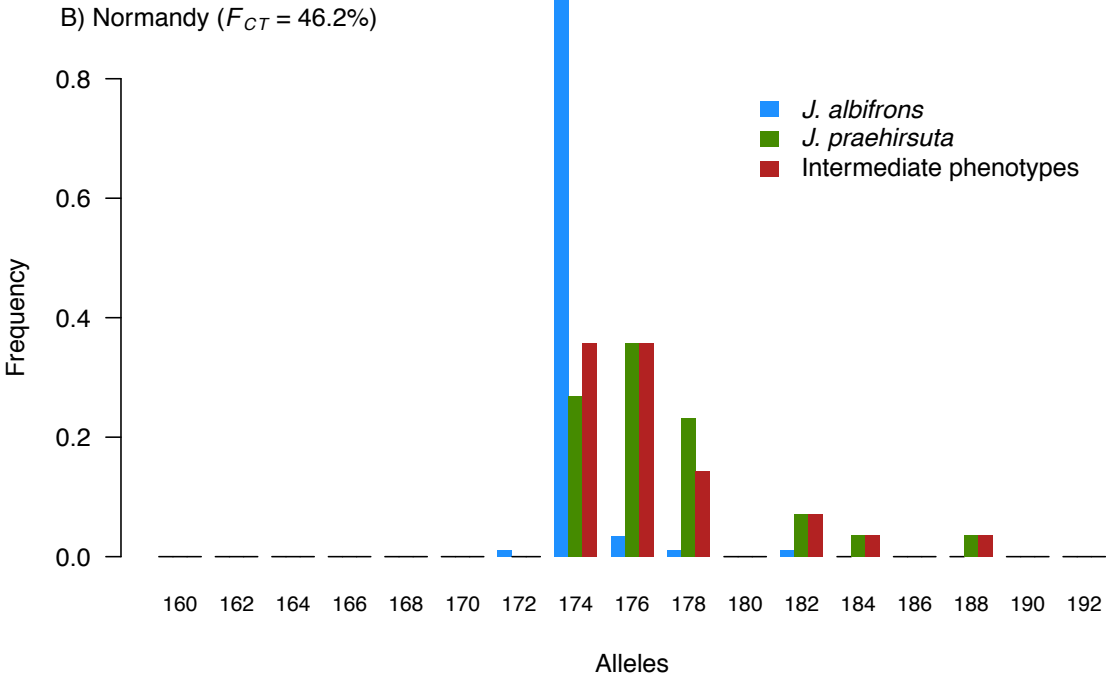
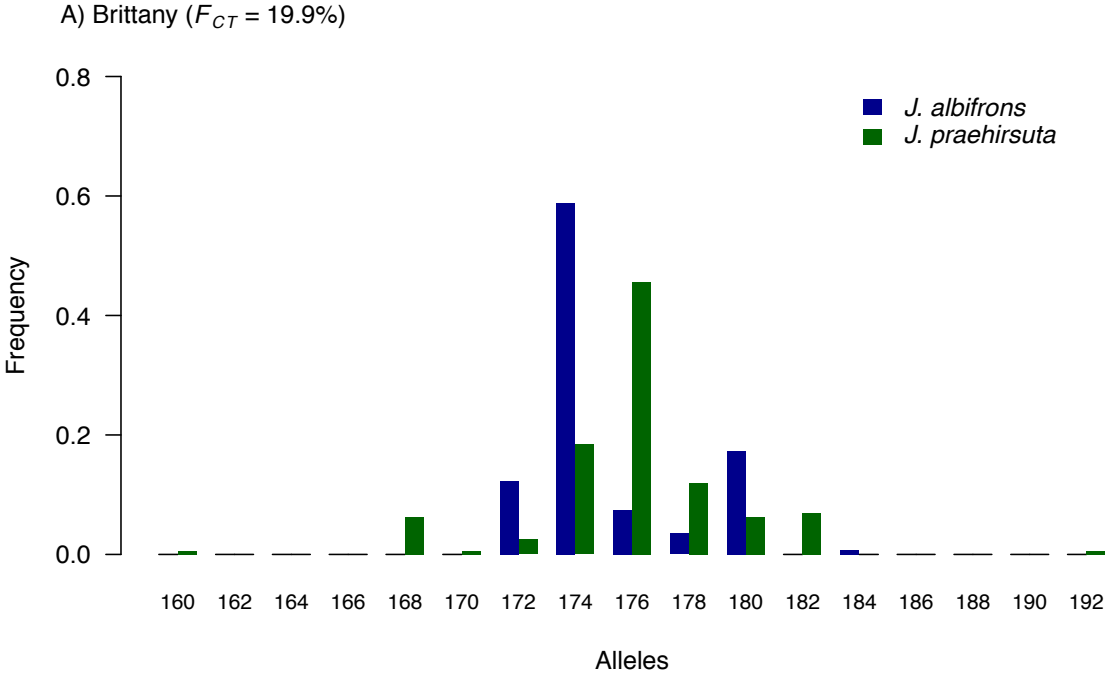


Figure S5

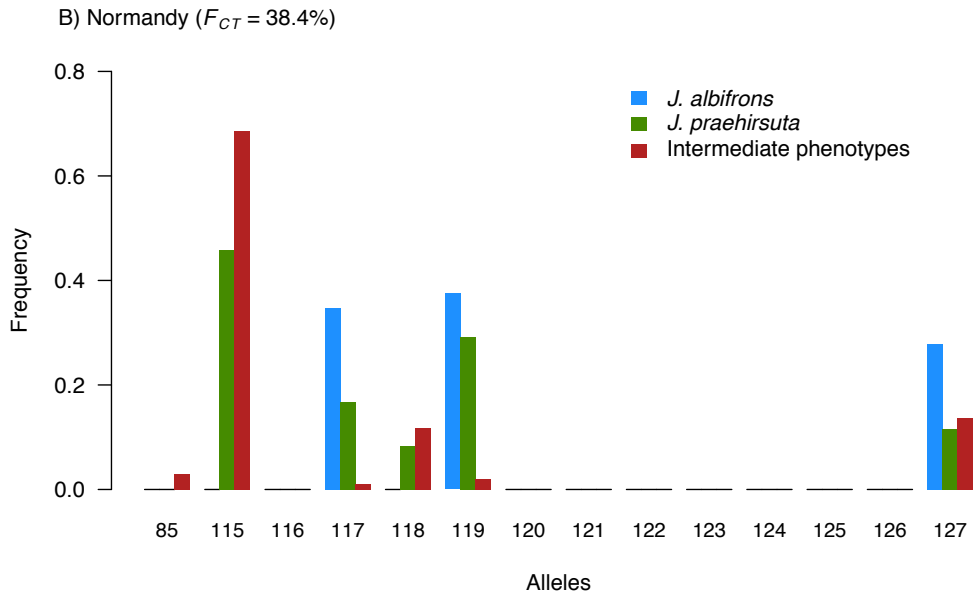
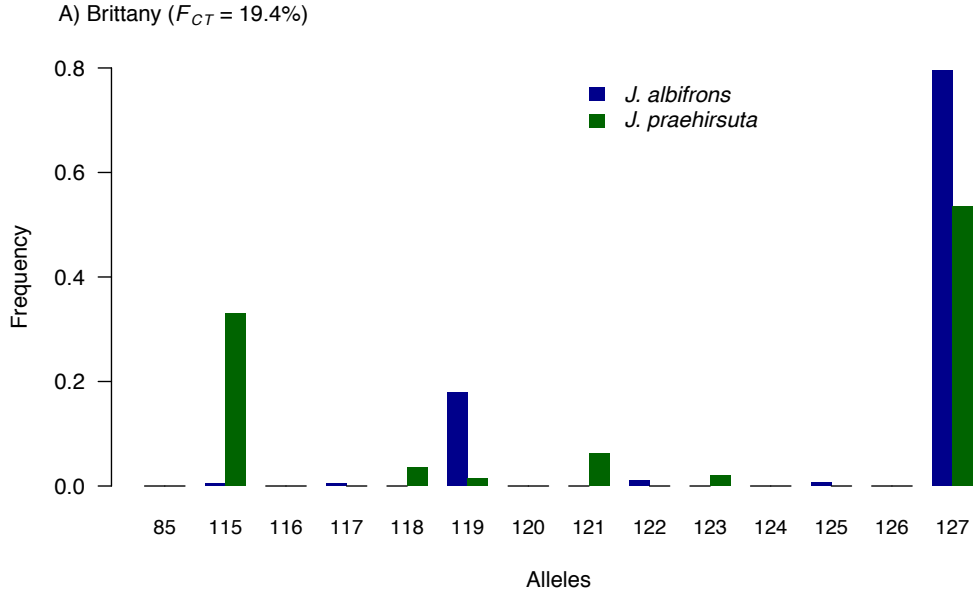
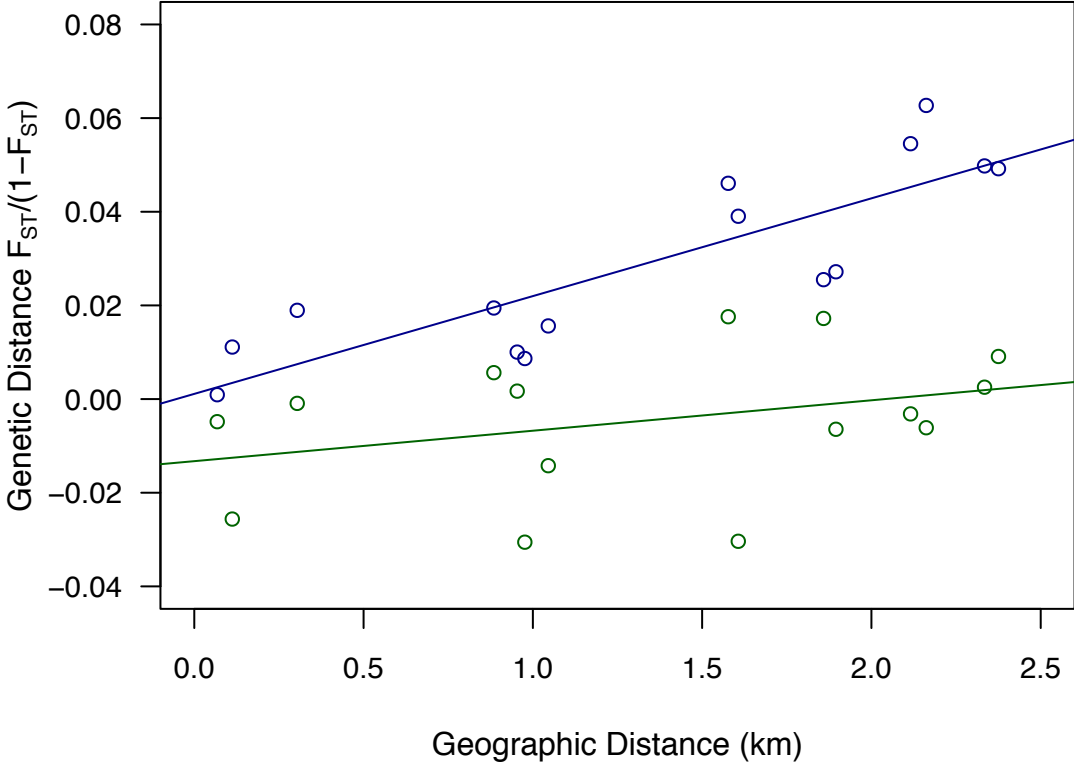


Figure S6



Chapitre 2

**STRUCTURE GENETIQUE
INTRASPECIFIQUE ET
INTERSPECIFIQUE AU SEIN DU
COMPLEXE *JAERA ALBIFRONS***

Dans le chapitre précédent, nous avons confirmé la présence de deux zones distinctes où les espèces *J. albifrons* et *J. praehirsuta* se trouvent en sympatrie, avec une faible différenciation génétique et présence d'individus hybrides dans l'une de ces zones (en Normandie), et forte différenciation génétique dans la seconde (en Bretagne).

Cette situation permettant une approche intéressante de l'isolement reproductif entre espèces, l'ensemble du travail de thèse sera largement basé sur la comparaison entre les situations Normande et Bretonne. Ce chapitre a pour objectif de préparer la suite du travail en 1) présentant la méthode de type RAD-seq qui permettra des analyses génétiques à l'échelle du génome, 2) identifiant les patterns inter-populationnels de divergences et migrations entre les espèces du complexe *Jaera albifrons* présentes le long des côtes nord-européennes, et 3) élargissant l'emprise spatiale de notre description de la structure génétique pour les espèces *J. albifrons* et *J. praehirsuta*.

1. Introduction

Dans la littérature sur la spéciation, les zones hybrides sont des cas d'étude courants, utilisées pour la compréhension de l'origine et du maintien des espèces ainsi que pour l'exploration des bases génétiques de l'isolement reproductif (par exemple, Teeter et al., 2010, Gagnaire et al., 2013, Soria-Carrasco et al., 2014). Sur le plan spatial, deux types de zones hybrides sont couramment décrites. Tout d'abord, les zones classiques, lorsque la rencontre de deux espèces parapatriques pour lesquelles l'isolement reproductif partiel (Barton & Hewitt, 1985, Endler, 1977) résulte en un gradient de fréquences allélique (cline). Le deuxième type de zone est dit en mosaïque (Harrison & Rand, 1989), car consistant en rencontres répétées dans l'espace entre deux espèces mères chacune adaptée à un habitat qui lui est propre (Bierne et al., 2003b, Butlin et al., 2014, Harrison & Larson, 2016). Les zones mosaïques se caractérisent par des inversions abruptes de fréquences alléliques dues à la distribution inégale des taxons différenciés et de leurs hybrides. Parfois certaines espèces se rencontrent à la fois dans des zones hybrides clinales et dans des zones hybrides mosaïques, comme les hippocampes *Hippocampus guttulatus* (Riquet et al., 2017). Quelle que soit leur nature, ces zones hybride impliquent toujours des cliniques de fréquences alléliques (Endler, 1977) dont le maintien est soumis à un équilibre entre différentes forces. Ainsi face à un apport d'individus parentaux avec une sélection contre les hybrides (Barton & Hewitt, 1985), la zone peut se maintenir. En revanche en fonction de l'adaptabilité et de la viabilité hybride, ces derniers pourraient être mieux adaptés à leur environnement que les individus parentaux et ainsi engendrer une spéciation hybride (Mallet, 2007). Enfin, si de nombreux loci à travers le génome montrent une migration et une taille efficace hétérogènes et que ces barrières sont asymétriques entre les deux espèces pures, et /ou que la fitness d'une des deux espèces parentales est plus faible que celle des hybrides alors un

déplacement de la zone hybride peut avoir lieu (Wielstra et al., 2017, Buggs, 2007). D'autres types de zones hybrides, non clinales, peuvent exister, tel qu'un cas de divergence allopatrique avec sous-structure génétique, suivi de plusieurs contacts secondaires, impliquant alors des interactions entre différents génotypes multilocus (Harrison & Larson, 2016). La compréhension de la structure spatiale chez les espèces étudiées est donc primordiale pour l'étude de la différenciation inter-spécifique. Dans le cas de complexe d'espèces, si une hybridation avec une autre espèce a eu lieu il peut y avoir un renforcement en cascade (Humphreys et al., 2016), pouvant augmenter l'isolement reproductif d'une des deux espèces parentales à un endroit donné. Il est donc important d'avoir une vue d'ensemble sur les différentes interactions et divergences entre espèces d'un même complexe.

L'étude génétique des zones hybrides a été révolutionnée par les méthodes de séquençage permettant une approche génomique. Le séquençage de marqueurs nucléaires associés à des sites de restriction (Rad-tag ; Baird et al., 2008) permet d'obtenir de nombreux fragments d'ADN de courtes séquences (~100–150 pb) flanquées par des sites coupés par une ou plusieurs enzymes de restriction, et fournissant ainsi plusieurs milliers de nucléotides polymorphes (SNPs) distribués à travers le génome. Cette méthode est donc très puissante pour travailler sur l'ensemble du génome chez des espèces non-modèles pour lequel aucun génome de référence n'est disponible, puissance confirmée par exemple par l'étude de Reitzel et al. (2013) qui ont montré des résultats similaires par analyses avec et sans génome de référence chez l'anémone de mer *Nematostella vectensis*.

Grâce au RAD-seq (restriction site-associated DNA), de nombreux défis dans l'étude de la structure génétique spatiale et de la spéciation peuvent maintenant être relevés. Ainsi, il a été possible de délimiter la structure de population à échelle fine chez une espèce à très grandes capacités de dispersion telle que le homard américain (Benestan et al., 2015) ou de résoudre des relations phylogénétiques dans un groupe de coraux de haute mer avec des divergences supérieures à 80 millions d'années (Herrera & Shank, 2016) ainsi que chez 16 espèces de poissons cichlidés composant une seule communauté insulaire rocheuse (Wagner et al., 2013). L'obtention d'un large nombre de SNPs est utile pour travailler sur les relations entre espèces d'un même complexe (par exemple, Dupuis et al., 2017 ou encore Manthey et al., 2017), ou analyser l'architecture génomique de l'isolement reproductif entre deux espèces (Gagnaire et al., 2013).

Dans le chapitre précédent, plusieurs sites présentant des individus *J. albifrons* et *J. prae-hirsuta* en sympatrie le long des côtes normandes ont été découverts éloignés les uns des autres de plusieurs kms, entraînant les questions de l'étalement et de la structure de la zone hybride, si l'on suppose une zone hybride limitée dans l'espace. Dans ce cas, et sur le postulat de Solignac (1981), il semblait important de rechercher i) la zone de coupure entre le contexte de Bretagne (microallopatrie sans

hybridation) et celui de Normandie (sympatrie avec hybridation introgressive), ii) l'étendue de cette zone hybride au nord de la France, et iii) d'éventuels nouveaux sites de sympatrie entre ces deux espèces. Le complexe *Jaera albifrons* regroupant 5 espèces, dont quatre fréquemment trouvées le long des côtes Nord de la France, un des objectifs de cette nouvelle recherche intensive était de récolter d'autres espèces du complexe *Jaera albifrons*. En effet, il a été démontré dans la littérature que certaines espèces du complexe *Jaera albifrons*, telles que *J. albifrons* et *J. ischiosetosa* peuvent parfois s'hybrider en milieu naturel (Solignac, 1981). Nous tenterons donc de savoir si d'éventuels événements de « migration » (flux de gènes interspécifiques) entre les différentes espèces ont pu influencer la différenciation génétique entre *J. albifrons* et *J. prae-hirsuta* dans l'une des deux régions que nous étudions. Enfin, nous analyserons la différenciation interspécifique entre les deux espèces modèles sur une plus large zone géographique que précédemment par l'analyse de nombreux marqueurs nucléaires afin d'affiner les résultats obtenus à l'aide de marqueurs microsatellites.

Les objectifs du travail présenté dans ce chapitre sont donc de i) rechercher si d'autres zones de sympatrie existent et s'il existe des limites à la zone hybride de Normandie ou s'il s'agit de plusieurs zones discrètes avec potentiel d'hybridation ii) estimer si la divergence génétique interspécifique entre ces deux espèces et les deux autres espèces du complexe *Jaera albifrons* présentes le long des côtes bretonnes et normandes est similaire ou non, et de détecter d'éventuels échanges génétiques entre celles-ci, iii) observer si la différenciation interspécifique entre *J. albifrons* et *J. prae-hirsuta* s'affaiblit le long d'un gradient géographique, de la Bretagne à la Normandie. Ces analyses descriptives de la structure génétique au sein du complexe *Jaera albifrons* seront utiles pour mieux comprendre le contexte dans lequel une analyse comparative de la paire *J. albifrons* / *J. prae-hirsuta* peut être interprétée.

2. Matériel et Méthodes

2.1 Zones de sympatrie et limites de la zone hybride

Afin d'évaluer l'étendue géographique de la zone hybride entre *J. albifrons* et *J. prae-hirsuta*, de nouvelles recherches de populations et de nouveaux échantillonnages ont été réalisés le long de la côte de Roscoff (Finistère) jusqu'à Saint-Pierre-en-Port (Seine-Maritime) en mars 2015 puis novembre 2016 (Figure 2.1) et aux îles Scilly (Royaume-Uni, juin 2015).

Le choix des sites visités s'est porté sur la présence d'habitats semblant favorables à la présence des espèces *J. albifrons* et *J. prae-hirsuta*, soit : des zones d'estran avec pierres (avec ou sans algues brunes), l'accessibilité, et la distance entre sites. Le temps moyen d'échantillonnage à un site occupé par des individus du complexe *Jaera albifrons* est d'environ 2h, et n'est possible qu'à

Tableau 2.1. Sites échantillonnés et individus prélevés. Sont donnés dans ce tableau l'ensemble des femelles échantillonnées par site (espèce non déterminable par critère morphologiques), et les mâles déterminés phénotypiquement comme appartenant à l'espèce *J. forsmanni*, *J. ischiosetosa*, *J. albifrons* ou *J. praeahirsuta* ou présentant des morphotypes intermédiaires. Les chiffres entre parenthèses représentent le nombre d'individus séquencés (RAD-seq) et analysés.

N° site	Site	Région	E	N	femelle	<i>J. forsmanni</i>	<i>J. ischiosetosa</i>	<i>J. albifrons</i>	<i>J. praeahirsuta</i>	Morph. Interm. <i>J. albifrons</i> - <i>J. praeahirsuta</i>
1	Prat	Bretagne	-4.02822	48.71217	27	7 (2)	15 (2)	5 (5)	0	0
2	Perharidi	Bretagne	-4.01209	48.7171	35	1	6	13	0	0
3	SBR_1	Bretagne	-3.989406	48.728	3	2	0	0	0	0
4	Sous_les_hotels	Bretagne	-3.98534	48.72732	8	1	0	0	0	0
5	Bloscon	Bretagne	-3.96969	48.71498	199	7	12	8	0	0
6	Pointe_Grande_Greve	Bretagne	-3.97198	48.70592	217	171 (2)	8 (1)	3	0	0
7	Pempoul	Bretagne	-3.95845	48.68753	174	4	49	49	0	0
8	Bouillennou	Bretagne	-3.953	48.67501	89	1	0	62 (10)	10 (10)	0
9	Saint-Jean	Bretagne	-3.95009	48.67246	38	0	2 (1)	26 (10)	9 (9)	0
10	Les_hauts-de-cheminees	Bretagne	-3.95205	48.65845	8	5	12	17	0	0
11	port_de_Penze	Bretagne	-3.93512	48.59804	0	0	1	0	0	0
12	Lingoz_1	Bretagne	-3.95083	48.65301	87	0	0	25 (10)	33 (10)	0
13	Lingoz_2	Bretagne	-3.95012	48.65342	153	0	1	45	64	0
14	Les_Cheminées	Bretagne	-3.94675	48.65918	69	0	0	0	72	0
15	Inizan_2	Bretagne	-3.94199	48.65931	87	0	0	22	49	0
16	Inizan_1	Bretagne	-3.94046	48.65947	70	8 (1)	0	25	14	0
17	Carantec	Bretagne	-3.90668	48.67015	111	0	0	34	0	0
18	StEfflam	Bretagne	-3.61272	48.67196	12	0	10	4	1	0
19	Tanguy	Bretagne	-3.531728	48.82330	9	1 (1)	1 (1)	2 (2)	7 (7)	0
20	Skevel	Bretagne	-3.472375	48.82868	10	2	0	0	0	0
21	Perros_Digue	Bretagne	-3.436739	48.80888	5	1	0	0	0	0
22	Port_Epine	Bretagne	-3.381654	48.81673	21	3 (2)	13 (2)	1	0	0
23	Saint_Briac	Bretagne	-2.151035	48.63265	2	0	0	1 (1)	0	0
24	Rotheneuf	Bretagne	-1.961066	48.68586	5	7	0	0	0	0
25	Port_Briac	Bretagne	-1.846318	48.69085	25	0	0	3 (3)	0	0
26	Sol_Roc	Normandie	-1.552497	48.73247	7	0	1	19	0	0
27	Granville	Normandie	-1.598987	48.83968	2	1 (1)	0	3 (3)	0	0
28	Goury	Normandie	-1.950256	49.71815	13	0	0	7 (5)	1 (1)	0
29	La Hougue	Normandie	-1.265909	49.58190	35	0	0	31 (8)	0	0
30	Grandcamp	Normandie	-1.034732	49.39035	130	0	0	41 (10)	19 (10)	2 (2)
31	Pointe_du_Hoc	Normandie	-1.01928	49.3937	100	0	37 (2)	1 (1)	0	0
32	Sainte_Honorine	Normandie	-0.79994	49.35487	207	0	0	21 (10)	70 (10)	13 (13)
33	Longues	Normandie	-0.6822	49.34829	21	0	0	15 (10)	1 (1)	0
34	Luc	Normandie	-0.34165	49.31541	5	0	0	0	0	0
35	Villerville	Normandie	0.123957	49.40290	1	0	0	0	0	0
36	Honfleur	Normandie	0.229547	49.42816	31	0	0	30 (5)	0	0
37	Yport	Normandie	0.310682	49.74008	32	0	0	161 (10)	14 (10)	11 (11)
38	Saint-Pierre-en-Port	Normandie	0.511711	49.81987	2	0	0	0	0	0
39	Iles_Scilly	UK	-6.337267	49.89415	123	0	20	67 (9)	27 (5)	1 (1)

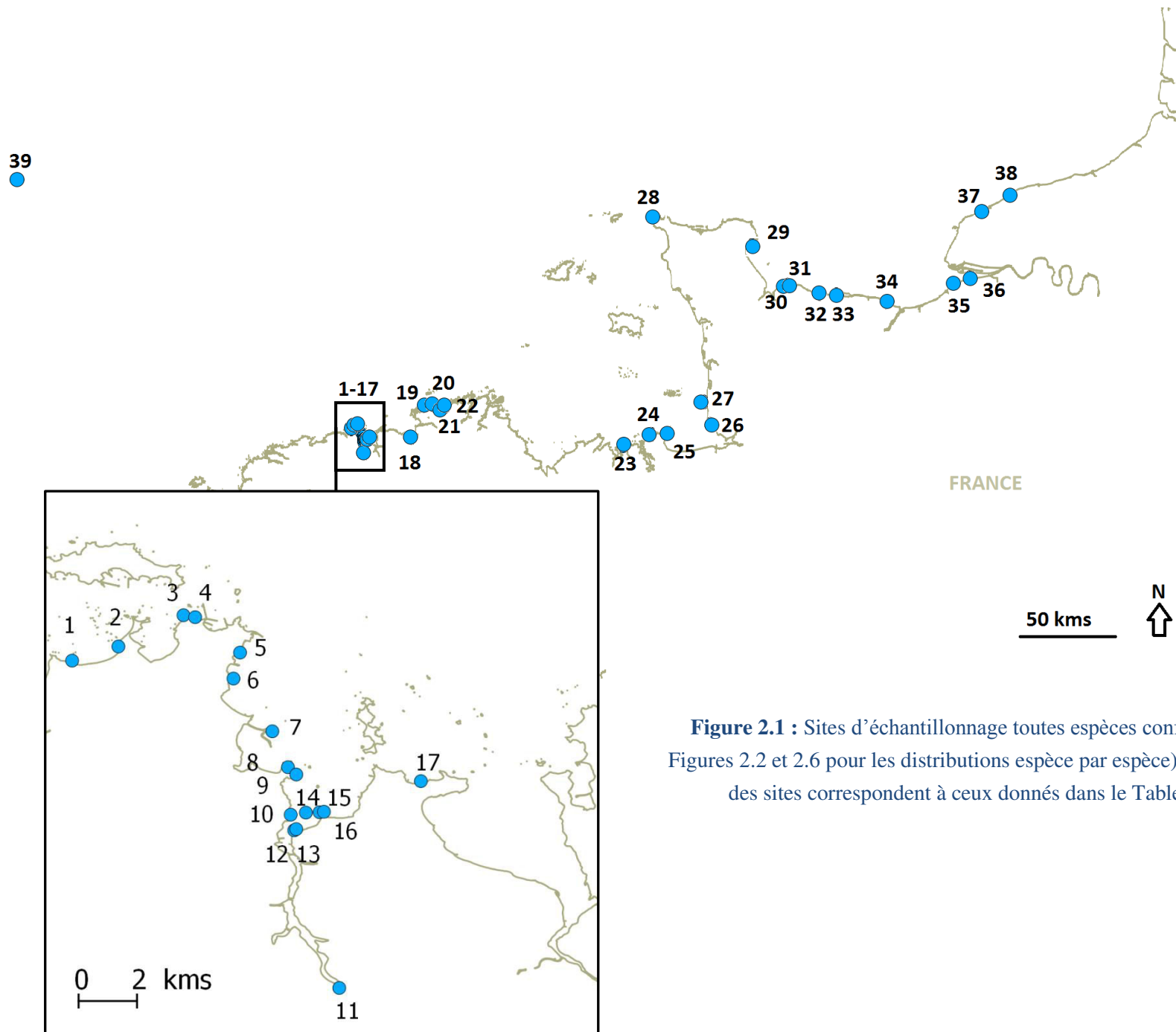


Figure 2.1 : Sites d'échantillonnage toutes espèces confondues (cf. Figures 2.2 et 2.6 pour les distributions espèce par espèce). Les numéros des sites correspondent à ceux donnés dans le Tableau 2.1.

marée basse (soit 1 à 2 site(s)/jour selon la densité des individus). Tous les sites échantillonnés entre janvier 2014 et mars 2017 ont été annotés dans le Tableau 2.1 et la Figure 2.1. Les méthodes utilisées pour les prélèvements et la détermination du sexe et de l'espèce sont celles décrites dans (Ribardière et al., 2017). Au total, 39 sites ont ainsi été échantillonnés avec un total de 2 173 femelles et 1 560 mâles prélevés (toutes espèces confondues – en effet l'identification spécifique nécessitant un travail à la loupe binoculaire au retour du terrain, les quatre espèces Européennes sont susceptibles d'avoir été échantillonnées).

2.2 Le complexe *Jaera albifrons*: Différenciation génétique entre les quatre espèces présentes sur les côtes européennes

- Echantillonnage

Au total 215 individus mâles *J. forsmanni* ont été trouvés sous les pierres dans 15 sites, presque tous en Bretagne, et 188 individus mâles *J. ischiosetosa* ont également été trouvés sous les pierres dans 15 sites, relativement mieux répartis le long de la côte que les *J. forsmanni* (Tableau 2.1). Une dizaine d'individus par espèce ont été gardés pour analyses génétiques à raison d'un à deux individus par site pris aléatoirement le long de la côte. Ainsi, 9 *J. forsmanni* provenant de 6 sites et 9 *J. ischiosetosa* provenant également de 6 sites (dont 4 en commun avec *J. forsmanni*) ont été gardés pour l'étude (Figure 2.2).

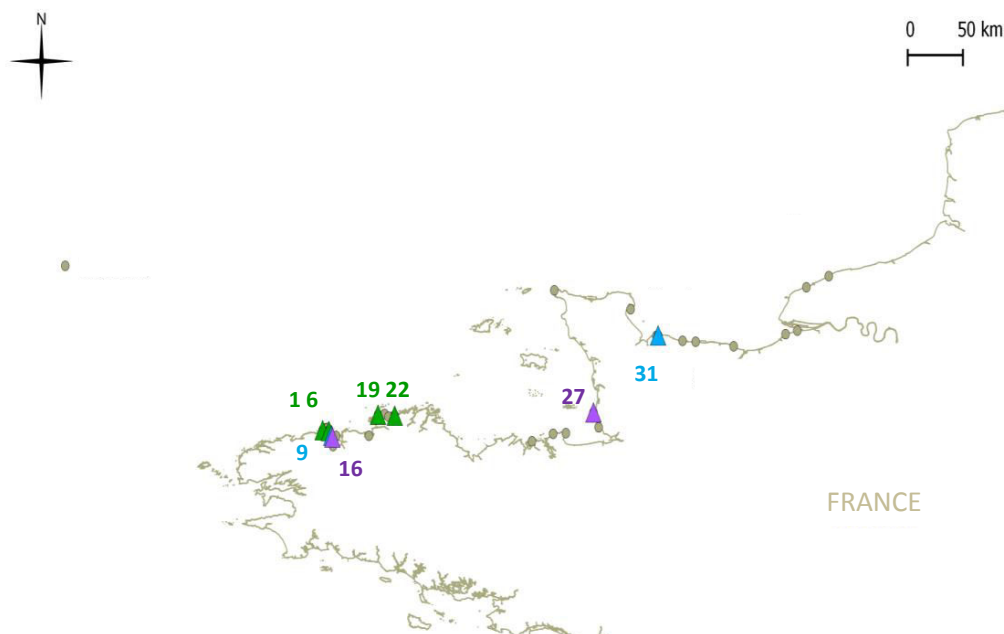


Figure 2.2 : Sites d'échantillonnage des espèces *Jaera forsmanni* et *Jaera ischiosetosa*. Les points gris représentent l'ensemble des zones d'échantillonnage, toutes espèces confondues (tel que représenté sur la Figure 2.1). Les triangles violets représentent les sites d'échantillonnage des individus *J. forsmanni*, les triangles bleus sont spécifiques des *J. ischiosetosa*, tandis que les 4 sites mixtes sont représentés par des triangles verts. Le numéro des sites correspond à celui donné dans le Tableau 2.1.

- Constructions des banques RAD et analyses bio-informatiques

Afin de comparer la différenciation sur l'ensemble des quatre espèces présentes le long des côtes européennes, une analyse de nucléotides polymorphes (SNPs, Single-Nucleotide Polymorphism) a été réalisée. Les SNPs ont été obtenus par séquençage de marqueurs nucléaires associés à des sites de restriction par double digestion (ddRAD), méthode faisant intervenir 2 enzymes de restriction avec sélection de taille précise permettant d'obtenir des fragments d'ADN (RAD-tags) répartis tout au long du génome (Peterson et al., 2012, méthode détaillée en annexe 1).

Pour cette étude, les femelles du complexe *Jaera albifrons* ne pouvant être discriminées sur la base de leur morphologie, seuls 241 mâles ont été séquencés et utilisés pour les analyses suivantes, dont 112 *Jaera albifrons*, 73 *Jaera praehirsuta*, 27 morphotypes intermédiaires à ces deux espèces, 9 *Jaera forsmanni* et 9 *Jaera ischiosetosa* (voir Tableau 2.1, chiffres entre parenthèses). 12 individus, dont 9 *J. albifrons*, 1 *J. praehirsuta* et 2 morphotypes intermédiaires ont été répliqués 1 fois afin de vérifier les erreurs possibles de séquençage. 10 témoins négatifs ont également été ajoutés, dont 4 témoins d'extraction, 4 témoins de construction de banques et 2 témoins de séquençage. Après réception des séquences, un protocole bio-informatique a été suivi (utilisation du programme *Stacks*) pour nettoyer et analyser ces séquences (Figure 2.3, voir aussi annexe 2).

Lors de l'utilisation du programme *Stacks* (Catchen et al., 2013), un *denovo_map* a été réalisé en l'absence de génome de référence, avec comme paramètres : m - un nombre minimal de 5 séquences (reads) identiques pour créer une pile (stacks) ; M - une distance maximale de 5 nucléotides entre deux stacks pour pouvoir les assembler en un même locus ; N - jusqu'à 7 nucléotides de différences pour ajouter les séquences secondaires (pas assez nombreuses pour créer une pile lors de l'étape précédente) aux stacks déjà créés ; n - un nombre maximum de 5 nucléotides différents entre individus pour considérer que deux stacks sont homologues lors de la construction du catalogue.

Afin de retirer les séquences qui sont peu couvertes, et/ou qui peuvent être dues à des erreurs de séquençages, une limite de log-likelihood de -12 a été appliquée via le programme *rxstacks*, un taux d'erreur maximal de 0.05 a été autorisé ainsi qu'un maximum de 75 % des individus ayant de multiple loci à séquences répétées.

Sur l'ensemble des banques et des séquences, le nombre moyen de reads par individu après séquençage était de 201 921 593. Le programme *populations* a été lancé sur l'ensemble des échantillons, dans un premier temps sans filtres, et 768 881 SNPs ont été obtenus, répartis sur 176 046 tags (groupes de SNPs).

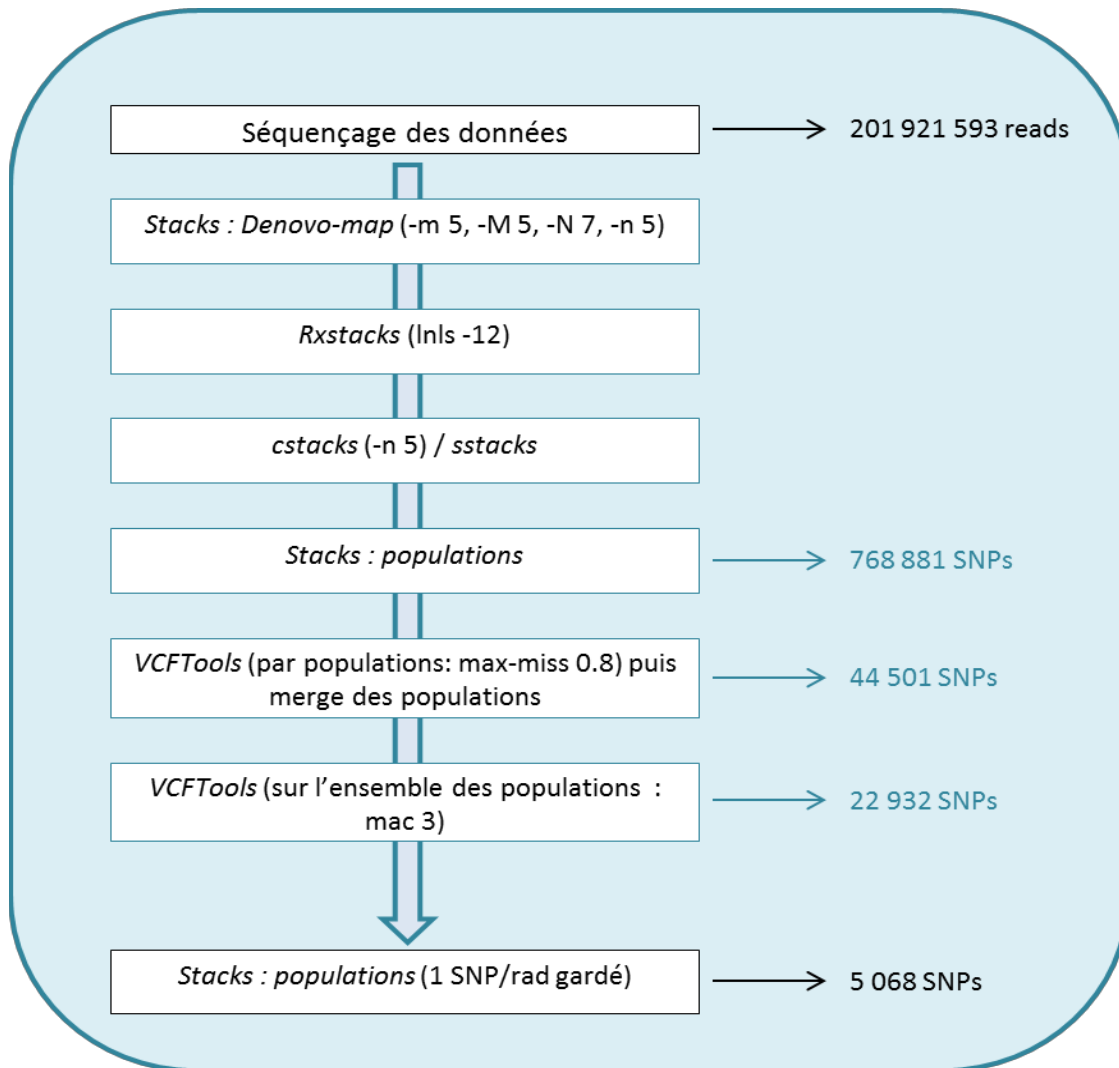


Figure 2.3 : Analyses bio-informatiques effectuées et nombre de SNPs obtenus à différentes étapes.

Deux morphotypes intermédiaires, provenant de Sainte-Honorine, et ayant plus de 95% de données manquantes sur l'ensemble des loci ont été retirés de l'analyse. Les témoins négatifs (>99% de données manquantes sur l'ensemble des loci) ont également été retirés de l'analyse. Les réplicas montraient en moyenne 96.5% de similarité sur l'ensemble des loci [93.4-98.4] et sur 2 réplicas effectués par individu, seul celui ayant le plus de données était gardé.

Via le programme *Vcftools* (Danecek et al., 2011) tous les individus ont été groupés selon leur population d'échantillonnage (cf. Tableau 2.2), puis 20% de données manquantes maximum ont été autorisées, *i.e.* chaque locus n'est retenu que s'il est génotypé chez au moins 80% des individus de la population. Tous les individus ont ensuite été réunis ensemble en ne gardant que les SNPs en commun (environ 44 000 loci à ce stade), et les loci dont un des allèles présentait moins de trois copies sur l'ensemble des individus ont été retirés du jeu de données, diminuant le jeu de données de moitié, pour obtenir 22 923 loci. Les loci restant ont été mis dans une liste à garder et *populations*

a de nouveau été lancé, après paramétrage pour ne garder qu'un SNP aléatoire par read parmi les loci de cette liste. Le jeu de données final comporte 5 068 SNPs, à raison d'un seul SNP par tag gardé aléatoirement. La profondeur de couverture moyenne de l'ensemble des loci sur l'ensemble des individus est de 44 (Figure 2.4) et la profondeur de couverture moyenne sur l'ensemble des individus de chaque population varie de 41 à 45 (Figure 2.5).

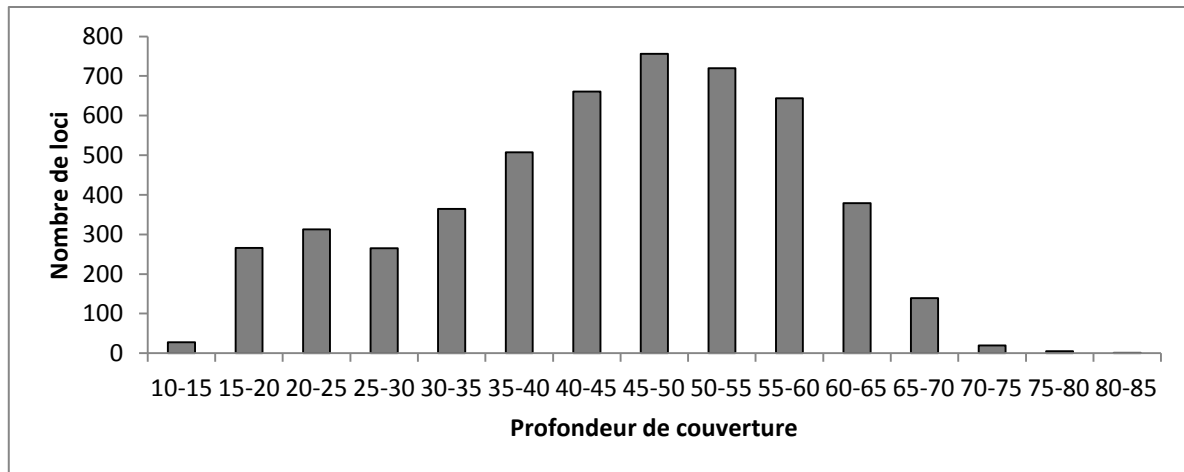


Figure 2.4 : Distribution de la profondeur de couverture moyennée sur l'ensemble des individus

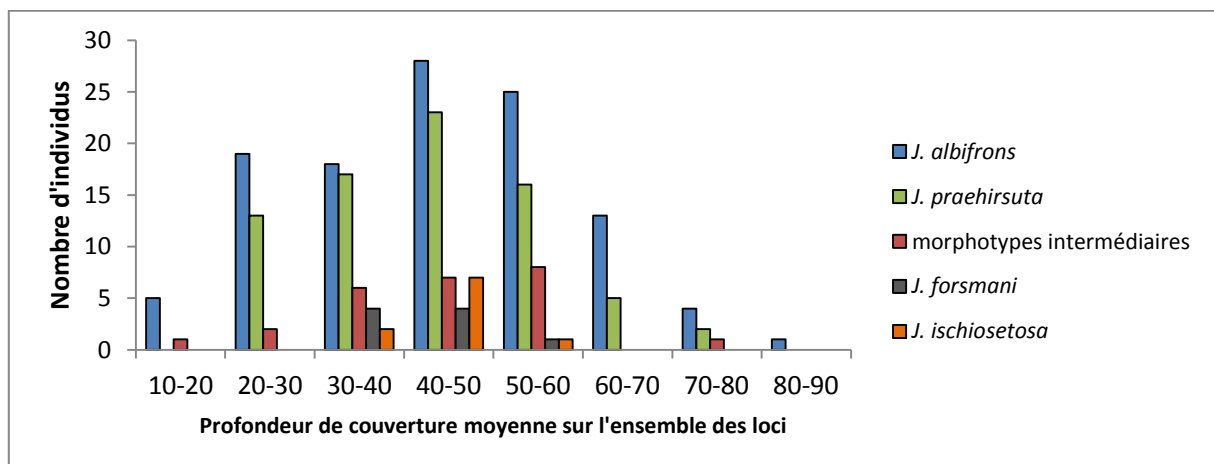


Figure 2.5 : Distribution de la profondeur de couverture moyenne obtenue sur l'ensemble des loci pour chaque individu

La différenciation interspécifique au sein des quatre espèces européennes du complexe *Jaera albifrons* a été observée sur une Analyse en Composantes Principales (ACP), réalisée à l'aide du programme adegenet dans R (Jombart, 2008, Core Team, 2016) sur le jeu de données final. Lors de cette analyse, le nombre de données manquantes a été vérifié, et s'est avéré faible comme attendu. Par la suite, des inférences d'évènements de divergence et de migration (flux de gènes interspécifiques) inter-populationnelles ont été réalisées à l'aide du programme Treemix (Pickrell &

Pritchard, 2012), en utilisant *J. ischiosetosa* pour enraciner l'arbre (sur la base des résultats d'ACP, voir ci-dessous), en générant 1000 bootstraps, et en appliquant une contre-correction afin de ne pas avoir de sur-correction due à certaines petites tailles de populations (conformément au manuel d'utilisation de *Treemix*). Pour cette analyse, les sites de Bouillennou, StJean, Lingo, Prat et Tanguy ont été réunis sous une population « Bretagne », les sites Port-Briac, Saint-Briac, Goury et Granville sous une population « Inter-région » les sites Grandcamp, Sainte-Honorine, Honfleur, Longues et Pointe du Hoc sous une population « Normandie », les sites Yport, La Hougue et Scilly sous leur propre nom, et enfin les populations de *J. forsmanni* ont été réunies ensemble, ainsi que celles de *J. ischiosetosa* (Tableau 2.2).

2.3 Différenciation génétique entre *Jaera albifrons* et *Jaera praeheirsuta*

Pour cette partie, les individus *J. albifrons*, *J. praeheirsuta* et les morphotypes intermédiaires ont été sous-échantillonnés du jeu de données final. Une ACP a de nouveau été réalisée en ne gardant que ces individus (populations utilisées décrites dans le Tableau 2).

Tableau 2.2. Sites d'échantillonnage des populations utilisées dans cette étude, et noms associés pour l'ACP et l'arbre des événements démographiques.

Site_ID	Site	E	N	populations Treemix	populations ACP
1	Prat	-4.02822	48.71217	Bretagne	bret_tot
8	Bouillennou	-3.953	48.67501	Bretagne	bret_tot
9	Saint-Jean	-3.95009	48.67246	Bretagne	bret_tot
12	Lingo_1	-3.95083	48.65301	Bretagne	bret_tot
19	Tanguy	-3.53173	48.82331	Bretagne	bret_tot
23	Saint_Briac	-2.15104	48.63265	Inter-région	norm_ouest
25	Port_Briac	-1.84632	48.69086	Inter-région	norm_ouest
27	Granville	-1.59899	48.83969	Inter-région	norm_ouest
28	Goury	-1.95026	49.71815	Inter-région	Goury
29	La Hougue	-1.26591	49.58191	Inter-région	norm_centre
30	Grandcamp	-1.03473	49.39035	Normandie	norm_centre
31	Pointe_du_Hoc	-1.01928	49.3937	Normandie	norm_centre
32	Sainte_Honorine	-0.79994	49.35487	Normandie	norm_centre
33	Longues	-0.6822	49.34829	Normandie	norm_centre
36	Honfleur	0.229547	49.42816	Normandie	norm_est
37	Yport	0.310682	49.74008	Yport	norm_est
39	Iles_Scilly	-6.33727	49.89415	Scilly	Scilly

3. Résultats

3.1 Zones de sympatrie et limites de la zone hybride

Nous avons trouvé 10 sites présentant les espèces *J. albifrons* et *J. praeheirsuta* en sympatrie, dont 4 avec des habitats différenciés (sous les pierres pour les *J. albifrons* et sur des algues pour les *J. praeheirsuta*) et 6 avec un même habitat partagé (sous les pierres) (Figure 2.6). Parmi les 6 sites où les espèces partagent un même habitat, la présence de morphotypes intermédiaires a été avérée (n=27) pour quatre d'entre eux. Deux de ces sites, obtenus lors du nouvel échantillonnage (brièvement mentionnés dans le chapitre 1), semblent particulièrement intéressants. Le premier, situé sur l'île St-Agnes dans l'archipel des Scilly (site numéro 39) est un site où les deux espèces sont en syntopie dans lequel un individu morphologiquement intermédiaire a été trouvé. Ce site est situé très loin au nord-ouest des populations Normandes (dans lesquelles un processus d'hybridation introgressive a été observé, voir chapitre 1). Le second, Yport (site numéro 37), est le site où les deux espèces sont en syntopie le plus à l'est des côtes françaises que nous ayons trouvées lors de nos échantillonnages, et le nombre de mâles morphologiquement intermédiaires détectés est de 5 %.

A l'autre extrémité de la zone d'hybridation a été échantillonné le site de Goury (numéro 28), intéressant car les deux espèces ont été trouvées sous les pierres mais aucun morphotype intermédiaire n'a été détecté parmi les individus mâles échantillonnés, soit 7 *J. albifrons* et 1 *J. praeheirsuta*. L'analyse génétique des individus trouvés à ce site devrait donc permettre de définir s'il présente une forte différenciation interspécifique telle que trouvée aux sites bretons, auquel cas la limite de la zone hybride, *i.e.* de la région géographique (Normandie) où plusieurs sites avec hybridation ont été décrits, serait à l'est de ce site ou bien si Goury présente très peu de différenciation interspécifique, auquel cas la limite pourrait se situer plus à l'ouest. Enfin, le nouvel échantillonnage n'a pas permis de détecter d'individu *J. praeheirsuta* de l'ouest de Goury jusqu'en Bretagne (soit entre les sites numéros 19 et 28), ou entre le site de Longues (33) et celui d'Yport (37).

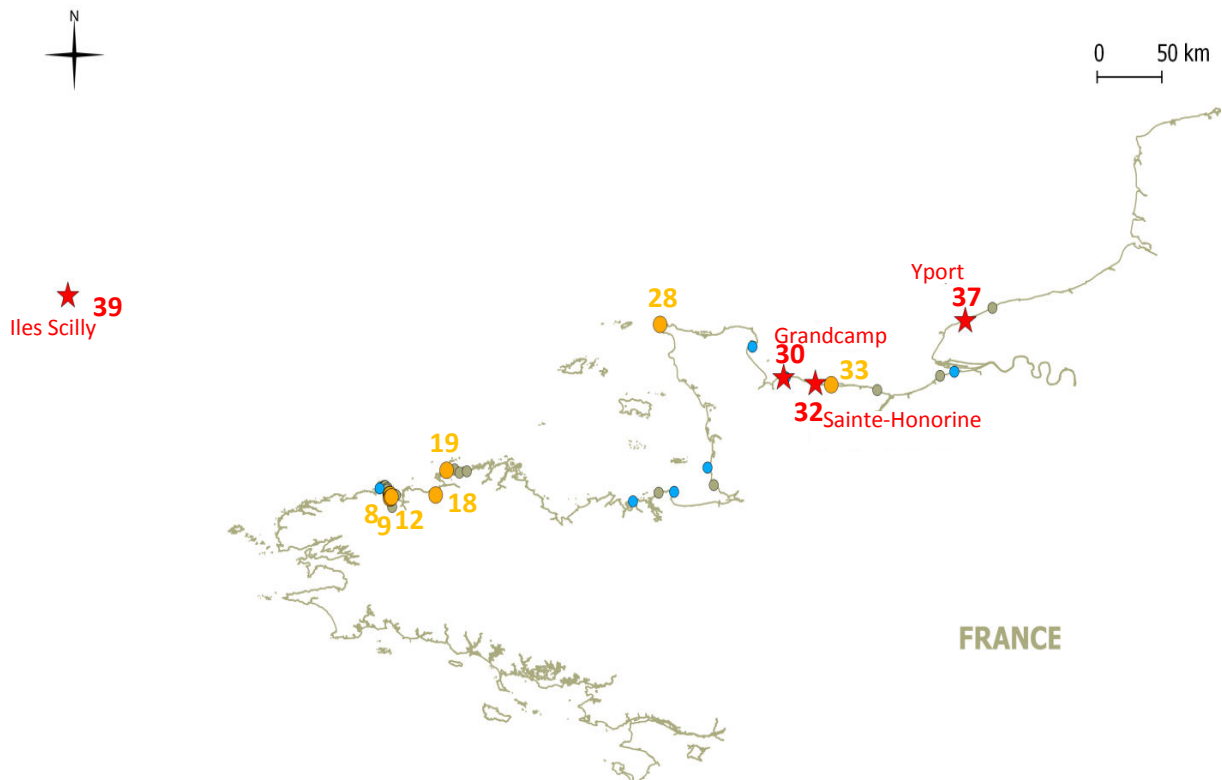


Figure 2.6 : Sites d'échantillonnage des espèces *Jaera albifrons* et *Jaera praeahirsuta*. Les points gris représentent l'ensemble des zones d'échantillonnage, toutes espèces confondues, les points bleus représentent la distribution de l'espèce *J. albifrons*, les sites où les deux espèces *J. albifrons* et *J. praeahirsuta* sont présentes sont représentés en orange, et enfin, les sites sympatriques où des morphotypes intermédiaires ont été détectés sont représentés par des étoiles rouges. Les triangles violets représentent les sites d'échantillonnage des individus *J. forsmanni*, les triangles bleus sont spécifiques des *J. ischiosetosa*, tandis que les 4 sites mixtes sont représentés par des triangles verts.

3.2 Le complexe *Jaera albifrons* : Différenciation génétique entre espèces des côtes européennes

- ACP

A partir du jeu de données final, une ACP a pu être réalisée avec toutes les espèces (5 068 loci en commun) et la forte divergence de *J. ischiosetosa* par rapport aux autres espèces est détectable sur cette ACP (Figure 2.7). Le long de l'axe deux, on détecte depuis le bas de la figure, un groupe de *J. albifrons* et *J. forsmanni*, puis un groupe de *J. praeahirsuta* provenant de sites situés proches des sites normands connus pour l'hybridation entre *J. albifrons* et *J. praeahirsuta* (voir sites 29 à 37 du Tableau 2.1), des morphotypes intermédiaires puis un groupe de *J. praeahirsuta* provenant de sites proche des sites bretons où aucune hybridation n'avait été détectée précédemment (voir sites 1 à 28 et site 39 du Tableau 2.1).

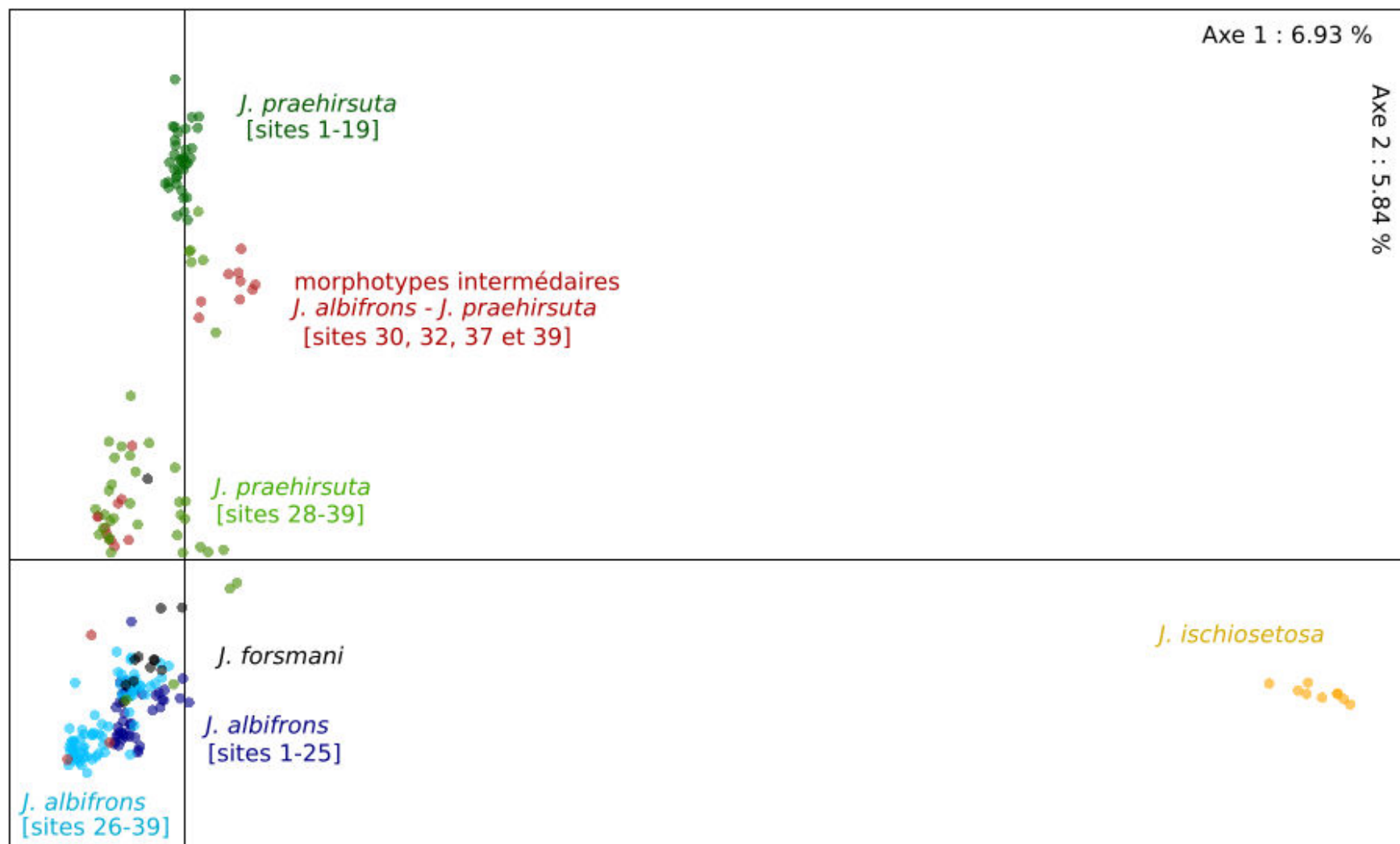


Figure 2.7 : ACP obtenues sur l'ensemble des individus, toutes espèces confondues. Les *Jaera albifrons* provenant de sites bretons (numéros 1 à 25) sont indiqués en bleu foncé, tandis que les individus provenant de sites normands et des îles Scilly (numéros 26 à 39) sont présentés en bleu clair. Les *J. praehirsuta* provenant de sites bretons (numéros 1 à 19) sont indiqués en vert foncé, tandis que les individus provenant de sites normands et des îles Scilly (numéros 28 à 39) sont présentés en vert clair. Les morphotypes intermédiaires *J. albifrons* - *J. praehirsuta* obtenus sur les sites 30,32, 37 et 39 sont en rouge, tous les individus *J. ischioetosa* sont en jaune et tous les individus *J. forsmani* sont en noir.

- Arbre de relations inter-populationnelles

Un arbre des relations entre espèces Européennes du complexe (Figure 2.8) a été réalisé après l'ACP, de telle sorte que les *J. ischiosetosa* ont pu être utilisés comme racine. L'arbre obtenu regroupe les individus appartenant à une même espèce entre eux, en accord avec l'identification morphologique des espèces. Deux évènements de flux de gènes inter-spécifiques apparaissent dans cet arbre, à chaque fois de *J. albifrons* vers *J. praeheirsuta*, à Yport, et *a priori* dans une moindre mesure sur les autres sites de Normandie (population définie ici comme le regroupement des sites Grandcamp, Pointe-du-Hoc, Sainte-Honorine, Longues, Honfleur, soit les sites numéros 30, 31, 32, 33 et 36).

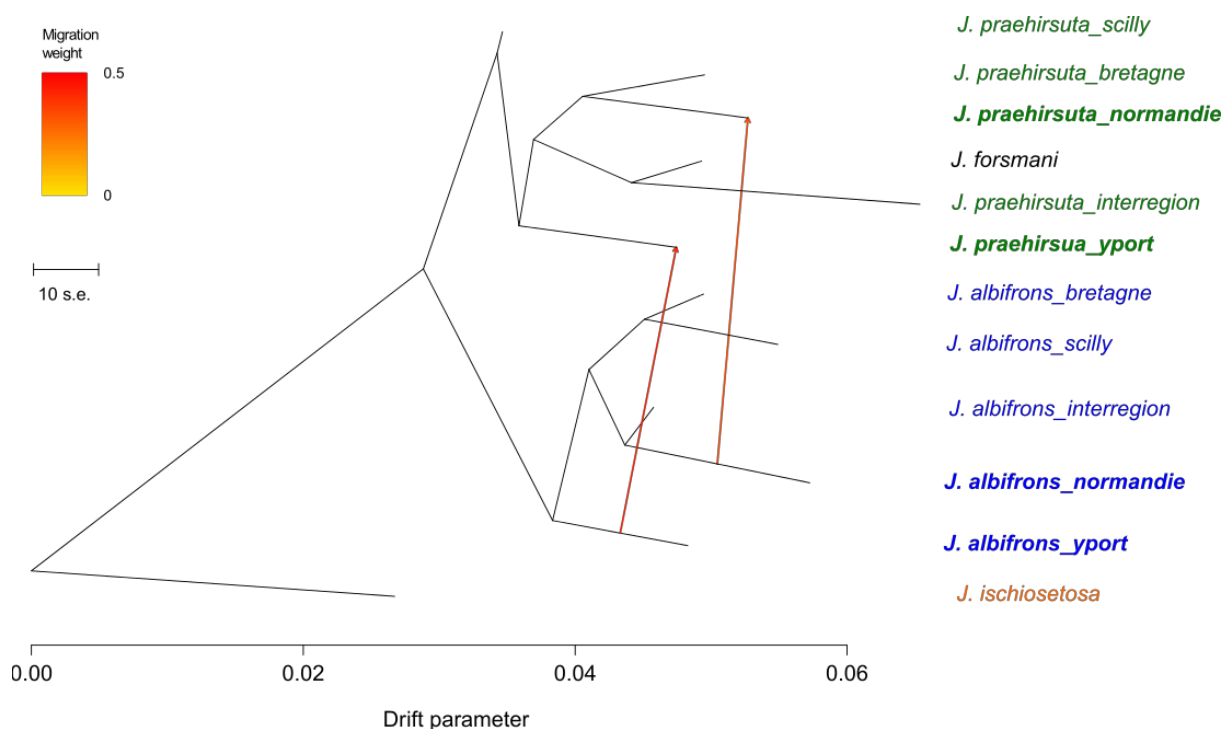


Figure 2.8 : Arbre de relations entre les 4 espèces du complexe *Jaera albifrons* des côtes européennes ; *J. praeheirsuta* (en vert), *J. forsmani* (en noir), *J. albifrons* (en bleu) et *J. ischiosetosa* (en orange, utilisé pour enraciner l'arbre). Les espèces en gras représentent celles impliquées dans les deux évènements de migrations.

3.3 Différenciation génétique entre *Jaera albifrons* et *Jaera praeheirsuta*

Une nouvelle ACP a été réalisée avec les individus *J. albifrons*, *J. praeheirsuta* et les morphotypes intermédiaires, *i.e.* sans les individus *J. forsmani* et *J. ischiosetosa* (Figure 2.9). L'axe 1 semble déterminé par la différenciation interspécifique et l'axe 2 par la géographie et/ou par la différenciation intraspécifique. En Bretagne (sites 1-22), les *J. praeheirsuta* (triangles verts foncés) et les *J. albifrons* (carré bleus foncés) sont très différenciés et aucun individu n'est génétiquement intermédiaire. En Normandie ces deux espèces sont génétiquement plus proches, et les *J. praeheirsuta* ont l'air d'être génétiquement « tirés » par les *J. albifrons* : les points sont plus proches du centre (sans pour autant avoir plus de données manquantes que les autres individus) ce qui est d'autant plus vrai pour les individus d'Yport (site 37, en violet). Lorsque l'on s'intéresse aux morphotypes intermédiaires de Normandie, on peut remarquer un pattern inversé : les morphotypes intermédiaires provenant de Yport (site 37, en violet) sont génétiquement plus proches des *J. albifrons*, alors que ceux provenant de Sainte-Honorine et Grandcamp (sites 32 et 30, respectivement en jaune et marron) sont plus proches des *J. praeheirsuta*. Le seul individu *J. praeheirsuta* trouvé à Longues (site 33, vert clair) montre une similarité génétique avec des individus de Sainte-Honorine. Concernant le site Goury (site 28, en rose), site sympatrique géographiquement situé entre nos sites bretons et normands, les individus *J. albifrons* et *J. praeheirsuta* montrent tous une différenciation intra-spécifique intermédiaire entre les individus bretons et normands. Cela est d'autant plus visible chez les *J. albifrons* pour lesquels nous avons eu plus de succès lors de l'échantillonnage. Ainsi nous pouvons observer de bas en haut (correspondant à un axe géographique d'ouest en est), les sites bretons Prat, Bouillenou, Saint-Jean, Lingoz et Tanguy en bleu foncé (sites 1-22), puis les sites de Saint-Briac, Port-Briac et Granville en bleu clair (sites 23-27) ainsi que Goury en rose (site 28) suivi par le site de La Hougue (site 29, groupe de turquoise isolé) également situé sur la pointe du Cotentin, et enfin les sites normands de Grandcamp, Pointe du Hoc, Sainte Honorine, Longue, Honfleur et jusqu'à Yport (sites 30-37) en mélange marron, turquoise, jaune, vert clair, violet. Enfin, les îles Scilly (en rouge) montrent un pattern de différenciation interspécifique semblable aux sites bretons, ainsi qu'un morphotype intermédiaire situé sur l'ACP au milieu des deux groupes d'espèces distincts.

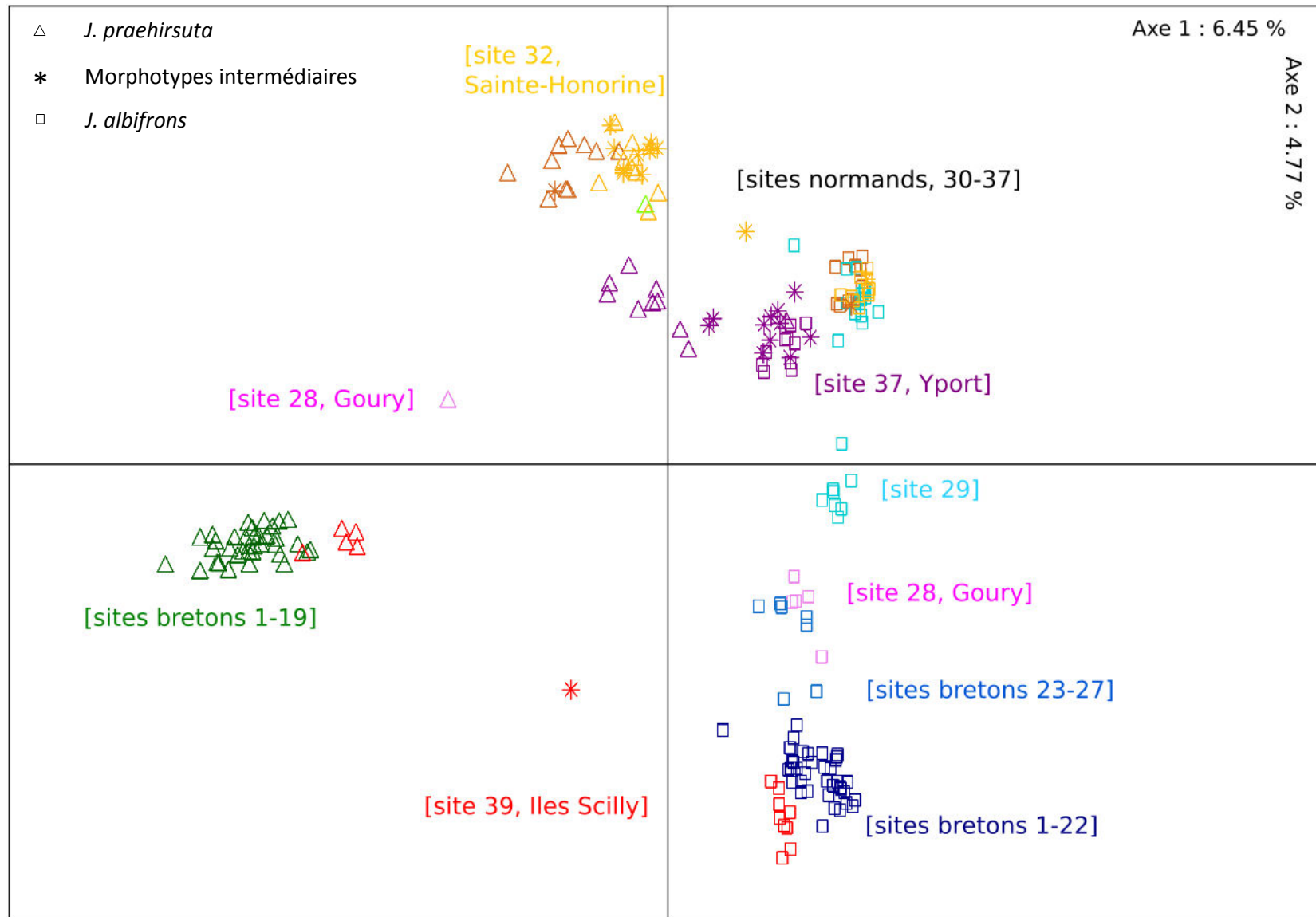


Figure 2.9 : ACP réalisée sur l'ensemble des individus *J. albifrons* (représentés par des carrés), *J. praeheirsuta* (représentés par des triangles) et morphotypes intermédiaires (représentés par des étoiles). Les sites avec hybridation sont en rouge pour les îles Scilly, en violet pour Yport, en marron pour Grandcamp et en jaune pour Sainte-Honorine. Le site de Goury est représenté lui en rose. Les autres sites sont représentés par un gradient de bleu ou de vert, qui suit un axe géographique d'ouest en est. Ainsi les sites bretons (Prat, Bouillenou, Saint-Jean, Lingoz et Tanguy) sont représentés en vert foncé et bleu foncé ; Saint-Briac, Port-Briac et Granville sont représentés en bleu clair ; La Hougue, Pointe du Hoc et Honfleur sont représentés en turquoise, et le site de Longues est représenté en turquoise pour les *J. albifrons* et en vert clair pour *J. praeheirsuta*.

4. Discussion

Le travail présenté dans ce chapitre avait pour but de mieux appréhender le contexte spatial et génétique des relations entre *Jaera albifrons* et *J. prae-hirsuta* décrites dans le chapitre précédent. Les résultats nouvellement acquis nous ont permis d'avoir une vision un peu plus détaillée de l'étendue et de la structure des zones de sympatrie avec et sans hybridation, associée à une description plus fine de la structure génétique de nos deux espèces focales sur cette aire d'étude étendue.

4.1 Distribution de *J. albifrons* / *J. prae-hirsuta*

L'ensemble des résultats du premier chapitre étaient basés sur deux zones réduites (baie de Morlaix en Bretagne et plages du débarquement en Normandie) séparées par une grande région non échantillonnée. L'échantillonnage beaucoup plus étendu qui est présenté ici nous a permis de découvrir de nouveaux sites présentant une population mixte *J. albifrons* / *J. prae-hirsuta* et nous avons pu constater la présence de quelques phénotypes intermédiaires dans certaines de ces populations. Une partie de ces résultats, basés uniquement sur l'identification morphologique des mâles échantillonnés, nous avait déjà permis d'évoquer lors de la rédaction de l'article formant le premier chapitre de cette thèse que la zone abritant des populations avec hybridation interspécifique est beaucoup plus étendue que ce qui était supposé (Solignac, 1969b, Solignac, 1969a, Solignac, 1978b). Nous pouvons confirmer ici que des phénotypes clairement intermédiaires ont été trouvés dans quatre sites, dont trois sur les côtes Normandes et un dans l'archipel des Scilly (étoiles rouges sur la Figure 2.9).

Le nouvel échantillonnage présenté ici nous renseigne également sur deux autres points intéressants concernant la distribution des espèces. Premièrement, nous n'avons trouvé aucune population constituée uniquement de *J. prae-hirsuta* (*i.e.* en absence de *J. albifrons* à proximité immédiate). Il se peut que de telles populations existent et n'aient pas été échantillonnées (*e.g.* zones à *Fucus vesiculosus* ou *Fucus serratus* ne présentant pas d'habitat de type galets à proximité immédiate), mais il est clair que les deux espèces sont souvent en mélange, que ce soit en Bretagne ou en Normandie, et cela est probablement vrai sur une très grande partie de leur aire de distribution (Solignac, 1978b). Deuxièmement, il semble qu'il y ait une coupure dans la distribution de *Jaera prae-hirsuta* au niveau de la côte Ouest du Cotentin, aucun individu n'ayant été trouvé entre Tanguy et Goury (sites 19 et 28, Figure 2.6). La distribution de *J. albifrons*, bien que discrète en raison de la nature de l'habitat de cette espèce, semble plus homogène sur l'ensemble des côtes de la Manche (tout du moins sur la portion que nous avons échantillonnée).

4.2 Structure génétique de la paire *J. albifrons* / *J. praeheirsuta*

Le premier résultat de cette nouvelle étude basée sur des données génomiques (5 068 SNPS) est la confirmation claire de la situation décrite avec 23 marqueurs microsatellites (chapitre 1) : la différenciation génétique inter-spécifique varie entre répliques de populations mixtes et reflète une hybridation introgressive forte dans nos deux populations Normandes, et inexistante ou faible dans les populations Bretonnes que nous avons analysées dans le chapitre 1 (populations de Grandcamp et Sainte-honorine, numéros 30 et 32 en Normandie et populations de Bouillenou, Saint-Jean, Lingoz_1, Lingoz_2, Inizan_1, Inizan_2, numéros 8, 9, 12, 13, 15 et 16 en Bretagne). La différenciation n'est cependant pas nulle dans ces populations Normandes (e.g. distribution des triangles et carrés marrons sur l'ACP, Figure 2.9), ce qui est en phase avec notre hypothèse de barrière semi-perméable avec flux de gènes hétérogènes entre loci (chapitre 1), la multiplicité et la distribution aléatoire des RAD-tags le long du génome permettant de capter efficacement cette hétérogénéité (cf. chapitres 5 et 6).

De plus, Les nouveaux sites échantillonnés nous apportent de nouvelles informations, dont la plus importante est la suivante : il semble exister des situations intermédiaires (ou différentes) entre les deux systèmes décrits ci-dessus. Ainsi la population de St-Agnes (îles Scilly), dans laquelle un individu à phénotype intermédiaire a été trouvé, montre une structure génétique inter-spécifique identique à la situation Bretonne (forte différenciation). Ce résultat est très intéressant car il indique qu'il existe des populations dans lesquels les deux espèces coexistent en mélange dans le même habitat (sous les pierres, comme en Normandie) avec de rares individus à phénotype intermédiaire (comme en Normandie) mais sans trace d'introgression (comme en Bretagne, où les individus n'occupent pas le même habitat). La position parfaitement centrale du génotype de l'individu à phénotype intermédiaire sur l'ACP (étoile de couleur rouge sur la Figure 2.9) semble compatible avec l'hypothèse qu'il s'agit d'un hybride de première génération. Cette nouvelle population apporte ce qui est peut-être le résultat le plus important de ce chapitre : les processus d'hybridation et d'introgression ne semblent pas strictement contraints par la disponibilité d'habitats différenciés. On peut par ailleurs spéculer sur la nature des barrières d'isolement en jeu. L'hybridation peut avoir lieu, signe peut-être que la barrière comportementale, bien que forte, n'est pas totale et permet l'apparition d'hybrides de première génération. Aucun signe d'introgression n'ayant été détecté, on peut supposer que les hybrides souffrent ensuite d'une fécondité réduite, soit en lien avec l'inadéquation de leur phénotype pour la reproduction (barrière post-zygotique de type extrinsèque selon Coyne and Orr (2004)), soit due à l'action supplémentaire de barrières post-zygotiques liés à des incompatibilités génétiques (cf. chapitre 3).

Par ailleurs, la nouvelle population mixte située sur la pointe du Cotentin (Goury, site numéro 28) est intéressante de par sa situation géographique intermédiaire entre nos sites d'intérêts Bretons et Normands. L'échantillon provenant de ce site est très limité (5 individus *J. albifrons* et 1 individu *J. praehirsuta*) mais l'ACP semble indiquer une différenciation génétique moyenne entre les deux espèces. Cette observation doit être confrontée à des analyses quantitatives (analyses de variance moléculaire et d'admixture), mais, si elle est confirmée, pourrait indiquer une autre situation d'introgession d'intensité moyenne. Il se peut également que la position de ces individus sur l'ACP reflète un effet de l'isolement géographique et non une situation d'hybridation introgressive intermédiaire entre espèces. Quel que soit le résultat des analyses qui seront conduites ultérieurement pour tester ces hypothèses, cette population devrait nous permettre de comprendre s'il existe une rupture géographique abrupte entre "zone hybride" contenant des populations introgressées et "zone d'isolement reproductif fort " ou bien si la frontière entre ces deux systèmes est plus graduelle.

Enfin, le nouveau site Yport (numéro 37, Figure 2.9) apporte également son lot d'informations. Ce site (situé à une centaine de kilomètres au Nord-Est des sites Normands étudiés dans le chapitre 1) abrite une population mixte dans un habitat unique (sous les pierres) dans laquelle des individus à phénotype intermédiaire ont été trouvés et dont l'analyse génétique indique clairement une situation d'hybridation introgressive. Il est possible que l'introgession soit particulièrement forte dans cette population, ce qui constitue une première observation intéressante car elle contribue également à l'idée qu'il existe autant de situations (avec plus ou moins d'hybridation et d'introgession) que de populations mixtes. La position des individus sur l'ACP interpelle également pour une autre raison. Les morphotypes intermédiaires semblent génétiquement plus proches des *J. albifrons* (étoiles et carrés violets, Figure 2.9) à Yport, à la différence de Sainte-Honorine (Figure 2.9, marques jaune) où ils sont plus proches des *J. praehirsuta*. Cette observation nous invite à essayer de comprendre si l'intensité de l'introgession asymétrique discutée dans le chapitre 1 (flux de gènes préférentiellement de *J. albifrons* vers *J. praehirsuta*) est seule responsable du patron observé par ACP.

Pour résumer, tel que décrit dans le papier de Saetre et al. (1999) portant sur les gobemouches (*Ficedula hypoleuca* et *F. albicollis*), il existerait une zone hybride (dans leur cas, zone hybride en cline) ainsi que des zones d'hybridation extérieures ne montrant pas le même pattern de différenciation. Cependant, à la différence de cette étude, où la fertilité hybride est moindre dans la zone en cline que dans la zone en « île », notre étude suggère que la fertilité hybride serait moindre aux îles Scilly qu'au niveau des sites normands. Il est cependant également possible que chaque zone où les deux espèces se trouvent en sympatrie soit une nouvelle situation pouvant déboucher sur plus ou moins d'hybridation. On n'aurait donc pas une zone

hybride et une zone non-hybride, mais un ensemble de réplicas avec plus ou moins d'hybridation.

4.3 Contexte de relations inter-populationnelles

Dans cette étude nous avons montré que *J. ischiosetosa* apparait comme le groupe le plus divergent entre les quatre espèces du complexe *Jaera albifrons* présentes le long des côtes européennes, permettant d'envisager une utilisation future de cette espèce comme outgroup lors d'inférences démographiques (voir chapitre 5). *Jaera forsmanni*, espèce souvent décrite comme proche des *J. praehirsuta* (Solignac, 1981 + Marius Wenzel et Stuart Piertney communication personnelle), se situe étonnamment près des *J. albifrons* sur la première ACP. Cependant, cela est dû à un effet d'observation des axes retenus, et une autre ACP réalisée sans *J. ischiosetosa* montre les *J. forsmanni* au sein des *J. praehirsuta* (ACP non montrée ici). De plus, les résultats de l'arbre de relations inter-populationnelles obtenu avec *Treemix* viennent confirmer ce lien entre *J. forsmanni* et *J. praehirsuta* (Figure 2.8).

Un seul évènement de divergence au sein de chacune de ces espèces semble avoir eu lieu au vue des résultats de *Treemix*, suivi par des divergences entre populations, qui pourraient suivre un gradient géographique, hypothèse soutenue par la distribution de *J. albifrons* observée sur la seconde ACP.

Enfin, les résultats obtenus par *Treemix* confirment les flux de gènes inter-spécifiques au sein de Yport et des sites normands en général décrits ci-dessus sur la base seule de l'ACP, avec comme information supplémentaire la direction, montrant une migration de *J. albifrons* vers *J. praehirsuta*, soit une asymétrie d'introgession, confirmée par l'ACP où l'on observe un décalage des individus *J. praehirsuta* vers les individus *J. albifrons*, signe que les *J. praehirsuta* sont plus introgressés par les *J. albifrons* que l'inverse.

4.4 Méthode RAD-seq

Le choix du protocole de séquençage de nos échantillons par ddRAD (Mse1/Pst1) s'est avéré judicieux, car le jeu de données final comportait plus de 5 000 SNPs ayant passé les différents filtres tels que le nombre de données manquantes ou le nombre minimal d'allèles autorisé. De plus, l'analyse des témoins et réplicas permet de confirmer la fiabilité de nos résultats. Les populations et nombre d'individus gardés pour cette analyse se sont révélés être suffisants pour observer des patterns particuliers de différenciation génétique. Le protocole RAD-seq décrit

dans ce chapitre (annexes 1 et 2) sera réutilisé à l'identique pour les analyses présentées en chapitres 5 et 6.

4.5 Résultats marquants et perspectives

Ce chapitre avait pour objectif de décrire la différenciation génétique entre *J. albifrons* et *J. praehirsuta* à une plus fine échelle génomique et sous une structure spatiale élargie, ainsi que de la contextualiser en terme de relations interspécifiques entre espèces du complexe *Jaera albifrons*.

Il a finalement été montré au cours de ce chapitre que :

- les sites échantillonnés confirment des niveaux de différenciation inter-spécifique différents avec une forte différenciation en Bretagne et de l'hybridation introgressive en Normandie.
- une introgression asymétrique de *J. albifrons* dans *J. praehirsuta* aurait lieu, et cette introgression semble suivre un certain gradient géographique de la Bretagne à la Normandie.
- Un hybride a été trouvé au site Saint-Agnès aux Iles Scilly, malgré une grande différenciation entre *J. albifrons* et *J. praehirsuta*, et une absence d'introgression visible.
- des barrières à la reproduction autres qu'écologique semblent être en œuvre dans le maintien de la différenciation interspécifique.
- le long des côtes européennes, l'espèce *J. ischiosetosa* est la plus divergente alors que *J. forsmanni* semble plus proche de *J. praehirsuta*.
- le protocole ddRAD est adapté à nos espèces.

Enfin, le nouvel échantillonnage n'a pas permis de détecter de *J. praehirsuta* à l'ouest de Goury jusqu'en Bretagne, ou entre le site de Longues et celui d'Yport. La question de la migration d'individus *J. praehirsuta* « pures » dans les sites avec hybridation pour le maintien des morphes parentaux est donc toujours d'actualité. Une analyse détaillée et quantitative de la structure génétique inter- et intra-spécifique, incluant une analyse d'isolement par la distance, l'analyse génétique de femelles et des préférences de choix de partenaires pouvant expliquer les patterns d'hybridation asymétrique trouvés en Normandie sont envisagés.

Plus globalement, comprendre l'histoire démographique qui a eu lieu entre ces deux espèces, quelles sont les barrières à la reproduction actives, ou encore rechercher l'emplacement de zones fortement différenciées maintenant la divergence interspécifique le long du génome seront les prochains objectifs abordés durant cette thèse.

Chapitre 3

ISOLEMENT REPRODUCTIF

Comprendre la nature et la force des barrières impliquées dans l'isolement reproductif est essentiel dans l'étude globale de la divergence entre espèces. Nous avons montré dans les deux chapitres précédents que la différenciation génétique interspécifique est variable en Bretagne et en Normandie, et nous nous sommes interrogés sur les barrières maintenant une telle différenciation en Bretagne, et les mécanismes en jeu dans le maintien des phénotypes parentaux dans le cas particulier des populations avec hybridation introgressive. Ce chapitre a pour objectifs de décrire les croisements réalisés entre *J. albifrons* et *J. praehirsuta* durant ma thèse, et les résultats obtenus afin d'analyser l'isolement reproductif entre ces deux espèces et tenter d'identifier les différentes barrières impliquées, tels qu'un isolement pré-copulatoire de type choix de partenaire ou post-copulatoire de type dépression hybride.

1. Introduction

Face à des niveaux d'isolement variables entre espèces sympatriques, identifier l'importance de chaque barrière contribuant au maintien de la divergence est un véritable défi, de multiples barrières pré-zygotiques et/ou post-zygotiques pouvant être impliquées dans l'isolement (Dobzhansky, 1937, Gavrillets, 2000, Schluter, 2001).

Ainsi Coyne and Orr (2004) proposent de classer les barrières en trois grandes classes (cf. introduction générale) : i) les barrières à la reproduction pré-copulatoires, telles que l'isolement comportemental, écologique ou encore mécanique ; ii) les barrières prézygotiques post-copulatoires, en particulier de type gamétique ; et iii) les barrières post-zygotiques de type extrinsèques (mauvaise adaptation à la niche écologique, ou "stérilité comportementale") ou intrinsèques (hybrides souffrant d'une viabilité ou d'une fertilité réduite).

Des situations sympatriques répétées avec un niveau d'isolement variables sont des cas intéressants pour rechercher la contribution relative de différentes barrières à la spéciation et au maintien de la divergence entre espèces (Rasanen & Hendry, 2014), cas présenté par le complexe *Jaera albifrons* et la situation décrite entre les espèces *J. albifrons* et *J. praehirsuta* en détail dans les chapitres précédents. Il est connu au sein de ce complexe qu'il existe une forte barrière comportementale, mais également des barrières écologiques et génétiques (*i.e.* post-zygotiques de type intrinsèque) (Solignac, 1981). En revanche, l'absence d'isolement temporel et d'isolement mécanique ont été démontrées par le passé entre les espèces du complexe *Jaera albifrons* (Jones & Naylor, 1971b, Bocquet, 1953) et une barrière de type prézygotique gamétique est assez peu probable, bien que cela n'ait jamais été étudié en détail (*e.g.* Solignac 1978 p. 80-82).

Comme cela a été décrit dans l'introduction générale et dans le chapitre 1 (Ribardière et al. 2017), la barrière comportementale consiste en un choix de partenaire, principalement fait par les femelles

envers les mâles, basé sur les caractères sexuels secondaires, *i.e.* soies sur les péréiopodes, de ceux-ci. Le mâle effectue ainsi une parade tactile différente selon son phénotype (propre à son espèce), en brossant les femelles, qui vont ainsi pouvoir distinguer les mâles conspécifiques des mâles hétérospécifiques. Il a été clairement établi que cette barrière a un rôle prépondérant dans l'isolement entre toutes les espèces du complexe *Jaera albifrons* (99 à 99.9% de l'isolement d'après Solignac, 1978 p. 85). Cependant il existe également une certaine variabilité dans la distribution des espèces entre habitats différenciés (liés principalement à l'étagement dans la zone intertidale sur l'estran) et certains auteurs ont associés les différentes espèces à des préférences écologiques assez clairement identifiées (Naylor & Haahtela, 1966, Carvalho, 1989a), au premier rang desquelles la différence d'habitat observée entre *Jaera albifrons* (sous les pierres) et *J. praehirsuta* (sur les algues brunes) le long des côtes Françaises (Solignac, 1978b). Cette distribution semble cependant loin d'être figée (*e.g.* nous avons échantillonné des *J. albifrons* sur les algues le long du littoral Canadien, et, bien entendu, des *J. praehirsuta* sous les pierres dans les différentes populations décrites en chapitre 1 et 2) et l'isolement écologique semble pour le moins d'intensité très variable le long de l'immense aire de distribution du complexe.

Enfin, une barrière de type post-zygotique liée à la ségrégation d'incompatibilités génétiques a été mise en évidence chez certaines paires d'espèces, le travail le plus complet sur ce sujet ayant été réalisé par M. Solignac avec *J. albifrons* et *J. ischiosetosa* (Solignac, 1978b). Par exemple chez cette paire d'espèces, les hybrides F1 produisent peu de descendants (fécondité normale des F1 mais développement anormal de certains des juvéniles qu'ils produisent). Aucune donnée chiffrée n'existe pour la paire d'espèces qui nous intéresse, mais Solignac pensait que le niveau d'introgression suspectée dans la population de Luc sur Mer (Normandie, site 34 - Figure 2.1, chapitre 2) était tout à fait incompatible avec une barrière post-zygotique forte, contrairement à ce qui a été écrit par erreur dans la discussion de l'article publié. Il est donc important de savoir quel est le niveau d'isolement potentiellement lié à des incompatibilités génétiques entre nos deux espèces focales.

Nous tenterons de détecter ici si la nature et l'intensité de l'isolement varient en fonction de nos situations avec et sans hybridation, et quelle(s) barrière(s) participe(nt) au maintien des deux espèces même dans des situations d'hybridation et d'introgression.

Pour cela, nous avons réalisé une série de croisements expérimentaux intra- et inter-spécifiques (F1 et backcross) avec des individus issus de deux types de populations (avec et sans hybridation). Les croisements décrits dans ce chapitre ne laissaient aucun choix aux individus (une seule femelle et un seul mâle isolés). Afin de pouvoir quantifier l'isolement reproductif entre *J. albifrons* et *J. praehirsuta*, le nombre de juvéniles obtenus à la première portée et leur taux de survie en fonction du type de croisements réalisés et de la région d'échantillonnage des individus (le facteur "région" désignant le

type de population : avec *versus* sans hybridation) ont été comparés. Pour le reste du manuscrit de thèse on appellera "F1" les juvéniles issus de croisements intra ou inter-spécifiques de première génération, et "backcross" les juvéniles issus de croisements en retour entre une femelle hybride F1 et un mâle de l'une ou l'autre espèce parentale. Pour l'ensemble des croisements réalisés il était supposé que les juvéniles F1 et Backcross obtenus provenant de Bretagne, zone où aucun hybride n'a été détecté en milieu naturel, soient moins nombreux et présentent un taux de survie moindre qu'en Normandie, zone d'hybridation et d'introgession (chapitre 2). En effet, une hypothèse est que la barrière comportementale et la barrière génétique agissent de concert en Bretagne, au vue de la forte différenciation génétique interspécifique, la barrière écologique n'étant pas suffisamment forte pour empêcher les individus de se retrouver, parfois, en syntenie (quelques individus de chaque espèce étant retrouvés sur l'habitat préférentiel de l'autre espèce). En Normandie, malgré l'introgession, la variabilité de phénotypes est assez limitée, avec des morphes parentaux qui semblent rester fixes, faisant penser qu'il existe un isolement, qui pourrait être d'ordre génétique.

Bien que l'isolement comportemental est probablement l'une des causes les plus fréquentes dans la spéciation animale (Coyne & Orr, 2004), les incompatibilités génétiques accumulées lors de la divergence peuvent amener à une réduction de la valeur adaptative hybride lors de remises en contact de ces espèces (Dobzhansky, 1937, Orr & Turelli, 2001, Maheshwari & Barbash, 2011).

Solignac (1976) avait par ailleurs déjà remarqué une dépression hybride lors de croisements expérimentaux effectués entre les espèces *J. albifrons* et *J. ischiosetosa* : la fertilité des hybrides F1 était diminuée. On peut donc se demander si entre les espèces *J. albifrons* et *J. praeheirsuta*, ce même genre de dépression hybride est présent, voir même si cela a pu conduire à un renforcement de l'isolement reproductif pré-zygotique (Kirkpatrick & Servedio, 1999, Servedio & Noor, 2003, Lemmon & Kirkpatrick, 2006).

Une autre possibilité dans l'étude de l'isolement de type comportemental est l'étude des préférences femelles, pour laquelle quelques tests ont été réalisés et présentés par la suite dans ce chapitre.

2. Matériel et méthodes

2.1 Croisements contrôlés

D'Avril 2015 à Juin 2016, trois types de croisements ont été réalisés en Bretagne et en Normandie : des croisements intraspécifiques *J. albifrons* - *J. albifrons* et *J. praeheirsuta* - *J. praeheirsuta*, des croisements Interspécifiques *J. albifrons* (mâle ou femelle) - *J. praeheirsuta* (femelle ou mâle), et des backcross femelle hybride - mâle *J. albifrons* ou *J. praeheirsuta* (Figure 3.1).

Pour ce faire, nous avons procédé dans les deux régions à un même protocole (Figure 3.2), échantillonnant tout d'abord des femelles gravides sur 5 sites en Bretagne (chapitre 2 - Tableau 2.1

et Figure 2.1 - Sites 8, 12, 13, 15, 16) et 6 sites en Normandie (chapitre 2 - Tableau 2.1 et Figure 2.1 - Sites 27, 29, 30, 32, 36, 37). Ces femelles dont l'espèce était inconnue, ont ensuite été élevées (seules) dans des plaques de 6 puits (un individu par puits, chaque puits a un diamètre d'environ 3,5 cm et contient environ 10 ml d'eau de mer filtrée à 3 µm) jusqu'à obtention de juvéniles. Suivant les conditions décrites dans Bocquet (1953, p. 212-213), chaque puits contenait de petits morceaux d'algues vertes (*Enteromorpha sp.*) et une petite feuille d'orme fournissant abris et nourriture. Les plaques ont été maintenues dans des armoires à thermostat à 17 °C avec un cycle de lumière / obscurité de 11h / 13h et l'eau de mer a été changée une fois par semaine (avec des algues et des feuilles au besoin). Les embryons se développent dans un marsupium (poche d'élevage contenant en moyenne environ 15 descendants) jusqu'à ce que la femelle les libère (observation tous les deux jours des femelles afin de connaître la date d'obtention des juvéniles). À cette étape, la progéniture mesure environ un demi-millimètre et ressemble beaucoup aux adultes (développement direct, pas de stade larvaire pélagique).

Les descendants ont ensuite été isolés un par un dès leur libération et ont été élevés individuellement dans les mêmes conditions que les adultes, à deux exceptions près: l'eau de mer n'a pas été changée au cours de la première semaine, et les feuilles d'orme n'ont été ajoutées que pendant la deuxième semaine. Dans ces conditions, les individus atteignent leur maturité, *i.e.* apparition des caractères sexuels secondaires chez les mâles, à environ 4 semaines, et peuvent ainsi être sexés et leur espèce déterminée. Les femelles ont été identifiées selon le phénotype, *i.e.* caractères sexuels secondaires spécifiques, de leurs frères. Tous les individus sont vierges et ont tous été obtenus et élevés dans les mêmes conditions de laboratoire. Ce sont ces individus (nommés "F0") qui ont été utilisés pour réaliser nos croisements intraspécifiques ou interspécifiques contrôlés. Afin de s'affranchir des problèmes de précédence dans le choix de partenaire pour la descendance (Robinson et al., 1994, Klibansky & McCartney, 2014) pouvant avoir lieu en milieu naturel dans des zones de sympatrie, chaque femelle ne s'est vu proposer qu'un mâle pour la reproduction, pour ainsi pouvoir quantifier l'isolement reproductif en l'absence de compétition.

En Bretagne, 116 femelles gravides ont été prélevées en milieu naturel. Sur 48 prélevées sur des algues, 19 femelles ont donné 1 portée. Pour chaque portée, jusqu'à 12 juvéniles ont été gardés en laboratoire. 68 femelles avaient été prélevées sous des pierres, et seules 16 d'entre elles ont eu une portée. Au total, sur les 35 portées, 14 étaient des portées de *J. albifrons* (57 femelles – 44 mâles), 12 des portées de *J. prae-hirsuta* (22 femelles – 39 mâles), 3 des portées de *J. forsmanni* (9 femelles et 15 mâles) et 6 des portées non identifiées pour cause de mortalité ou d'absence de mâle dans la portée. Sur l'ensemble des portées le sexe ratio ne semblait pas biaisé (cf. chapitre 4) et le taux de mortalité à maturité, à 5 semaines environ, était de 35%.

En Normandie, 100 femelles gravides ont été prélevées en milieu naturel sous des pierres. 31 femelles ont eu une portée, et pour chaque portée, jusqu'à 12 juvéniles ont été gardés en laboratoire. Au total, sur les 31 portées, 14 étaient des portées de *J. albifrons* (72 femelles – 65 mâles), 16 des portées de *J. praehirsuta* (75 femelles – 73 mâles), 1 portée n'a pu être identifiée pour cause d'absence de mâle dans la portée. Sur l'ensemble des portées le sexe ratio ne semblait pas biaisé, et le taux de mortalité à maturité, soit 5 semaines environ, était de 21%.

Une fois ces juvéniles F0 devenus adultes, soit après environ 5 semaines, un total de 72 croisements intraspécifiques et interspécifiques ont été réalisés (Tableau 3.1). Par la suite, les nouveaux juvéniles hybrides femelles, *i.e.* issues des croisements interspécifiques, ont été croisées à leur maturité avec un individu mâle F1 pure *J. albifrons* ou *J. praehirsuta* afin d'obtenir des individus de type backcross (30 croisements). Le protocole utilisé pour ces backcross a été légèrement modifié puisque deux sœurs ont été mises en contact avec un même mâle (voir Figure 3.2) avant d'être ré-isolées pour pouvoir assigner les juvéniles à leur mère. Ce design des croisements backcross a été conçu pour pouvoir être utilisé ultérieurement dans la recherche de QTL (données non présentées dans le manuscrit de thèse).

Ainsi 43 croisements intraspécifiques ont été réalisés : 16 croisements *J. albifrons* et 5 croisements *J. praehirsuta* en Bretagne, et 16 croisements *J. praehirsuta* et 6 croisements *J. albifrons* en Normandie. En Bretagne, 14 croisements interspécifiques ont été réalisés, dont 8 croisements de type femelle *J. albifrons* – mâle *J. praehirsuta* et 6 de type femelle *J. praehirsuta* – mâle *albifrons*. En Normandie, 15 croisements interspécifiques ont été réalisés, dont 9 croisements de type femelle *J. albifrons* – mâle *J. praehirsuta* et 6 de type femelle *J. praehirsuta* – mâle *albifrons*. Enfin 4 croisements de type backcross ont été réalisés en Bretagne, alors que 26 backcross ont été réalisés en Normandie.

Pour chaque croisement, la région d'origine des individus, le type de croisement réalisé (intraspécifique, interspécifique ou backcross), la durée du croisement (de la mise en contact des deux parents à l'obtention de la première portée) et, pour chaque portée (plusieurs portées ayant généralement été isolées jusqu'au décès de la mère), le nombre de juvéniles puis leur survie à exactement 35 jours et à environ 70 jours (+ ou - 5 jours) ont été notés. Les femelles des deux espèces focales peuvent stocker du sperme pendant plusieurs mois, ce qui nous a permis d'observer la production de portées successives même si le mâle d'un croisement donné était mort avant la femelle. Cependant, dans cette étude nous ne nous intéresserons qu'à la première portée de chaque famille, sauf cas particulier précisé dans le texte, le nombre de portées suivantes obtenues étant directement lié à la survie des femelles et à des contraintes expérimentales (certaines portées n'ont pas été conservées). Enfin, tous les croisements (n=9) dont la mère a survécu moins de 24 jours (durée de vie minimum d'une mère ayant produit des juvéniles) ont été supprimés des analyses pour ne pas interpréter de faux résultats négatifs.

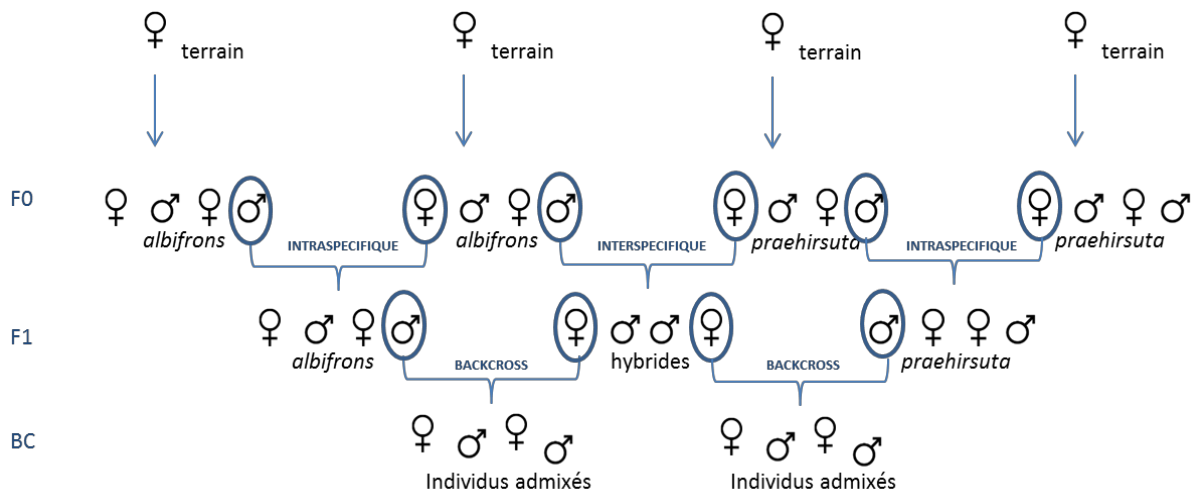


Figure 3.1 : Principe général des différents types de croisements réalisés en laboratoire

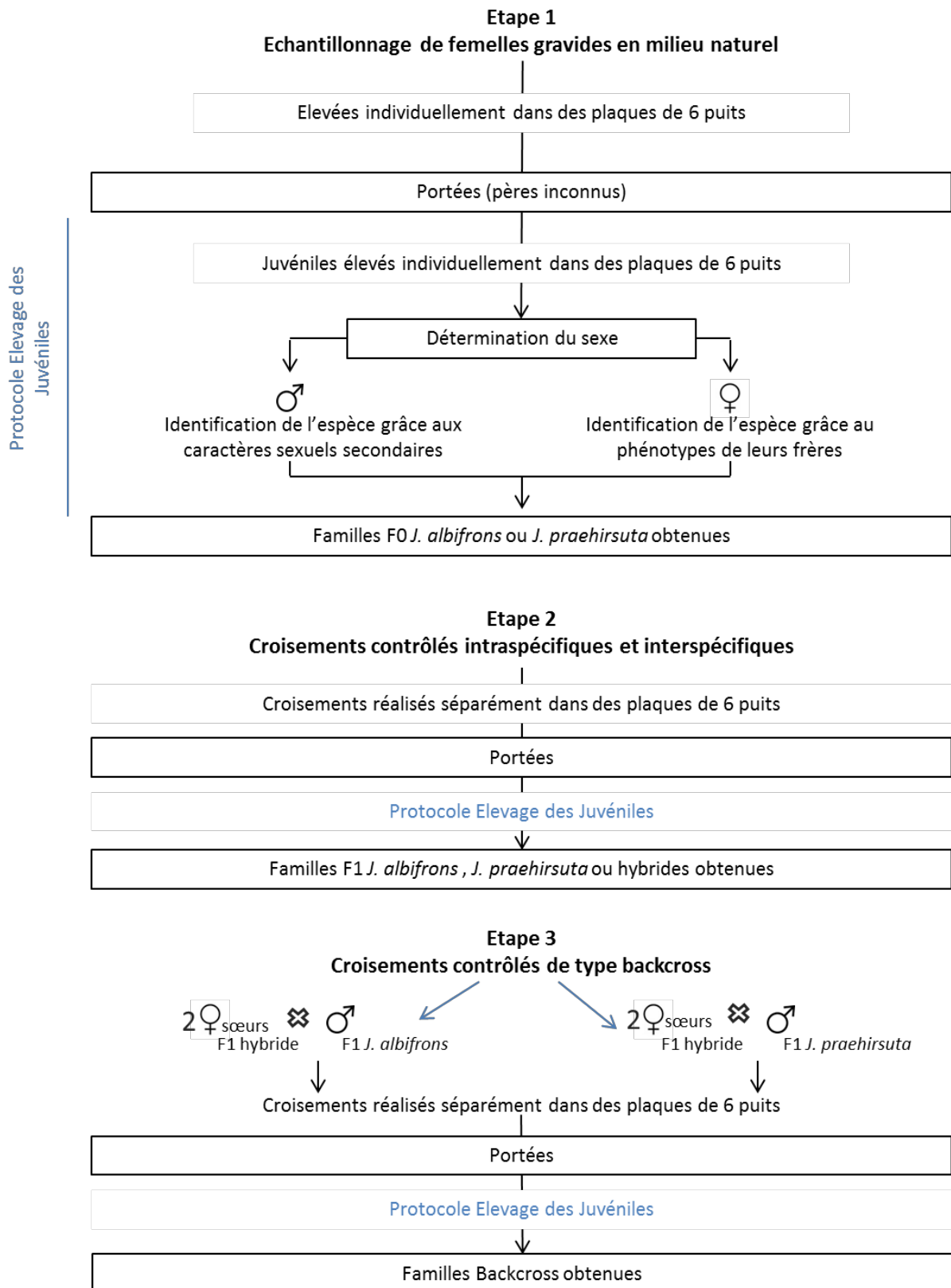


Figure 3.2 : Protocole d'élevage en laboratoire

2.2 Préférences femelles

Il est connu dans la littérature qu'il existe une barrière comportementale importante au sein du complexe *Jaera albifrons* (Solignac, 1978a, Solignac, 1978b, Solignac, 1981). Afin de détecter en Bretagne la force de l'isolement comportemental entre *J. praehirsuta* et *J. albifrons*, 650 tests pour détecter la préférence de femelles ont été réalisés à partir de 65 femelles issues de Backcross. Dans le but de détecter cette éventuelle préférence femelle envers les mâles faisant une parade sexuelle, nous avons placés deux à deux chaque individu femelle successivement avec 5 mâles *J. albifrons* puis 5 mâles *J. praehirsuta* (ou inversement, aléatoirement) pendant 3 min, puis avons noté son comportement (1 : acceptation ou 0 : rejet). Le nombre d'acceptation (de mâles *J. albifrons*, puis *J. praehirsuta*) sur le nombre de test total est calculé. Entre chaque test, *i.e.* nouveau mâle proposé, la femelle avait un temps de repos d'au moins 5h. Lorsque le mâle n'effectuait pas de parade, *i.e.* action de grimper sur la femelle et d'effectuer des mouvements d'aller-retour à l'aide des périopodes portant les soies spécifiques à l'espèce (cf. introduction générale), aucune action de la part de la femelle ne pouvant alors être détectée, la rencontre n'était pas prise en compte (Figure 3.3, A). Si le mâle effectuait une parade mais que la femelle donnait des coups de pléotelson afin de faire partir ce dernier pour signifier son refus d'accouplement (Solignac, 1981), la rencontre était marquée comme rejetée de la part de la femelle (Figure 3.3, B). Enfin, lorsque le mâle effectuait la parade et que la femelle ne se débattait pas, laissant ainsi la parade avoir lieu, la rencontre était notée comme acceptée de la part de la femelle (Figure 3.3, C).

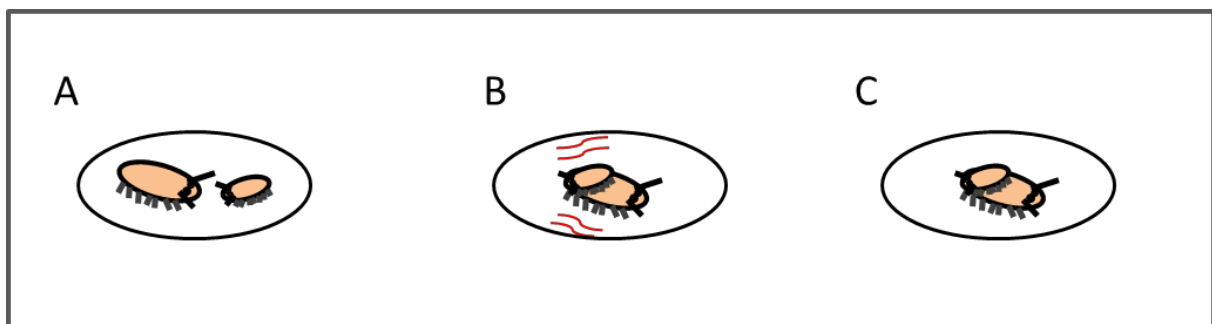


Figure 3.3 : Actions possibles lors de la rencontre entre le mâle (plus petit) et la femelle (plus grosse).
A) pas d'action, B) rejet du mâle par la femelle à coups de pléotelson, C) parade acceptée par la femelle.

3. Résultats

3.1 Croisements contrôlés

3.1.1 Isolement pré-zygotique

En Bretagne, relativement peu de croisements intraspécifiques *J. praehirsuta* ont pu être réalisés (Tableau 3.1), cependant ceux-ci semblent avoir fonctionné aussi bien que ceux entre individus *J. albifrons*. En effet chez *J. albifrons* 15 croisements sur 16 réalisés ont donné des juvéniles, contre 4 sur 5 croisements intraspécifiques *J. praehirsuta* (test exact de Fisher $p=0.429$). Malgré un faible nombre de croisements interspécifiques effectués, une certaine asymétrie entre espèces est remarquable puisque sur 8 croisements effectués, aucun n'a donné de portée dans le sens femelle *J. albifrons* – mâle *J. praehirsuta*, alors que 2 portées sur 6 ont donné des juvéniles dans le sens femelle *J. praehirsuta* – mâle *J. albifrons* (cette tendance n'est cependant pas significative, $p=0.165$). Enfin, de par le faible rendement des croisements interspécifiques, seulement deux croisements backcross ont pu être réalisés, tous deux donnant cependant lieu à des juvéniles.

En Normandie, sur 6 croisements intraspécifiques *J. albifrons*, 5 ont donné des juvéniles, alors que chez *J. praehirsuta* 11 croisements sur 16 ont données des juvéniles ($p=0.634$). Comme en Bretagne, malgré un faible nombre de croisements interspécifiques effectués, une tendance (non-significative) à l'asymétrie dans le même sens que précédemment est à nouveau notée puisque 2 sur 9 croisements effectués ont donné des portées dans le sens femelle *J. albifrons* – mâle *J. praehirsuta*, alors que 3 sur 6 ont donné des juvéniles dans le sens femelle *J. praehirsuta* – mâle *J. albifrons* ($p=0.329$). Enfin, une bonne réussite des croisements backcross est à noter car 83 à 85% des croisements ont donné des juvéniles (soit 10/12 femelle hybride – mâle *J. albifrons* et 12/14 femelle hybride – mâle *J. praehirsuta*).

Les données analysées région par région et espèce par espèce s'accompagnent nécessairement d'effectifs réduits, et il est intéressant d'observer si des patrons généraux se dessinent. Sur l'ensemble des deux régions et des deux espèces, on peut observer (Tableau 3.1) un nombre de croisements à succès plus élevé pour les croisements intraspécifiques (35 sur 43) qu'interspécifiques (7 sur 29, $p<0.001$), et cela reste vrai dans chaque région séparément ($p<0.001$ en Bretagne et $p=0.0233$ en Normandie). Il semble que les croisements inter-spécifiques soient un peu plus faciles en Normandie (5 sur 15) qu'en Bretagne (2 sur 14) mais cette tendance n'est pas significative ($p=0.390$). De plus, dans les deux régions, une même asymétrie est observée entre les croisements interspécifiques. Ainsi toutes régions confondues et avec 2 croisements réussis sur 17, les croisements femelle *J. albifrons* – mâle *J. praehirsuta* semblent avoir eu un moins bon succès que les croisements femelle *J. praehirsuta* – mâle *J. albifrons*, avec 5 couples ayant donné des juvéniles sur 12, bien que cette tendance ne soit pas significative (test exact de Fisher $p=0.092$).

Tableau 3.1. Quantification de l'isolement reproductif en Bretagne et Normandie, à l'aide de croisements intraspécifiques, interspécifiques et Backcross. Le temps moyen d'obtention des juvéniles (en jours), le nombre moyen de juvéniles et leur survie sont basés uniquement sur la première portée de chaque famille.

	Nombre de croisements effectués	Nombre de croisements réussis	Temps moyen d'obtention de juvéniles	Nombre moyen de juvéniles	Survie à 35 jours	Survie à 70 jours
Bretagne						
Intraspécifique						
Femelle <i>J. albifrons</i>	16	15	40.7	6.7	82.2%	64.4%
Mâle <i>J. albifrons</i>						
Femelle <i>J. praehirsuta</i>	5	4	36.0	4.0	62.8%	38.3%
Mâle <i>J. praehirsuta</i>						
Interspécifique						
Femelle <i>J. albifrons</i>	8	0	-	-	-	-
Mâle <i>J. praehirsuta</i>						
Femelle <i>J. praehirsuta</i>	6	2	31.5	5.5	95.0%	80.0%
Mâle <i>J. albifrons</i>						
Backcross						
Femelle Hybride	2	2	30.0	8.5	75.7%	41.4%
Mâle <i>J. albifrons</i>						
Femelle Hybride	2	2	33.5	5.5	56.7%	56.7%
Mâle <i>J. praehirsuta</i>						
Normandie						
Intraspécifique						
Femelle <i>J. albifrons</i>	6	5	40.0	8.6	88.3%	55.6%
Mâle <i>J. albifrons</i>						
Femelle <i>J. praehirsuta</i>	16	11	44.3	9.6	83.9%	64.1%
Mâle <i>J. praehirsuta</i>						
Interspécifique						
Femelle <i>J. albifrons</i>	9	2	71.5	20.5	92.4%	100%
Mâle <i>J. praehirsuta</i>						
Femelle <i>J. praehirsuta</i>	6	3	68.7	4.3	61.7%	53.3%
Mâle <i>J. albifrons</i>						
Backcross						
Femelle Hybride	12	10	33.1	8.4	54.5%	39.3%
Mâle <i>J. albifrons</i>						
Femelle Hybride	14	12	26.6	11.8	71.3%	42.5%
Mâle <i>J. praehirsuta</i>						

- Temps d'obtention des juvéniles issus de la première portée

Le temps d'obtention des juvéniles montre des différences significatives entre croisements intra-spécifiques et inter-spécifiques, toutes régions confondues (Figure 3.4), et entre croisements backcross et intra-spécifiques (modèle linéaire généralisé, famille quasi-poisson, $p=0.0167$ et $p=0.00717$). Les juvéniles issus de backcross sont les plus rapides à obtenir, et les croisements interspécifiques montrent une grande variance (Figure 3.4, boxplot de gauche). Ce résultat est inchangé si on ne considère que les résultats obtenus en Normandie (résultat non montré), et il n'est pas possible de vérifier ce qu'il en est en Bretagne (trop peu de croisements ayant produit des juvéniles dans cette région).

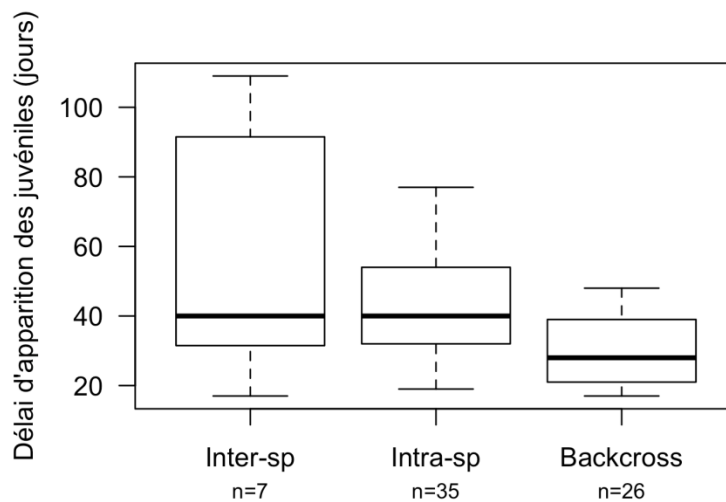


Figure 3.4 : Délai d'obtention des juvéniles en fonction du type de croisement effectué, toutes régions confondues

3.1.2 Isolement post-zygotique

- Nombre de juvéniles obtenus lors de la première portée

La taille des portées est similaire quel que soit le type de croisement effectué, toutes régions confondues (Figure 3.5, modèle linéaire généralisé, $p=0.356$).

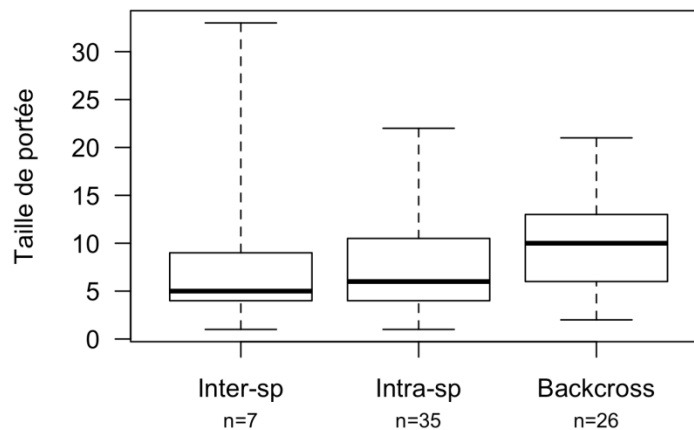


Figure 3.5 : Taille des portées (nombre de juvéniles) en fonction du type de croisement effectué sur l'ensemble des sites observés.

En Bretagne, le nombre de juvéniles par portée semble similaire entre les croisements intraspécifiques et interspécifiques (respectivement 5.3 juvéniles/portée vs 5.5), et légèrement plus grand pour les croisements backcross avec en moyenne 7 juvéniles/portée (chiffres basés sur 4 portées, pas de test réalisé).

En Normandie, les croisements intraspécifiques et Backcross ont donné en moyenne le même nombre de juvéniles par portée (respectivement 9.1 et 10.1), oscillant entre 8.4 à 11.8 juvéniles par portée selon le type de croisement considéré (Tableau 3.1). Une plus grande disparité est observable entre les croisements interspécifiques puisque les croisements femelle *J. albifrons* – mâle *J. praehirsuta* ont eu en moyenne une vingtaine de juvéniles sur la première portée, alors que les croisements femelle *J. praehirsuta* – mâle *J. albifrons* ont eu en moyenne 4.3 juvéniles par portées. Cependant cela n'est basé que sur respectivement 2 et 3 croisements réussis (pas de statistiques réalisés sur ces petits échantillons).

- Survie des juvéniles

Chez les espèces du complexe *Jaera albifrons*, il est connu que les individus deviennent matures à environ 4 semaines (Solignac, 1978b). Cette maturité se présente sous formes du pléotelson bien développé et des caractères sexuels secondaires chez le mâle bien discriminants. Le taux de survie a donc été mesuré dans un premier temps à 35 jours, une fois les juvéniles à maturité. Cependant, les individus issus des backcross ne présentaient pas de caractères suffisamment développés pour discriminer sans aucun doute les espèces à 35 jours. Il a donc été décidé de les garder jusqu'à

l'obtention de phénotypes similaires à ceux observés dans les croisements F1, ce qui a été obtenu environ 4 semaines plus tard. Le taux de survie a donc été recalculé au terme de 70 jours environ. Cependant, le taux de survie à 70 jours est bien entendu calculé sur un nombre d'individus plus restreint qu'à 35 jours.

La survie est en moyenne similaire sur l'ensemble des régions entre les croisements intra-spécifiques et inter-spécifiques à 35 et 70 jours (modèles linéaires généralisés, famille quasi-binomiale, $p=0.639$ et $p=0.482$). En revanche, à 35 jours, la survie est significativement plus faible chez les juvéniles issus de backcross, comparée à celle des croisements intraspécifiques ($p=0.0367$; Figure 3.6). A 70 jours, la survie des juvéniles issus de backcross semble également plus faible (Figure 3.7), bien que cette différence ne soit pas significative ($p=0.179$, avec une taille d'échantillon presque deux fois plus petite).

La survie entre 35 et 70 jours diminue en moyenne de 17.2 % (*i.e.* elle passe de 74.9 % à 35 jours à 57.8 % à 70 jours) sur l'ensemble des croisements dans les deux régions (soit 18.3 % en Bretagne et 16.2 % en Normandie), et chaque croisement a une survie à 35 jours similaire entre les deux régions (Tableau 3.1). Il y a trop peu de croisements pour analyser finement les différences entre espèces entre régions, bien qu'il semblerait qu'en Bretagne il y ait une légère tendance dans la différence de survie des croisements backcross selon le sens de croisement, puisqu'à 70 jours, la survie a diminué de 34 % dans le sens femelle hybride – mâle *J. albifrons*, alors qu'il est resté le même dans le sens femelle hybride – mâle *J. praehirsuta*.

La mesure du taux de survie à 70 jours n'est pas une mesure qui a été réalisée de manière aussi précise qu'à 35 jours (+ ou - 5 jours) mais elle correspond globalement à la durée d'élevage qui a été nécessaire pour que les individus mâles issus de croisements en retour acquièrent des caractères sexuels secondaires différenciés. Il est donc intéressant de comparer la survie "à maturité" des juvéniles issus des différents types de croisement (soit 35 jours pour les croisements intra- et interspécifiques et 70 jours pour les backcross). Nous obtenons 81 % de survie moyenne "à la maturité" pour les croisements intra- et interspécifiques des deux régions contre 45% de survie chez les backcross. Cela est observable en Normandie (81.6% vs 41%) et en Bretagne (80 % vs 49 %), bien que la différence soit moins importante et basée que sur 4 familles issues de backcross seulement (contre 22 familles en Normandie).

Le sex-ratio est équilibré et ne présente pas de différences entre croisements intraspécifique, interspécifique et backcross (présenté dans le chapitre 4, voir notamment la Figure 4.1, page 133), la règle de Haldane ne s'applique donc pas.

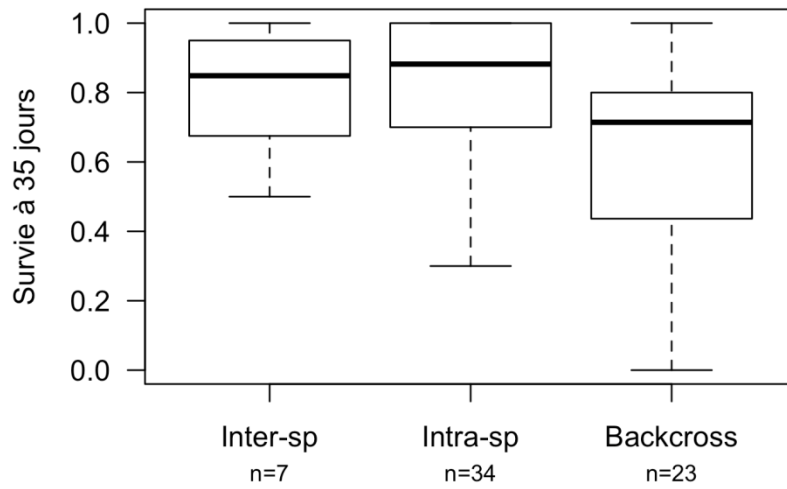


Figure 3.6 : Survie à 35 jours pour chaque type de croisement effectué, sur l'ensemble des sites observés

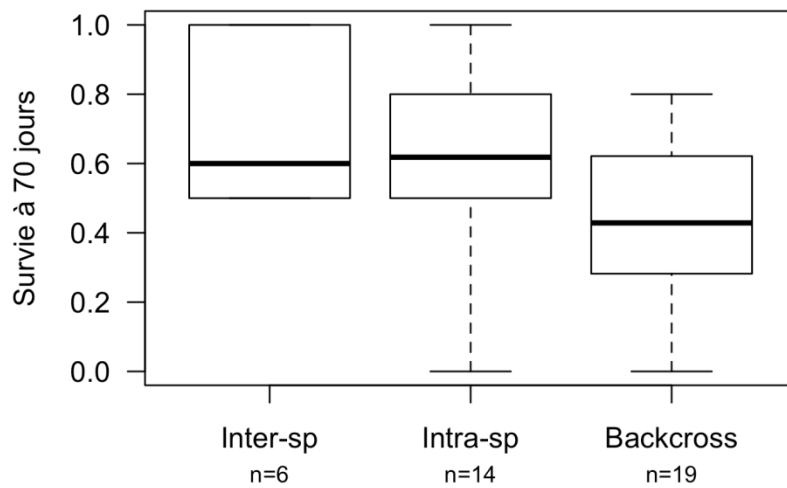


Figure 3.7 : Survie à 70 jours pour chaque type de croisement effectué, sur l'ensemble des sites observés

- Nombre de portées et temps moyen d'obtention des juvéniles sur l'ensemble des portées en fonction des types de croisements et des régions

Sur l'ensemble des régions et des croisements, les premières portées semblent avoir été généralement plus longues à obtenir que les suivantes (Figure 3.8). Enfin, la Normandie montre un nombre de juvéniles par portée supérieur à la Bretagne (10.2 vs 7.8) en moyenne et pour tous les types de croisements.

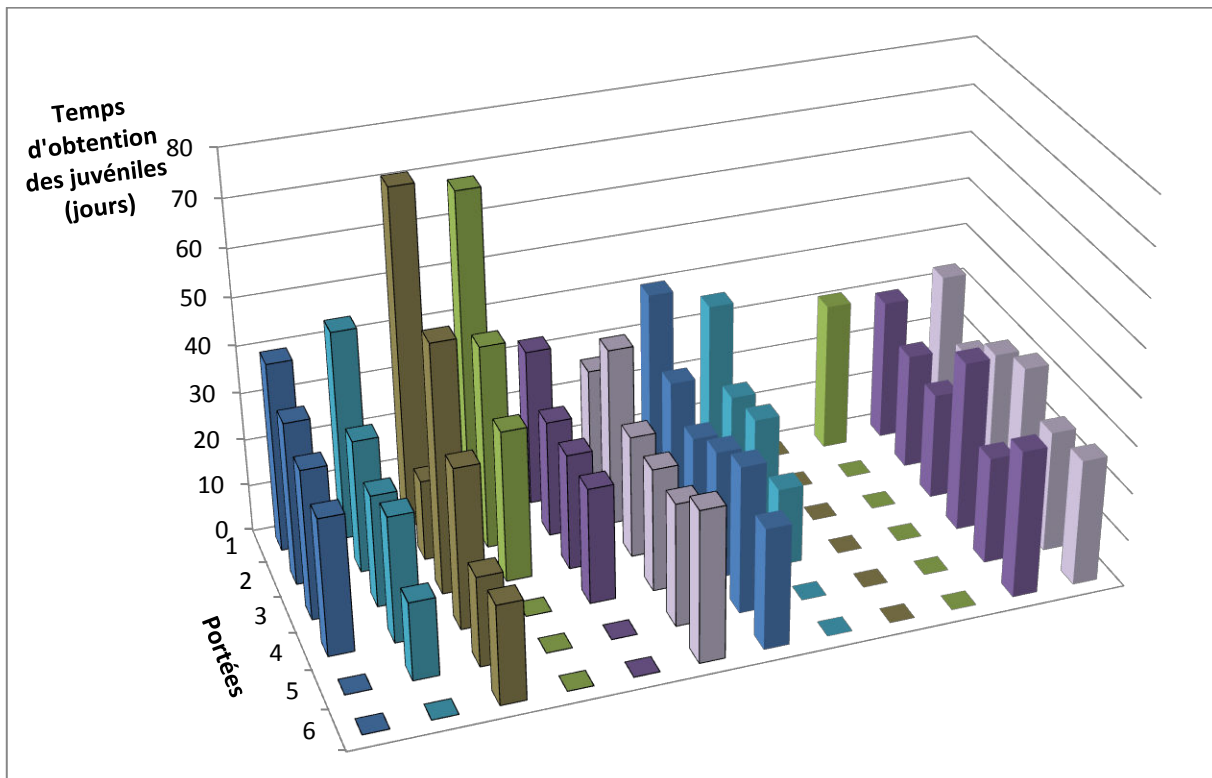


Figure 3.8 : Temps d'obtention des juvéniles par portées selon les croisements réalisés, Intraspécifique *J. albifrons* en bleu foncé, Intraspécifique *J. praeheirsuta* en bleu clair, Interspécifique femelle *J. albifrons* – mâle *J. praeheirsuta* en marron, Interspécifique femelle *J. praeheirsuta* – mâle *J. albifrons* en vert, Backcross femelle hybride – mâle *J. albifrons* en violet foncé et Backcross femelle hybride – mâle *J. praeheirsuta* en violet clair. Les croisements faits avec des individus normands sont encadrés en noir (1^{ère} moitié du graphique), tandis que ceux effectués avec des individus bretons ne sont pas encadrés (2^{ème} moitié du graphique).

3.2 Préférences femelles

Sur l'ensemble des 65 femelles bretonnes testées, le taux d'acceptation pour un mâle *J. albifrons* était de 64,2% contre 37,5% pour un mâle *J. praehirsuta*. Les mâles *J. albifrons* semblent donc avoir la préférence des femelles, et cela est toujours vrai lorsque l'on analyse séparément les croisements femelle hybride – mâle *J. albifrons* et femelle hybride - mâle *J. praehirsuta*. En effet, les préférences sont respectivement de 64,5% et 64,7% pour les mâles *J. albifrons*, et respectivement de 29% et 48,9% pour les mâles *J. praehirsuta*. On peut toutefois noter que 9 femelles issues de croisements de type femelle hybride – mâle *J. albifrons* n'ont jamais été abordées par un mâle *J. praehirsuta*. Le traitement statistique de ces données préliminaires sera réalisé ultérieurement, en combinaison avec une autre expérience de ce type réalisée par une étudiante de M2 au printemps 2016 (Elsa Pabion).

4. Discussion

Cette étude, portant sur 3 grands types de croisements et dans 2 régions différentes, fait ressortir des résultats intéressants sur la compréhension de l'isolement reproductif au sein du complexe *Jaera albifrons* et sur les forces relatives des barrières pré-zygotiques de type comportementales vs post-zygotiques.

Les résultats acquis avant cette étude dans les deux régions que nous étudions impliquent des niveaux d'isolement très différents : un fort isolement en Bretagne, supposé écologique, comportemental et génétique (Bocquet, 1972, Solignac, 1976), et des barrières semi-perméables avec introgression des génomes en Normandie (cf. chapitres 1, 2, 5 et 6), laissant supposer que les barrières post-zygotiques doivent être faibles. La réussite des croisements intraspécifiques, interspécifiques et backcross a été mesurée dans les deux régions, ainsi que le temps d'obtention et le nombre de juvéniles de la première portée de chaque croisement et la survie de ceux-ci à 35 jours et à environ 70 jours. Sur l'ensemble de cette étude, nous n'avons principalement présenté ici que les analyses basées sur la première portée, et non sur les portées suivantes, car la comparaison ne peut être stricte, l'effort d'élevage n'ayant pas été le même en fonction de l'utilisation des individus pour différentes expérimentations réalisées par la suite (notamment cartes génétiques présentées en chapitre 6). Enfin, la survie à 35 jours a été déterminée par l'atteinte à la maturité décrite dans la littérature (Solignac, 1978b), tandis que la survie à 70 jours a été déterminée en fonction de l'atteinte à la maturité observée dans cette étude chez les individus backcross (développement retardé des caractères sexuels secondaires).

4.1 Isolement sexuel

Un isolement sexuel fort a été détecté dans les deux régions, visible par le succès des croisements interspécifiques (7 croisements sur 29 ont donnés des juvéniles) diminué par rapport aux

croisements intraspécifiques (35/43) et backcross (26/30). De plus, les croisements interspécifiques ont également montré une plus grande variabilité dans le temps d'obtention des juvéniles que les deux autres types de croisements. Cet isolement est incomplet, et montre une tendance à être asymétrique. Ainsi un plus grand nombre de croisements effectués dans le sens femelle *J. praehirsuta* – mâle *albifrons* ont donnés des juvéniles que dans le sens inverse (respectivement 5/12 vs 2/17).

Ces patterns se retrouvent lorsque l'on regarde les deux régions séparément, bien que l'isolement sexuel semble être un peu moins fort en Normandie (non significatif).

Ainsi, en Bretagne, 19 croisements intraspécifiques sur 21 ont donné en moyenne 5 juvéniles, obtenus en 38 jours (moyenne sur l'ensemble des premières portées). Ces résultats sont semblables à ceux obtenus pour les croisements backcross, puisque les deux croisements backcross réalisés ont donné en moyenne 7 juvéniles en 32 jours (moyenne sur l'ensemble des premières portées). Cependant, la puissance d'analyse est loin d'être la même puisque la taille des échantillons est bien supérieure pour les croisements intraspécifiques que backcross. Un fort isolement comportemental a en revanche été détecté lors des croisements interspécifiques, total dans le sens femelle *J. albifrons* – mâle *J. praehirsuta* (sur 8 croisements tentés), et de 33% (seulement 2 croisements sur 6 ont donnés des juvéniles) dans le sens femelle *J. praehirsuta* – mâle *J. albifrons*. Ces croisements semblent n'avoir donné qu'une seule portée (contre 4 à 6 portées pour les autres croisements, voir Figure 3.8), mais cela est dû à un artefact de manipulation, puisque ces femelles ont été fixées directement après leur première portée.

En Normandie, 16 croisements intraspécifiques sur 22 ont donné en moyenne 9 juvéniles, obtenus en 42 jours (moyenne sur l'ensemble des premières portées). Ces résultats sont également semblables à ceux obtenus pour les croisements backcross, puisque 22 sur 26 croisements backcross réalisés ont donné en moyenne 10 juvéniles en 30 jours (moyenne sur l'ensemble des premières portées). Les croisements interspécifiques dans le sens femelle *J. albifrons* – mâle *J. praehirsuta* montrent étonnement une forte barrière comportementale, seuls 22% des croisements ayant donné des juvéniles. Les croisements interspécifiques femelle *J. praehirsuta* – mâle *J. albifrons* quant à eux montrent une barrière comportementale assouplie (50% des croisements ayant donné des juvéniles).

4.2 Isolement post-zygotique

Un isolement post-zygotique a été détecté mais est faible en Normandie et de manière plus surprenante également en Bretagne. D'une part, aucune différence dans les tailles des portées n'a été détectée entre croisements intraspécifiques, interspécifiques et backcross toutes régions confondues, indiquant l'absence de problème en termes de fécondité, avec une tendance de taille de portées plus grandes pour les croisements backcross. D'autre part, la "survie à maturité" semble

moins élevée pour les individus issus de backcross (45 %) que pour les individus issus de croisements intra- (79 %) et interspécifiques (83 %). Cette observation reste vraie si l'on compare la survie de tous les types de croisement à la date précise de 35 jours (intraspécifique 79 %, interspécifique 83 %, backcross 64 %). Ces résultats témoignent donc de l'existence d'une barrière d'isolement de type post-zygotique qui se manifeste par une survie légèrement moins bonne des juvéniles produits par des femelles hybrides, et également par un retard de développement des caractères sexuels secondaires des mâles issus de ces mêmes femelles hybrides (bien que ce retard n'est pas été quantifié avec précision lors de ce travail).

Cependant, cet effet n'apparaît seulement qu'à partir des croisements backcross (absence de problème dans la fécondité et la survie des juvéniles issus des croisements interspécifiques), et la mortalité à maturité ne se traduit pas par une grosse différence en nombre de descendants matures car les portées issues de BC étaient légèrement plus grandes (Figure 3.9). Ainsi, le nombre de descendants moyen par portée "à maturité" (soit 35 jours pour les croisements intra- et interspé, et 70 jours pour les BC) était de 6.18 individus issus des croisements intraspécifiques, 7.71 individus issus des croisements interspécifiques, et 4.53 individus issus des croisements backcross.

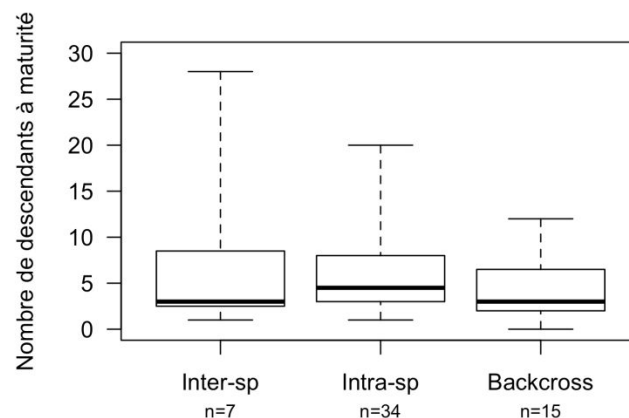


Figure 3.9 : Nombre de descendants à maturité pour chaque type de croisement effectué sur l'ensemble des régions

Une barrière post-zygotique semble donc bien exister, mais semble *a priori* de faible intensité. Nous n'avons pas mesuré l'efficacité de reproduction des individus issus de backcross, si bien qu'il nous manque une composante de l'isolement post-zygotique (isolement post-zygotique extrinsèque lié à la sélection sexuelle). Cependant la faiblesse de la barrière post-zygotique mesurée ici rend improbable la possibilité d'un processus de renforcement.

Isolement asymétrique

En Normandie, les barrières comportementales au niveau des croisements interspécifiques ne sont pas suffisantes pour empêcher totalement l'apparition d'hybrides, et l'absence de barrière écologique en Normandie permet à ces quelques hybrides F1 de pouvoir se reproduire avec des individus « purs » permettant une introgression, visiblement plus forte dans le sens *J. albifrons* vers *J. praehirsuta* (chapitre 1, 2 et 5). Cette asymétrie peut être expliquée au vue des résultats de ce chapitre par une plus forte introgression de *J. albifrons* vers *J. praehirsuta* due à l'asymétrie apparente de la barrière comportementale en F1 interspécifique (les femelles *J. praehirsuta* semblant accepter plus facilement les mâles *J. albifrons* qu'inversement). Il y a donc plus d'individus ayant un génome majoritairement *J. albifrons* qui peuvent se reproduire et ainsi potentiellement passer leur fond génétique que d'individus à génotype majoritairement *J. praehirsuta*. Bien que la fertilité des individus issus de backcross n'ait pas été testée, on ne peut totalement exclure une moins bonne fitness de la descendance hybride. Une réduction de la fitness hybride peut se traduire de différentes manières et notamment par la faible capacité des hybrides à concurrencer les espèces mères lors de la parade sexuelle. Ce point pourrait être abordé dans une étude ultérieure.

Cependant, une légère asymétrie dans l'isolement post-zygotique semble visible, mais ne peut être considérée comme telle de par le faible nombre de croisements effectués. Une étude de Larson et al. (2012) avait montré ce type de résultat, un croisement fécond et viable dans un sens, mais l'apparition d'hybrides non viables dans l'autre sens chez les criquets du genre *Gryllus*. Cette asymétrie semblait être due à un problème de fertilisation de l'œuf (mauvaise utilisation du sperme dans la spermathèque ou incompatibilités sperme-œuf), amenant à des incompatibilités post-copulation pré-zygotiques. L'isolement asymétrique en Normandie est-il uniquement dû à la barrière comportementale ou d'autres barrières viennent t'elles la renforcer? Il serait nécessaire de faire des analyses supplémentaires afin de détecter l'origine de cette asymétrie dans nos croisements.

Enfin, l'asymétrie est régulièrement trouvée puisque (Solignac, 1978b) avait réalisé 19 croisements femelles *J. albifrons* – mâle *J. praehirsuta* qui n'avait donné aucun juvéniles et 46 croisements femelle *J. praehirsuta* – mâle *J. albifrons* dont 10 croisements ayant eu des juvéniles (les individus *J. albifrons* provenant de Kiel – Allemagne, et les *J. praehirsuta* de Roscoff). On pourrait donc supposer que cette asymétrie à évoluée en même temps que la spéciation pour être présente dans des répliques indépendants, à la fois en situations de syntopie (cette étude), de micro-allopatrie (cette étude) et d'alopatrie (Solignac, 1978b).

4.3 Perspectives

Ce chapitre a permis de mettre en lumière certaines barrières impliquées dans l'isolement reproductif.

En Bretagne, nous avons confirmé, comme supposé par Solignac (1978a, 1981) que la principale barrière est de type sexuel, supposé comportemental. Bien qu'il y ait un faible isolement post-zygotique, le fort isolement comportemental, couplé à un isolement de type écologique en milieu naturel (les deux espèces n'occupant essentiellement pas les mêmes habitats), semble suffisant pour maintenir la forte différenciation interspécifique. L'isolement post-zygotique bien qu'existant ne semble pas jouer de rôle important dans le maintien de la différenciation au sein de cette région.

En Normandie la barrière comportementale semble également être la principale barrière à la reproduction. D'autres analyses concernant l'étude la barrière post-zygotique, qui semble ici faible, seraient à prévoir afin d'évaluer son impact sur l'introgression asymétrique.

Nous avons tout au long de ce chapitre, discuté de l'action d'une barrière pré-zygotique de type comportementale, car il est connu au sein de ce complexe d'espèces qu'il s'agit d'une barrière importante, mais nous ne pouvons exclure l'éventuel rôle d'une barrière de type gamétique (post-copulation pré-zygotique). Un exemple est l'étude de Yeates et al. (2013) qui ont montré que les saumons et truites femelles avait un fluide ovarien attirant le sperme de mâles conspécifiques.

Nos résultats montrent un isolement sexuel, que l'on suppose comportemental, ainsi qu'un isolement post-zygotique faible. On peut ainsi citer deux possibilités de sélection contre hybrides, par sélection sexuelle, et/ou par incompatibilités génétiques. Un point intrigant dans cette étude et dans les deux régions, est la segmentation des barrières à la reproduction : une barrière principalement comportementale va agir durant les croisements intraspécifiques et interspécifiques, sans diminution de la fécondité ou survie, alors que la barrière post-zygotique va apparaître en backcross, alors qu'il ne semble plus y avoir de choix de partenaires. Cette barrière post-zygotique pourraient être expliquées par des incompatibilités génétiques de type Dobzhansky–Muller (iDM) (Orr & Turelli, 2001).

Des recherches de telles incompatibilités seraient une piste à suivre dans les études futures. Il serait également nécessaire de quantifier l'intensité des barrières avec des méthodes permettant par la suite de les comparer (Sobel & Chen, 2014). Cette quantification est en cours, sur de plus grandes tailles d'échantillons (nouveaux croisements réalisés).

Une autre perspective concerne le rôle possible d'une toute autre barrière, liée l'action de bactéries endosymbiontes de type *Wolbachia*. En effet, ces bactéries peuvent participer à un isolement reproductif (Telschow et al., 2005) sous forme d'incompatibilités cytoplasmiques (Bordenstein et al., 2001), et pourraient être liées à l'asymétrie discutée plus haut. Ce thème fait l'objet du chapitre suivant.

Enfin, des réarrangements chromosomiques peuvent également être à la base de l'isolement reproductif (Coyne & Orr, 2004). Ainsi, si les gènes en lien avec l'isolement reproductif se situent sur des régions chromosomiques réarrangées entre taxa, le peu de recombinaison ne va pas éliminer ces gènes, maintenant la divergence entre espèces lors de contact secondaire (Noor et al., 2001, Rieseberg, 2001, Navarro & Barton, 2003, Feder & Nosil, 2009). Aucun génome n'est disponible pour nos espèces d'études, cependant dans cette étude, de nombreux croisements intra-spécifiques ont été obtenus et ont pu être utilisés pour créer des cartes de liaison (Gonen et al., 2014, Kakioka et al., 2013) afin de vérifier s'il existe des réarrangements chromosomiques au sein des populations dont les individus ont été utilisés ici. L'analyse de ces cartes génétiques est présentée dans le chapitre 6.

Chapitre 4

**LA BACTERIE *WOLBACHIA*
IMPLIQUEE DANS L'ISOLEMENT
REPRODUCTIF ?**

Dans ce chapitre seront posées les questions de la présence de bactéries endosymbiotes du genre *Wolbachia* et de leur implication potentielle dans l'isolement reproductif entre *J. albifrons* et *J. praehirsuta* fluctuant en fonction des zones géographiques d'échantillonnage. En effet, cette bactérie est entre autres connue pour féminiser certains de ses hôtes, ainsi que pour agir comme barrière à la reproduction entre espèces *via* des incompatibilités cytoplasmiques. Deux approches seront menées ici, l'étude du biais de sex-ratio et la recherche de cette bactérie à l'aide d'outils moléculaires.

Les analyses de ce chapitre ont été menées en partie pendant les stages de licence d'Arnaud Dano et de master 1 de Julia Centanni, et en collaboration avec Claire Daguin-Thiébaud, Marius Wenzel, Stuart Piertney et Richard Cordaux.

1. Introduction

Wolbachia est une alpha-Protéobactérie intracellulaire présente dans le cytoplasme des cellules de ses hôtes et se transmettant principalement de manière verticale *via* les gamètes femelles (Hertig & Wolbach, 1924, Werren et al., 2008), bien que plusieurs études démontrent une transmission également horizontale entre espèces, permettant sa propagation aux différents clades d'invertébrés (Vavre et al., 1999, Cordaux et al., 2001).

Wolbachia est retrouvée chez certains nématodes (Bandi et al., 1998, Taylor et al., 2005) et plus particulièrement chez les arthropodes, infectant les acariens (Breeuwer & Jacobs, 1996), les crustacés (Bouchon et al., 1998) et les insectes (Werren & Windsor, 2000). La prévalence de *Wolbachia* est difficilement estimable, mais elle a été trouvée dans environ 20% des espèces d'insectes (Werren & Windsor, 2000) et selon une étude plus récente (Zug & Hammerstein, 2012), chez 40% des arthropodes terrestres. Une étude récente (Sazama et al., 2017) basée sur la recherche de *Wolbachia* chez des insectes aquatiques et terrestres estime à 50 % la prévalence de *Wolbachia* chez les insectes aquatiques étudiés et à 60 % chez les insectes terrestres. Peu d'espèces de crustacés marins semblent être infectés (Bouchon et al., 1998), mais une étude plus récente a montré la propagation de *Wolbachia* a des crustacés marins amphipodes tels que les puces de sable (*Talitrus saltator*), des isopodes du Genre *Sphaeroma* ou encore des cirripèdes comme l'anatife (*Lepas anatifera*) (voir Cordaux et al., 2012).

Cette forte prévalence fait considérer *Wolbachia* comme l'une des plus grandes pandémies de l'histoire de la vie (Werren et al., 2008). Afin de favoriser sa transmission, cette bactérie va intervenir dans le processus de reproduction de ses hôtes, induisant dans certains cas des incompatibilités cytoplasmiques interspécifiques. Cela se traduit par exemple en une incompatibilité entre un spermatozoïde infecté par une souche *Wolbachia* et un ovocyte non infecté ou infecté par une autre souche, en produisant des descendants non viables (Rousset et al., 1992, Werren, 1997, Werren et

al., 2008). *Wolbachia* peut ainsi jouer un rôle important dans la spéciation (Telschow et al., 2007). Pour exemple, des études sur des guêpes du genre *Nasonia* (Bordenstein & Werren, 1998, Bordenstein et al., 2001, Bordenstein & Werren, 2007) ont montré que *Wolbachia* peut causer l'isolement reproductif entre deux de ces espèces, par incompatibilité bidirectionnelle. De plus *Wolbachia* peut également induire un effet de renforcement de l'isolement comportemental des espèces non infectés chez la drosophile (Jaenike et al., 2006).

D'autres effets sur la reproduction peuvent être induits par *Wolbachia*, tels que l'induction de la parthénogénèse, ou le *male-killing* lorsque le mâle reproductif n'est pas infecté (Hurst et al., 1999). Enfin, *Wolbachia* est surtout connu pour être impliquée dans la transformation de mâles génétiques en néo-femelles phénotypiques fonctionnelles, un processus appelé féminisation (Rigaud et al., 1999, Bouchon et al., 2008, Werren et al., 2008), induisant un biais de sex-ratio dans les populations. Or, l'étude théorique de l'évolution des sex-ratio prédit un état stable lorsque le nombre de mâles et de femelles est équilibré (Fisher, 1999), avec des investissements égaux dans la production de descendants mâles ou femelles nécessaires pour maintenir un ratio de sexe 1: 1 au niveau de la population, et un avantage d'accouplement pour le sexe le moins répandu conduisant à un retour à l'équilibre si un biais se produit. Il existe cependant des contraintes à la régulation de l'équilibre (West & Sheldon, 2002), et il est fréquent d'observer un sex-ratio biaisé à l'état sauvage (Hardy, 2002), pouvant être dû en particulier à la compétition entre individus pour l'accès aux partenaires sexuels (Werren, 1983), une dérive méiotique sur les chromosomes sexuels (Jaenike, 2001) ou encore des distorceurs cytoplasmiques de sex-ratio (Taylor, 1990), tel que *Wolbachia*.

Les espèces du complexe *Jaera albifrons* présentent i) un isolement reproductif variable impliquant dans une certaine mesure des incompatibilités (chapitres 1-3), avec particulièrement un croisement interspécifique femelle *J. albifrons* - mâle *J. praehirsuta* ne donnant jamais de juvéniles et ii) des biais très forts de sex-ratio. Ainsi Solignac (1978b) a souligné une différence entre le sex-ratio secondaire, *i.e.* à la naissance, qui semble équilibré, et le sex-ratio tertiaire, *i.e.* à l'état adulte, où un avantage numérique pour les femelles existe. Selon (Jones & Naylor, 1971a), il existe un biais sexuel important allant jusqu'à 14:1 (F:M) pour les espèces *ischioetosa* en Grande-Bretagne, variant selon les saisons, et plus faible et stable chez *J. praehirsuta* que chez les trois autres espèces de la côte nord de l'Europe. Le déterminisme du sexe semble être purement génétique chez le complexe *Jaera albifrons* (système hétérogamétique femelle ZW) et l'origine des biais de sex-ratio en faveur des femelles est inconnue.

Wolbachia pourrait donc être un facteur d'incompatibilité entre nos deux espèces d'intérêt (*J. albifrons* et *J. praehirsuta*) et induire un biais de sex-ratio. Enfin, *Wolbachia* semble avoir un impact sur les cas de multi-paternité chez *Armadillidium vulgare* (Valette et al., 2017), hypothèse

intéressante dans notre cas puisque quelques cas de multi-paternité ont été retrouvés lors de croisements expérimentaux entre espèces du complexe *Jaera albifrons* (Pabion, communication personnelle). Enfin, la propagation de *Wolbachia* peut conduire à une propagation des haplotypes mitochondriaux associés (balayage sélectif), en modifiant les fréquences des haplotypes mitochondriaux des populations hôtes sans avoir d'effets sur la distribution de la variation génétique nucléaire (Shoemaker et al., 2004, Ritter et al., 2013). Selon Mifsud (2011) et S. Piertney (communication personnelle) *Wolbachia* pourrait ainsi expliquer la faible différenciation mitochondriale chez nos espèces malgré un fond génétique nucléaire hétérogène (Ribardière et al., 2017), et la présence d'un haplotype mitochondrial largement partagé (Mifsud, 2011). En effet, les résultats des analyses de la phylogénie chez les espèces du complexe *Jaera albifrons* basée sur l'ADN mitochondrial obtenus par D. Mifsud, M. Wenzel et moi-même concordent quant au fait que les clades obtenus sont polyphylétiques. Cependant, la forte diversité d'haplotypes ne me paraît pas cohérente avec l'hypothèse d'un balayage sélectif mené par la présence de *Wolbachia*, et d'autres processus tel qu'une sélection sur le génome mitochondrial mériterait d'être pris en compte dans de futures études sur la discordance entre les patterns de différenciation obtenus à partir de marqueurs nucléaires et mitochondriaux (Bonnet et al., 2017).

En résumé, dans le contexte de notre travail sur les barrières à la reproduction entre *J. albifrons* et *J. praehirsuta*, il paraît très important de vérifier la présence de *Wolbachia* dans le complexe d'espèces *Jaera albifrons*, où l'on observe à la fois des forts biais de sex-ratio (en faveur des femelles) en population naturelle, des niveaux d'isolement reproductif variables (impliquant dans certains cas des croisements hétérospécifiques peu féconds), et une discordance nette entre la structure génétique cytoplasmique mitochondriale (homogène) et nucléaire (structurée entre espèces et populations).

Selon Bouchon et al. (1998), il n'y a pas de *Wolbachia* chez *J. albifrons* (3 individus testés), mais ces auteurs reconnaissent que les résultats négatifs pour la détection de *Wolbachia* peuvent être dus à une insuffisance d'échantillonnage. Les objectifs de ce chapitre sont de rechercher la présence de *Wolbachia* afin de déterminer si ce type de bactéries peut jouer un rôle dans l'isolement reproductif des espèces du complexe *Jaera albifrons*. Une première partie du travail sera consacrée à l'étude du sex-ratio, en distinguant sex-ratio secondaire (proportion de mâles au sein de portées élevées au laboratoire) et tertiaire (proportion de mâles au sein de populations naturelles). La seconde partie vise à vérifier la présence de bactéries du genre *Wolbachia* par amplification par PCR de séquences spécifiques.

2. Matériel & Méthodes

2.1 Analyse du sex-ratio chez le complexe *Jaera albifrons*

Les analyses de sex-ratio sont basées sur deux types de données qui ont été comparées : les sex-ratio des individus adultes prélevés en milieu naturel (sex-ratio tertiaire), et le sex-ratio des juvéniles élevés en laboratoire après acquisition des caractères sexuels. Cette seconde valeur donne une estimation du sex-ratio à la naissance (sex-ratio secondaire) étant donnée la faible mortalité des individus en cours d'élevage (cf. résultats).

2.1.1 Sex-ratio des adultes en populations naturelles

Pour pouvoir estimer le sex-ratio nous avons soigneusement échantillonné dans 3 sites connus pour n'héberger qu'une seule espèce par habitat en Bretagne (Lingoz 48°39'12.31"N - 3°57'0.43"W, Pempoul 48°41'16.9"N - 3°57'27.6"W et Pointe-Grande-Grève 48°42'22.6"N - 3°58'18.7"W) tous les individus situés dans plusieurs zones délimitées d'environ 50 cm par 50 cm, les mâles étant plus petits que les femelles et par conséquent plus difficilement détectables. Les espèces *J. albifrons*, *J. ischiosetosa*, et *J. forsmanni* vivant sous les pierres tandis que *J. praehirsuta* étant principalement trouvé sur des espèces d'algues telles que *Fucus vesiculosus* et *Ascophyllum nodosum*, les échantillonnages ont donc été réalisés dans des zones délimitées regroupant des pierres et des algues. Tous les individus ont été ramenés en laboratoire où le sexe a été déterminé, ainsi que l'espèce grâce aux caractères sexuels secondaires des mâles. Des tests binomiaux ont été utilisés afin de détecter un éventuel biais de sex-ratio à l'aide du logiciel R (Core Team, 2016).

2.1.2 Sex-ratio des juvéniles obtenus en laboratoire

Les juvéniles provenaient de deux cadres expérimentaux différents : soit les mères provenaient directement du terrain et les juvéniles étaient récupérés et élevés en laboratoire, soit les naissances faisaient suite à des croisements expérimentaux contrôlés de familles *J. albifrons* ou *J. praehirsuta* (incluant notamment les croisements utilisés pour les chapitres 3 et 6).

Tous les individus ont été élevés individuellement dans des plaques 6 puits avec de l'eau, de petits morceaux d'algues vertes (*Enteromorpha* sp., Bocquet, 1953) et une petite feuille d'orme fournissant alimentation et abri. Les plaques ont été conservées dans des armoires avec thermostat à 17 ° C ainsi qu'un cycle lumière / obscurité de 11h / 13h et l'eau de mer a été changée une fois par semaine (avec des algues et des feuilles au besoin). Les descendants ont été isolés individuellement quelques jours après leur libération du marsupium (espèces à développement direct sans stade larvaire pélagique) et ont été élevés dans les mêmes conditions que les adultes jusqu'à être sexés vers l'âge de 4 à 5 semaines.

Dans l'ensemble, nous avons utilisé des données provenant de 375 portées élevées en laboratoire entre 2013 et 2016. Une grande partie des données (243 portées) provenaient de Normandie, région d'hybridation naturelle entre *J. albifrons* et *J. praehirsuta* (Ribardière et al., 2017) et ont été analysées séparément (voir ci-dessous). L'autre partie des données correspond à 132 portées de quatre espèces, initialement échantillonnées dans la région Bretagne où aucune hybridation n'a été détectée (Ribardière et al., 2017) : *J. albifrons* (n = 72 portées), *J. praehirsuta* (n = 25), *J. ischiosetosa* (n = 12), et *J. forsmani* (n = 23).

Deux analyses ont été effectuées afin d'étudier les conséquences de l'hybridation introgressive sur le sex-ratio chez *J. albifrons* et *J. praehirsuta*. Tout d'abord, nous avons comparé le sex-ratio chez les portées de chaque espèce dans les régions de Bretagne (sans hybridation) et de Normandie (hybridation). Pour cela, nous avons utilisé des portées issues de femelles échantillonnées dans la nature (*J. albifrons* en Bretagne, n = 34 portées, Normandie, n = 22; *J. praehirsuta* en Bretagne, n = 15, Normandie, n = 39). Deuxièmement, nous avons comparé le sex-ratio chez les portées résultant de croisements contrôlés intra et inter-spécifiques. Pour cela, nous avons utilisé des croisements intra-spécifiques de *J. albifrons* (n = 131 portées), des croisements intra-spécifiques de *J. praehirsuta* (n = 134), des hybrides de première génération (n = 17) et des backcross (femelles hybrides F1 croisées avec des mâles *J. albifrons* ou *J. praehirsuta*, n = 58). Des modèles linéaires généralisés (famille binomiale) ont été utilisés pour tester les différences de sex-ratio entre catégories.

2.2 Détection moléculaire de *Wolbachia*

La présence de *Wolbachia* peut être détectée grâce à différents marqueurs génétiques spécifiques (Simoes et al., 2011). Dans cette étude, nous avons testé 3 différents marqueurs afin d'augmenter notre capacité de détection (détails donnés dans le Tableau 4.1). La région du gène *wsp* (*Wolbachia* surface protein) qui code pour une protéine de surface cellulaire est l'une des cibles les plus courantes pour la détection de *Wolbachia*, en particulier pour les analyses de relations phylogénétiques, de par sa grande variabilité (Zhou et al., 1998). Le marqueur ribosomique 16S-2 a également été choisi car il devrait produire un signal plus fort en dépit du risque plus élevé de faux négatifs (Simoes et al., 2011). Enfin, nous avons également utilisé les amorces *FbpA* pour amplifier le gène de la fructose-bisphosphate aldolase, même si le manque de spécificité peut être plus sensible (Simoes et al., 2011). Deux protocoles différents ont d'abord été testés pour le locus *wsp* (*wsp-1*, *wsp-2*), suivi d'un autre protocole, *Long-wsp*, à partir d'amorces *wsp* rallongées de quelques bases afin de ne pas perdre toute spécificité et avec 50 cycles de PCR (Tableau 4.1, protocole suggéré par Marius Wenzel). Après la découverte d'une séquence inédite de *Wolbachia* chez deux individus du complexe *J. albifrons* (cf. résultats), des amorces spécifiques ont été définies et utilisées à la place des amorces *wsp* universelles.

Au total, 817 individus appartenant au complexe *Jaera albifrons*, dont 321 femelles et 496 mâles, ont été testés. Les 5 espèces du complexe *Jaera albifrons* analysées ont été échantillonnées dans 5 populations de Bretagne (France), 4 populations de Normandie (France), 2 populations du Québec (Canada) 1 population des îles Scilly (Angleterre) et 1 population de Tromsø (Norvège) (détails donnés dans le Tableau 4.2). De plus, 37 individus de 6 espèces différentes d'isopodes et amphipodes ont également été analysés (Tableau 4.3). Certains individus ont été testés plusieurs fois avec différentes amorces, amenant le chiffre total d'échantillons testés à 1 595 (nombre excluant les individus testés plusieurs fois avec un même protocole).

L'ADN total (de *Jaera spp.* et potentiellement de *Wolbachia sp.*) des individus testés a soit été obtenu par extraction avec le kit Nucleospin® Tissue (Macherey–Nagel) soit par extraction dite « au sel » selon le protocole suivant : après avoir été entièrement écrasés, les individus ont été incubés dans 500 µl de tampon de lyse (contenant 1% de SDS, 150 mM de NaCl, 1 mM d'EDTA et 20 mM de TRIS) avec 1 mg de protéinase-K (Macherey-Nagel, Duren, Allemagne) à 56 ° C pendant 3h30 sous agitation à 120 tr/min. Afin de détruire les ARNs éventuellement extraits, 0,5 mg de ribonucléase A (Macherey-Nagel, Duren, Allemagne) ont été ajoutés à chaque tube qui ont ensuite été incubés pendant 20 minutes sous agitation à température ambiante. 200 µl de tampon de NaCl (5 M) ont ensuite été ajoutés et les tubes ont été centrifugés à 16 000 g pendant 10 min. Les surnageants ont été récupérés et mélangés à un volume égal d'isopropanol froid, puis conservés pendant une nuit à -20 °C. Après centrifugation à 16 000 g pendant 15 min, le surnageant a été retiré puis 500 µl d'éthanol à 80% a été rajouté. La même opération a été recommencée après centrifugation cette fois avec 70% d'éthanol froid. Enfin tout le volume contenu dans les tubes a été retiré puis l'ADN a été séché à l'air libre avant d'être remis en suspension dans 100 µl de TE 0.1X.

Tableau 4.1. Description des 6 loci testés dans cette étude : nom du locus, référence, noms des amorces et taille du produit PCR attendu.

Marqueur	Reference	Amorce (F/R)	Taille du produit PCR (pb)
<i>wsp-1</i>	Zhou et al., 1998	81F/691R	610
<i>wsp-2</i>	Braig et al., 1998	81F/691R	650
<i>Long-wsp</i>	Jeyaprakash & Hoy, 2000	LongF/LongR	657
<i>Jaera-wsp</i>	Cette étude	JwspF/JwspR	464
<i>16S-2</i>	Werren et al., 2000	Wspecf/Wspcr	438
<i>FbpA</i>	Baldo et al., 2006	FbpA_F1/FbpA_R1	509

Tableau 4.2. Individus testés pour *Wolbachia* par PCR et migration sur gel d'agarose. Le nombre d'individus testés avec plusieurs protocoles est donné entre parenthèses.

		<i>Jaera albifrons</i>	<i>Jaera praehirsuta</i>	<i>Jaera forsmani</i>	<i>Jaera ischiosetosa</i>	<i>Jaera posthirsuta</i>	NA	morphotypes intermédiaires <i>J. albifrons</i> - <i>J. praehirsuta</i>
France - Bretagne (5 sites)								
Femelles	<i>wsp-1</i>	8(8)	8(8)					
	<i>wsp-2</i>	11(9)	8(8)	1	5(2)		12	
	<i>long-wsp</i>	24(24)	24(24)		30(2)			
	<i>16S-2</i>	15(15)	1(1)					
	<i>FbpA</i>	1(1)	15(15)		1			
Mâles	<i>wsp-1</i>	11	9	8(8)	3(3)			
	<i>wsp-2</i>	5			2(2)			
	<i>long-wsp</i>	64(49)	43(27)	16(16)	39(8)			
	<i>Jaera-wsp</i>	49(49)	27(27)	8(8)	8(8)			
	<i>16S-2</i>	6	3					
	<i>FbpA</i>			1(1)				
France - Normandie (2 sites)								
Femelles	<i>wsp-1</i>	15(15)	1(1)					
	<i>wsp-2</i>	16(16)	12(8)				73	
	<i>long-wsp</i>	15(15)	16(16)					
	<i>16S-2</i>	1(1)	15(15)					1
	<i>FbpA</i>	1(1)						
Mâles	<i>wsp-1</i>	8(8)	11(8)					
	<i>long-wsp</i>	163(24)	76(15)	1				23(10)
	<i>16S-2</i>	3(3)						1(1)
	<i>FbpA</i>	13(13)	7(7)					8(8)
Canada - Côte Est (2 sites)								
Femelles	<i>wsp-2</i>						53	
Mâles	<i>wsp-1</i>				5(2)	8(8)		
	<i>wsp-2</i>				2(2)			
	<i>long-wsp</i>	2(2)				8(8)		
	<i>Jaera-wsp</i>	2(2)						
Angleterre - Iles Scilly								
Femelles	<i>wsp-2</i>						8(8)	
	<i>long-wsp</i>						10(8)	
Mâles	<i>long-wsp</i>	7	12					
Norvège (1 site)								
Femelles	<i>wsp-2</i>						43	
Femelles (321)		50	45	1	33	45	191	1
Males (496)		258	146	17	44	8	0	23
Total (817)		308	191	18	77	53	191	24

Tableau 4.3. Espèces d'isopodes marins (*J. hopeana*, *J. nordmani*, *Janira maculosa*) et d'amphipodes marins (*Ampelisca spinipes*, *Haploops nira* et *Sphaeroma serratus*) testés pour *Wolbachia*.

espèces	nombre d'individus
<i>Ampelisca spinipes</i>	8
<i>Haploops nira</i>	8
<i>Sphaeroma serratus</i>	4
<i>Jaera hopeana</i>	8
<i>Jaera nordmani</i>	4
<i>Janira maculosa</i>	5
Total	37

Les conditions de réaction en chaîne de la polymérase (PCR) sont détaillées dans les Tableaux 4.4 et 4.5. L'enzyme utilisée dans les différents mix est la GoTaq pour protocoles *16S-2*, *FbpA*, *wsp-1* et *wsp-2* et la Q5 hot start DNA polymérase (New England Biolabs, MgCl₂ de 2 mM inclus dans le buffer) pour les protocoles *Long-wsp* et *Jaera-wsp*. Cette enzyme a été testée car elle est réputée améliorer l'amplification de très petites quantités d'ADN. Les PCR ont été réalisées par programmation d'un thermocycleur T100 (Bio-RAD). L'ADN d'un cloporte *Oniscus asellus* infecté par *Wolbachia* (fourni par R. Cordaux) a également été utilisé comme témoin positif pour détecter la présence de *Wolbachia*. Les produits de PCR ont été soumis à une électrophorèse sur gel d'agarose 1,5% colorée au bromure d'éthidium et visualisée sous lumière UV.

45 individus ont été testés avec le protocole *16S-2*, 47 avec *FbpA*, 95 avec *wsp-1*, 251 avec *wsp-2*, 610 avec *Long-wsp*, et 45 avec *Jaera-wsp*. Les 37 individus d'espèces autres que celles appartenant au complexe *Jaera albifrons* ont tous été testés avec le protocole *Long-wsp*.

Au total, 75 produits de PCR (12 produits de PCR avec *wsp-2*, 62 produits de PCR avec *Long-wsp*, et le contrôle positif), ont été envoyés pour séquençage direct chez Eurofins Genomics (Ebersberg, Allemagne). Les séquences obtenues ont été assemblées et vérifiées avec le logiciel Codon Code Aligner (Codon Code corporation, Centerville, MA) puis comparées à la base de données Genbank Nucleotide en utilisant le logiciel «BLAST» (Basic Local Alignment Search Tool).

Des ré-amplifications de produit de PCR ont ensuite été réalisées afin d'augmenter le niveau de seuil de détection et détecter tout produit d'amplification très faible de la première PCR. Ainsi les produits de PCR de 95 individus amplifiés une première fois avec *wsp-1* et 177 avec *wsp-2* ont été utilisés pour une seconde PCR avec *wsp-2*. 161 individus amplifiés une première fois avec *Long-wsp* ont également été ré-amplifiés avec *Long-wsp* et enfin 20 individus amplifiés avec *Long-wsp* ont été ré-

amplifiés avec les amorces spécifiques *Jaera-wsp*. 12 produits de ré-amplification *wsp-2/wsp-2*, 12 produits de ré-amplification *Long-wsp/ Jaera-wsp* et 13 produits de ré-amplification *Long-wsp/Long-wsp* ont été séquencés puis comparés à la base de données Genbank.

Tableau 4.4. Composition du mix de PCR pour chacune des cinq paires d’amorces testées.

Nom	MgCl ₂ (mM)	amorces (μ M)	dNTPs (mM)	Taq (U)	ADN (μ l)	Volume final (μ l)
<i>wsp-1</i>	2.5	5	0.25	1	3	20
<i>wsp-2</i>	1.5	2.8	0.05	0.75	2	15
<i>Long-wsp</i>		10	0.2	1	0.6	20
<i>Jaera-wsp</i>		10	0.2	1	0.6	20
<i>16S-2</i>	1.5	2.8	0.2	1	3	20
<i>FbpA</i>	1.5	10	0.2	1	3	20

Tableau 4.5. Détail des programmes de la PCR; T: Température (°C); D: Durée (sec). Un programme “touchdown” a été utilisé pour l’amorce *16S-2*, consistant à réduire graduellement la température d’hybridation.

Nom	Dénaturation initiale		Dénaturation		Hybridation		Elongation		Nombre de cycles	Dénaturation finale	
	T	D	T	D	T	D	T	D		T	D
<i>wsp-1</i>	94	120	94	60	55	60	72	60	35	72	600
<i>wsp-2</i>	94	180	94	30	55	30	72	60	35	72	600
<i>Long-wsp</i>	98	30	98	15	55	20	72	60	50	72	600
<i>Jaera-wsp</i>	98	30	98	15	54	20	72	60	35	72	600
<i>16S-2</i>	95	120	95	120	60	60	72	60	2	NA	NA
			95	30	60	60	72	45	35	72	300
<i>FbpA</i>	94	120	94	30	59	45	72	90	36	72	600

3. Résultats

3.1 Sex-ratio des adultes provenant de populations naturelles

Entre 64 et 115 individus adultes ont été échantillonnés dans chacune des quatre populations choisies pour les estimations du sex-ratio. Trois espèces sur quatre présentaient un sex-ratio biaisé (Tableau 4.6 et Figure 4.1), *J. ischiosetosa* présentant le biais le plus fort (20 mâles, 94 femelles, test binomial, p -value <0.001). En revanche, *J. praehirsuta* n'a révélé aucun biais (30 mâles, 34 femelles, $p = 0.71$). Tous les mâles prélevés dans un échantillon (combinaison site / habitat) appartenaient à une seule espèce, à l'exception de trois mâles *J. praehirsuta* qui ont été trouvés avec un échantillon de *J. albifrons* (c'est-à-dire sous les roches). Ces trois mâles ont été éliminés de toutes les analyses, mais il existe une faible possibilité que certaines des femelles utilisées pour estimer le sex-ratio de *J. albifrons* soient en réalité des *J. praehirsuta* (voir la discussion).

Tableau 4.6. Individus du complexe *Jaera albifrons* échantillonnés en populations naturelles. Les biais de sex-ratio (proportion de mâles) ont été testés par test binomial au sein de chaque site (** $p < 0.01$, *** $p < 0.001$).

Espèces	Site	Habitat	Mâles	Femelles	Sex ratio
<i>J. albifrons</i>	Lingoz	pierres	15	49	0.23***
<i>J. praehirsuta</i>	Lingoz	algues	30	34	0.47
<i>J. ischiosetosa</i>	Pempoul	pierres	20	94	0.18***
<i>J. forsmanni</i>	Pointe Grande Grève	pierres	41	74	0.36**

3.2 Sex-ratio chez les juvéniles élevés en laboratoire

A partir de 132 portées provenant de quatre espèces initialement échantillonnées en Bretagne, 1 349 descendants ont été élevés (Tableau 4.7). La survie moyenne en laboratoire (c'est-à-dire de la naissance à la date de détermination du sexe) était de 77,5%, ce qui signifie que 1 046 individus ont pu être sexés avec succès pour les quatre espèces. Nous avons constaté un fort biais en faveur des femelles chez les portées *J. ischiosetosa* (sex-ratio=0.29). Les trois autres espèces n'ont pas montré de biais de sex-ratio significatif (Tableau 4.7 et Figure 4.1). Dans le cas de *J. ischiosetosa*, nous avons élevé 130 juvéniles provenant de 12 portées avec un taux de survie de 85%. Avec 20 décès sur 130 individus, le biais de sex-ratio ne peut-être expliqué par la mortalité de certains juvéniles. Même si tous les individus morts étaient des mâles, alors le sex-ratio à la naissance serait encore fortement biaisé (32 mâles, 78 femelles, test binomial, p -value = 0,029). Le même raisonnement s'applique aux quatre espèces et nous considérons donc le biais de sex-ratio en laboratoire comme une bonne approximation du sex-ratio secondaire (sex-ratio à la naissance, Figure 4.1).

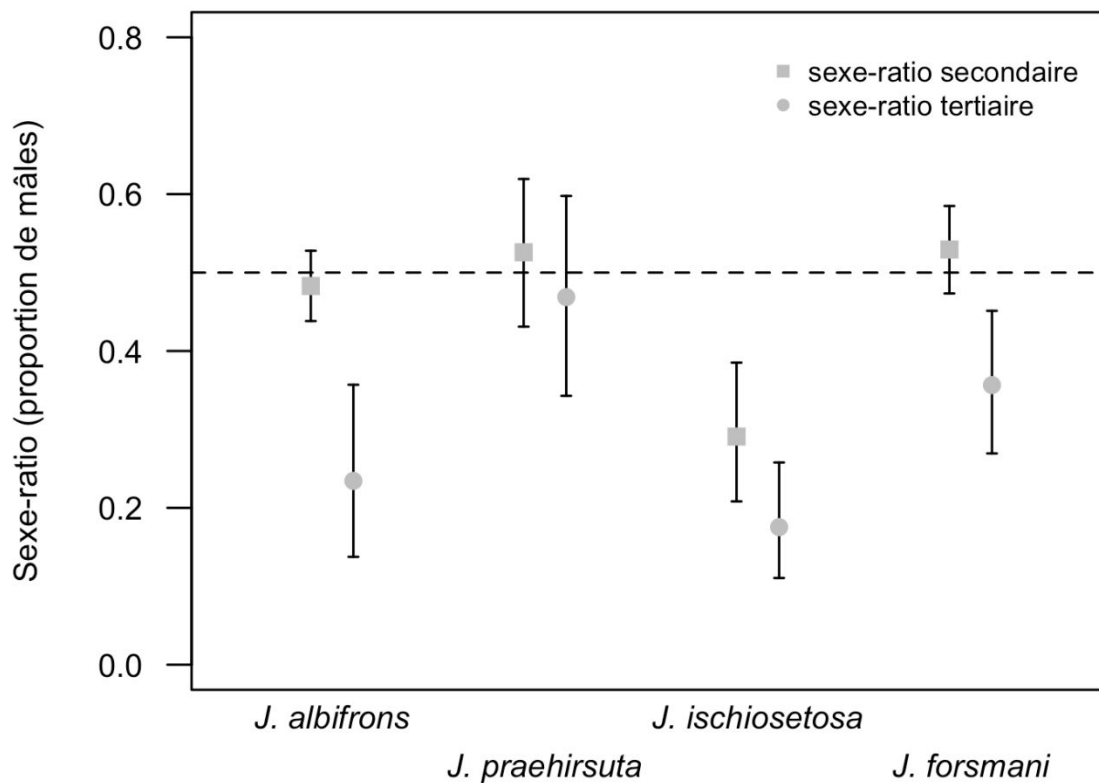


Figure 4.1 : Sex-ratio (proportion de mâles) au sein des quatre espèces européennes du complexe *Jaera albifrons*. Les barres représentent les intervalles de confiance à 95%. Toutes les espèces à l'exception de *J. praehirsuta* montrent un biais significatif envers les femelles au stade adulte en populations naturelles (tailles d'échantillons dans les Tableaux 4.6 et 4.7).

Tableau 4.7 : Nombre d'individus des quatre espèces du complexe *J. albifrons* obtenus et élevés en laboratoire, ainsi que le nombre de mâles et de femelles utilisés dans le calcul du sex-ratio (proportion de mâles) testés avec un test binomial au sein de chaque site (***) $p < 0.001$.

	<i>n</i>	mâles	femelles	sex-ratio
<i>J. albifrons</i>	655	240	257	0.48
<i>J. praehirsuta</i>	211	61	55	0.53
<i>J. ischiosetosa</i>	130	32	78	0.29***
<i>J. forsmanni</i>	353	171	152	0.53

Sexe-ratio dans le cas d'hybridation introgressive entre J. albifrons et J. praehirsuta

Pour cette analyse, nous avons séparé les analyses pour la région de Bretagne (sans hybridation) et celle de Normandie (avec hybridation introgressive), et sur 798 individus (sur 1070 juvéniles isolés) nous n'avons trouvé aucun biais de sexe-ratio significatif au sein de chaque espèce dans chaque région (tests binomiaux, p -values > 0.05) et aucune différence significative entre les espèces et les régions (GLM, p -value > 0.05 pour les deux facteurs et leur interaction). Nous avons abouti à la même conclusion lors de l'analyse des descendants des croisements expérimentaux avec *J. albifrons*, *J. praehirsuta*, les hybrides F1 et les hybrides issus de backcross (Figure 4.2, 1902 individus sexés sur 2744 juvéniles élevés). Encore une fois, nous n'avons trouvé aucun biais de sexe-ratio significatif au sein de chaque classe (test binomiaux, p -values > 0.05) et aucune différence significative entre les classes (GLM, p -value > 0.05).

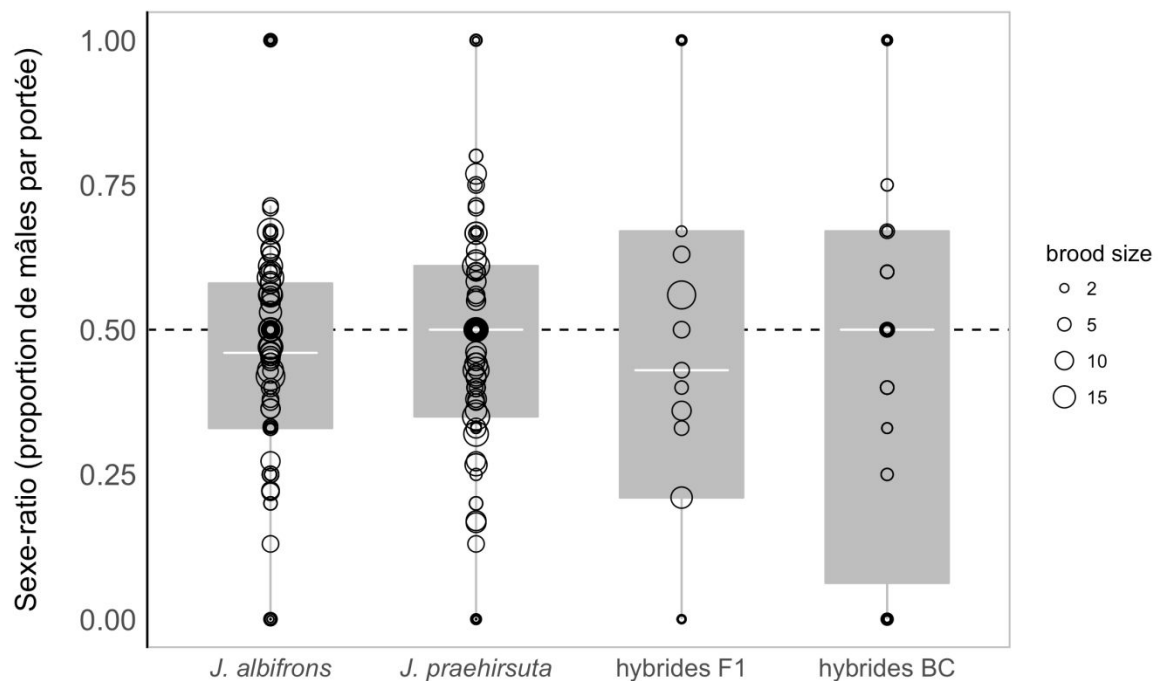


Figure 4.2 : Distribution du sexe-ratio chez les descendants des élevages issus de croisements expérimentaux (parents connus) de *Jaera albifrons* ($n=131$ portées), *J. praehirsuta* ($n=134$), hybrides de première génération ($n=17$), et hybrides issus de croisements backcross (femelles hybrides de première génération croisées avec un mâle *J. albifrons* ou *J. praehirsuta*, $n = 58$).

3.3 Détection moléculaire de *Wolbachia*

Une vaste recherche par amplification de gènes présents dans le génome de la bactérie endosymbionte *Wolbachia* au sein du complexe d'espèces *Jaera albifrons* a été menée sur 1 595 tests impliquant 817 individus, dont 417 ont été testés entre 2 et 17 fois et parmi lesquels 163 ont été testés avec au moins 2 amorces différentes.

- protocoles *wsp-1*, *wsp-2*, *FbpA*, *16S-2* : absence de détection

Au locus *FbpA*, aucune amplification n'a été détectée pour aucun des individus testés, et au locus *16S-2*, seuls 2 individus ont montrés une amplification, de taille moitié moins grande que celle attendue pour *Wolbachia*. Le protocole *wsp-1*, a permis quant à lui de détecter quatre individus avec un signal d'amplification très faible à environ 300 pb et deux individus avec deux bandes distinctes à 300 pb et 600 pb faibles également (Figure 4.3a). Avec le protocole *wsp-2*, 11% des individus ont montré une amplification faible.

- protocole de ré-amplification à partir de produits de PCR : absence de détection

Une ré-amplification des produits de PCR de 95 individus amplifiés avec *wsp-1* et 177 individus amplifiés avec *wsp-2* a donc été réalisée, ce qui a révélé un signal pour 43% des échantillons amplifiés, de 300 pb jusqu'à 650 pb, dont 8% montrant deux bandes. Après cette double amplification, 11 produits de PCR ré-amplifiés *wsp-2/wsp-2* montrant une bande forte et unique autour de la taille attendue 600-650 pb, plus le contrôle positif ont été séquencés directement. Aucune des 11 séquences n'a correspondu à une séquence de *Wolbachia* dans Genbank. 4 séquences n'étaient pas suffisamment bonnes pour être informatives, 6 séquences ne correspondaient à aucune séquence connue, et une séquence correspondait à une séquence d'une bactérie marine, *Rhodopirellula baltica*. La séquence obtenue avec notre contrôle positif correspondait bien à une séquence de *Wolbachia* (numéro d'accès AJ276601.1).

- protocole *Long-wsp* : détection de 2 individus positifs à *Wolbachia* (0.33% de réussite)

Un nouveau protocole a donc été mis en place avec des amorces *Long-wsp* plus sensibles, testé sur 610 individus. Parmi les nombreuses amplifications détectées, dont la plupart montraient plusieurs bandes pour un même individu (voir exemple de gel, Figure 4.3b, ligne du haut) ce qui exclut la possibilité de les séquencer par séquençage direct, 62 produits de PCR montrant des bandes uniques avec tailles d'amplifications proches de celle attendue ont été envoyés à séquencer (Tableau 4.8). Parmi les séquences obtenues, deux d'entre elles correspondaient à une séquence de *Wolbachia*, inédite. En effet, cette séquence est identique à 99% à des séquences de *wsp* de *Wolbachia* publiées dans Genbank (3 mutations sur 540 nucléotides). Par la suite, et malgré plusieurs tentatives avec une quinzaine d'amplifications et le séquençage des deux individus concernés, cette séquence n'a jamais pu être retrouvée par amplification simple (mais voir plus loin).

- protocole de ré-amplification avec *Long-wsp* : pas de détection

Des produits de pcr de 161 individus obtenus avec *Long-wsp* ont été de nouveau amplifiés avec ces mêmes amorces (voir exemple de gel, Figure 4.3b, partie de droite). Les produits PCR de ré-amplification de 13 de ces individus ont été séquencés sans trace de *Wolbachia*. La ré-amplification de ces mêmes produits de PCR avec les amorces universelles permettent d'amplifier d'autres souches de bactéries, notamment des Flavobactéries.

- protocole avec amorces spécifiques *Jaera-wsp* : pas de détection

De nouvelles amorces ont donc été définies sur la base de la séquence *Wolbachia* préalablement trouvée chez nos 2 individus positifs à cette bactérie, et 94 individus ont été testés avec, ne donnant aucun signe d'amplification, même pour les deux individus supposés infectés par *Wolbachia*.

- protocole de ré-amplification des PCR *Long-wsp* avec les amorces spécifiques *Jaera-wsp* : détection de 8 individus (40% de réussite)

En revanche l'utilisation de ces nouvelles amorces pour ré-amplifier 20 individus préalablement amplifiés avec l'amorce *Long-wsp* a finalement permis de détecter 12 individus montrant des bandes fortes et à la taille attendue (voir exemple de gel, Figure 4.3b, parties de gauche et du centre).

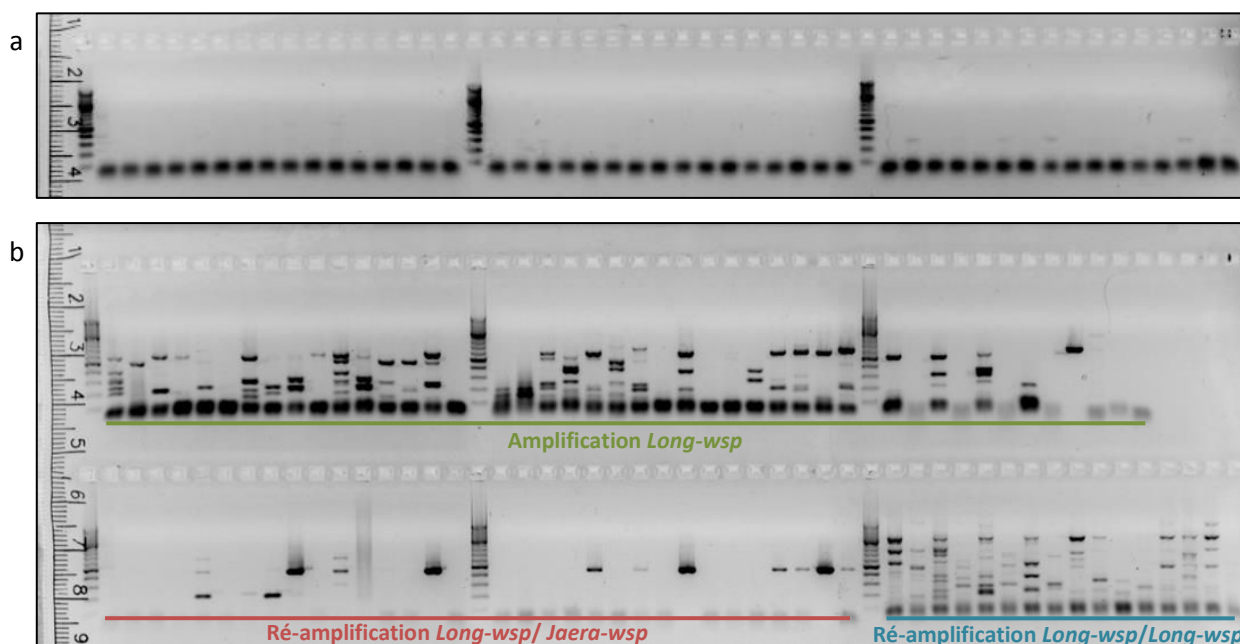


Figure 4.3 : Electrophorèse des produits d'amplification avec a) une amplification directe par *wsp-1*, b) une amplification par *Long-wsp* (en haut), une ré-amplification par *Long-wsp* puis *Jaera-wsp* (en bas, à gauche et au centre), une ré-amplification *Long-wsp/Long-wsp* (en bas à droite).

Parmi ces 12 individus se trouvaient les 2 individus préalablement détectés comme ayant *Wolbachia*. La séquence de *wsp* spécifique de *Jaera* a été retrouvée chez 8 individus (dont nos deux individus positifs). Les produits PCR des 4 autres individus ont été détectés comme trop faibles pour être correctement séquencés. Avec cette méthode, 40% des 20 individus testés (choisis car ils

présentaient une bande de taille attendue) se révélèrent positifs à *Wolbachia*, tandis qu'avec la méthode *Long-wsp*, seulement 0.3 % des individus ont été amplifiés pour *Wolbachia* (Figure 4.4).

Les 8 individus positifs à *Wolbachia* sont toutes des femelles : 1 femelle que l'on suppose *J. praeirsuta* car trouvée sur les algues, à Lingoz (Bretagne, chapitre 2 - site 12, Tableau 2.1 et Figure 2.1), 4 *J. albifrons* et 2 *J. praeirsuta* de Sainte-Honorine (Normandie, chapitre 2 - site 32, Tableau 2.1 et Figure 2.1) et 1 femelle non déterminée ayant donnée des juvéniles à morphotype intermédiaire entre *J. albifrons* et *J. praeirsuta* (voir Tableau 4.8). Un de ses juvéniles a également été testé pour *Wolbachia* avec le protocole de ré-amplification *Long-wsp/Long-wsp*, et c'est une séquence de Flavobactérie qui a été obtenue.

Sur l'ensemble des séquences obtenues certaines correspondaient à des Flavobactéries, à des alpha-Protéobactéries autres que *Wolbachia*, ou bien à des séquences inconnues de la banque de données. Des individus positifs à *Wolbachia* en utilisant le protocole de ré-amplification spécifique ont parfois été trouvés également positifs à une flavobactérie ou à une alpha-Protéobactérie autres que *Wolbachia* en utilisant des protocoles différents de la ré-amplification spécifique.

Afin de comparer notre séquence de *Wolbachia* spécifique de *Jaera* avec les séquences publiées, une phylogénie a été construite avec quelques séquences de *Wolbachia* connues en supplément de celle trouvée dans cette étude (Figure 4.5). Cela montre que cette souche de *Wolbachia* appartiendrait au Super-groupe A, groupe RUG (Figure 4.5, Cordaux et al., 2001), qui contient les endosymbiontes de type *Wolbachia* infectant les arthropodes marins (Cordaux et al., 2012).

Pour tous les tests réalisés, les contrôles positifs (*i.e.* les échantillons de cloportes terrestres infectés par *Wolbachia*) et les contrôles négatifs (eau) ont toujours été respectivement positifs (bande de la bonne taille et dont la séquence correspond bien à l'attendu) et négatifs (absence totale d'amplification). Certains réplicas ont montré des résultats différents pour un même individu testé en fonction de l'amorce utilisée lors de l'électrophorèse (absence ou présence de bande) et/ou du séquençage, *e.g.* détection de Flavobactéries ou de *Wolbachia* en fonction de l'amorce utilisée.

Enfin, 1 seul individu sur l'ensemble des individus des 6 espèces autres que *Jaera albifrons* montrait une bande sur gel semblant exploitable, mais le séquençage de celle-ci n'a donné aucun résultat.

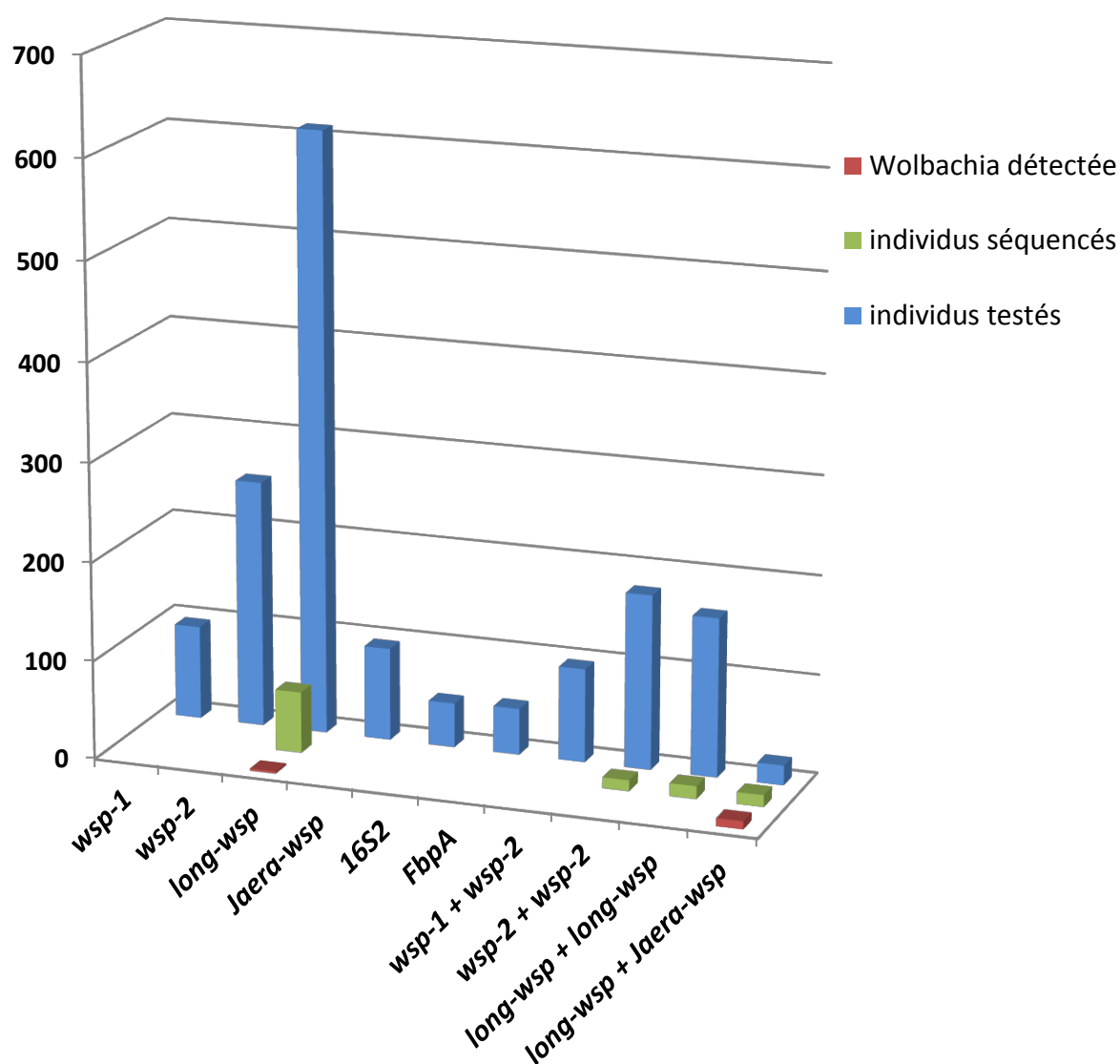


Figure 4.4 : Nombre d'individus testés (en bleu), séquencés (en vert) ou pour lesquels *Wolbachia* a été détectée (en rouge), en fonction des différents protocoles et amorces utilisés.

Tableau 4.8. Individus séquencés en fonction de l'espèce et de la région. Le nombre d'individus détectés pour *Wolbachia* est donné entre parenthèses.

	<i>Jaera albifrons</i>	<i>Jaera prae-hirsuta</i>	<i>Jaera forsmanni</i>	<i>Jaera ischiosetosa</i>	NA	Hybrides <i>J. albifrons</i> - <i>J. prae-hirsuta</i>	<i>Jaera hopeana</i>	<i>Sphaeroma serratus</i>
France – Bretagne (5 sites)								
Femelles		3 (1)		7	1(1)			
Mâles	2	8	1	8				
NA							3	1
France – Normandie (2 sites)								
Femelles	4(4)	3(2)			1	1		
Mâles	12	3	1			2		
Canada – Côte Est (2 sites)								
Femelles					3			
Angleterre – Iles Scilly								
Femelles					3			
Mâles	1							
Norvège (1 site)								
Femelles					6			
Femelles (32)	4	6	0	7	14	1	0	0
Males (38)	15	11	2	8	0	2	0	0
NA (4)	0	0	0	0	0	0	3	1
Total (74)	19	17	2	15	14	3	3	1

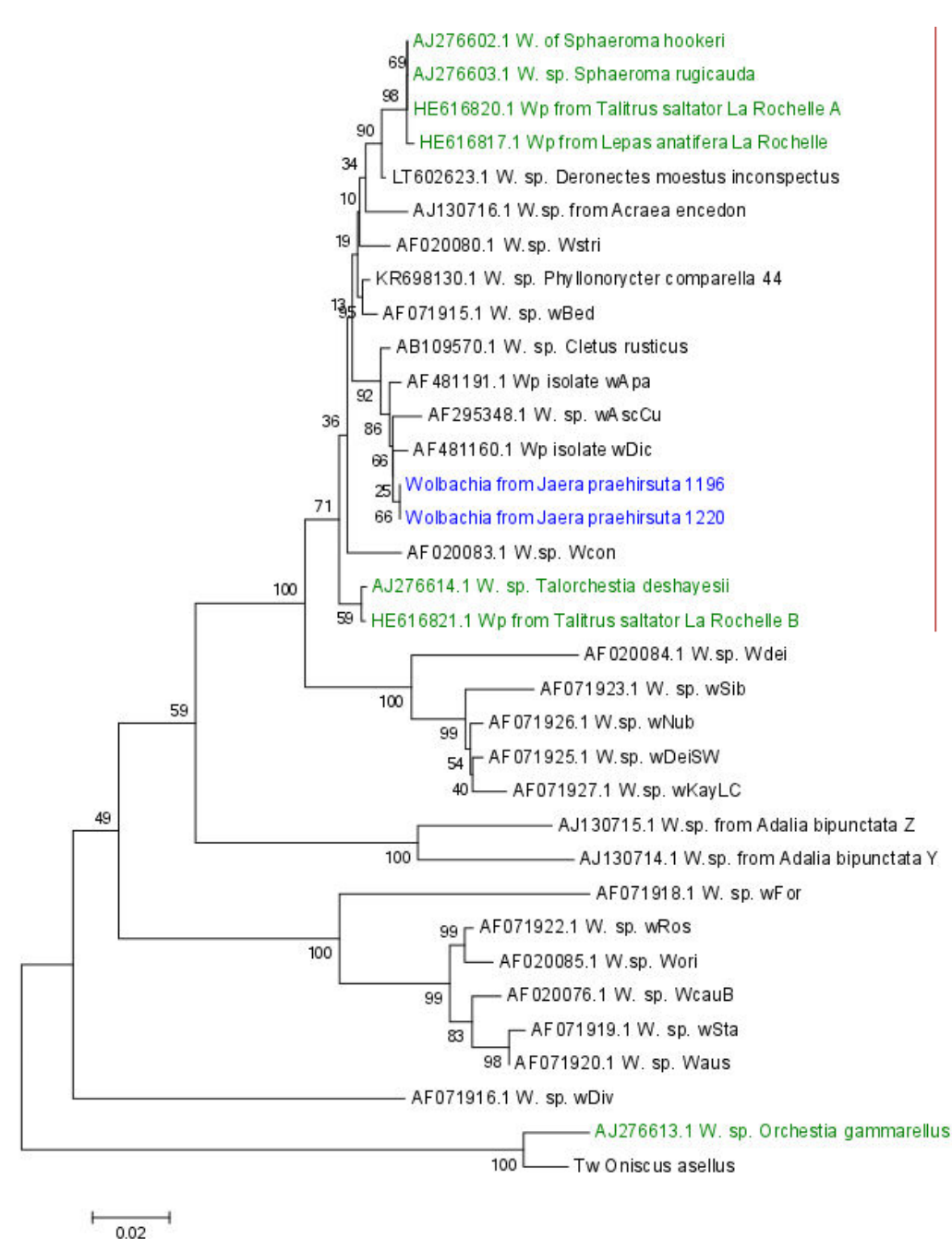


Figure 4.5 : Arbre construit par la méthode du neighbor-joining à partir des distances de kimura 2P, avec quelques séquences de *Wolbachia* du Super-groupe A, dont certaines infectant des espèces de crustacés marins (en vert), et la séquence de nos deux premiers individus positifs à la présence de *Wolbachia*. Les séquences obtenues pour les 6 autres individus *Jaera* positifs sont strictement identiques à celle représentée dans cet arbre au sein du groupe RUG.

4. Discussion

Le travail présenté dans ce chapitre avait pour objectif de rechercher la présence de l'alpha-Protéobactérie endosymbionte *Wolbachia*, à travers des analyses de détection moléculaire. Parallèlement à ce travail, nous avons mené une étude du sex-ratio au sein des quatre espèces Européennes du complexe *Jaera albifrons*, avec une attention particulière pour la paire d'espèces qui est l'objet de ce travail de thèse. Les résultats nous ont permis de définir une méthode pour détecter *Wolbachia* au sein du complexe *Jaera albifrons*, et de détecter un biais de sex-ratio intrigant chez une des espèces du complexe.

Dans cette étude, 1 595 échantillons ont été testés à partir de 10 protocoles distincts impliquant 5 amorces différentes. Étonnamment le protocole *wsp* (ici *wsp-1* et *wsp-2*) couramment utilisé pour la détection de *Wolbachia* (Braig et al., 1998, Zhou et al., 1998, Baldo & Werren, 2007) n'a pas donné de résultats positifs chez nos espèces. L'utilisation de l'amorce *wsp* par *Long PCR* (Jeyaprakash & Hoy, 2000) a permis de détecter 2 individus positifs à la présence de *Wolbachia*, mais ce résultat n'est jamais pu être répété par amplification simple malgré un grand nombre de tests d'amplification et de séquençage sur ces individus (quinze et dix-sept fois chacun). De plus, de nombreuses bandes d'amplification étaient obtenues par individu sur le gel d'électrophorèse, empêchant le séquençage direct de la bande de poids moléculaire attendu. Une sélection des bandes par découpage et purification sur le gel ou par clonage aurait pu être réalisée mais au vu du nombre d'individus, nous avons préféré développer une paire d'amorces spécifiques à la séquence de *Wolbachia* obtenue avec nos échantillons.

Finalement, ce n'est qu'en ré-amplifiant les produits de PCR issu du protocole *Long-wsp* avec ces amorces spécifiques de la séquence *wsp* trouvée chez *Jaera*, qu'il nous a été possible de détecter de façon répétable la présence d'une souche de *Wolbachia* chez 8 individus. La ré-amplification de ces mêmes produits de PCR avec les amorces universelles permettent d'amplifier d'autres souches de bactéries, notamment des Flavobactéries.

D'après les résultats des tests d'amplifications, nous pouvons donc conclure que certains individus des espèces *Jaera albifrons* et *Jaera praeahirsuta* sont bien infectés par *Wolbachia*, sans toutefois pouvoir définir une prévalence de cette bactérie au sein du complexe *Jaera albifrons* dans cette étude. Pour cela il faudrait utiliser ce protocole de ré-amplification par *Jaera-wsp* sur l'ensemble des individus testés avec les différents protocoles dans cette étude (travail en cours). De plus, seules des femelles ont été testées avec le protocole de ré-amplification *Long-wsp/ Jaera-wsp*. Il faudrait maintenant tester des mâles afin de savoir si l'absence de détection de *Wolbachia* est due à l'utilisation du mauvais protocole, ou à un phénomène biologique.

Par la suite, l'utilisation d'une enzyme de restriction spécifique coupant le produit d'amplification *wsp* de *Jaera sp.*, mais pas celui d'autres souches de bactéries permettrait d'obtenir des résultats

conclusifs sans devoir séquencer tous les échantillons, si nous sommes certains qu'une seule souche de *Wolbachia* est présente au sein des individus du complexe *Jaera albifrons*.

Il faudrait également essayer d'adapter le protocole afin de ne pas avoir à faire de ré-amplifications, mais pouvoir utiliser l'amorce *Jaera-wsp* directement, en adaptant le protocole à la manière de ce qui a été fait avec *Long-wsp*. Enfin, un autre protocole d'extraction pourrait également être envisagé, favorisant l'amplification concomitante de l'ADN bactérien avec l'ADN eucaryote.

D'après nos résultats, les individus *Jaera albifrons* et *Jaera praeheirsuta* portent la même souche de *Wolbachia* intra-site et inter-régions, semblant donc une quelconque incompatibilité cytoplasmique bidirectionnelle au moins en Normandie, et aucune présence de la bactérie n'a été détectée chez des individus mâles pour le moment.

Cependant, des analyses supplémentaires seront nécessaires avant de pouvoir apporter des conclusions définitives, d'autant plus que i) l'infection de *Wolbachia* peut dépendre des populations étudiées et du génotype de son hôte (Cooper et al., 2017), ii) une même souche a été trouvée en Normandie chez les deux espèces, pouvant être due à un transfert horizontal (Avtzis et al., 2014), permettant par la suite l'hybridation entre ces espèces (Bordenstein et al., 2001), mais une seule des deux espèces a été retrouvée infestée par *Wolbachia* en Bretagne. Les séquences obtenues concordent avec les séquences du groupe RUG qui contient les endosymbiontes de type *Wolbachia* infectant les arthropodes marins. Il est possible que la souche de *Wolbachia* présente chez *Jaera sp.*, provienne d'un transfert horizontal d'un autre arthropode marin par contact (Rigaud & Juchault, 1995), d'autant plus que certaines souches du Super-groupe A semblent se transmettre facilement et régulièrement par voie horizontale (Werren et al., 1995).

L'hypothèse la plus probable est que nos individus soient infectés de différents endosymbiontes, dont parfois *Wolbachia* en plus faible quantité, rendant sa détection difficile. On suppose donc que la première amplification par *Long-wsp* permettrait d'amplifier toutes les bactéries (y compris celle en faible quantité comme *Wolbachia*), et que seule la ré-amplification par *Jaera-wsp* permettrait ensuite d'amplifier spécifiquement *Wolbachia*. Les autres espèces testées dans ce protocole devront également être testées avec le protocole *Long-wsp/ Jaera-wsp* au cas où la même hypothèse de difficulté d'amplification due à la faible infection par rapport à d'autres bactéries ait lieu chez ces espèces.

La difficulté à obtenir des amplifications dues à *Wolbachia* laisse penser que l'infection est faible, ce qui signifierait qu'elle n'engendre pas d'incompatibilités cytoplasmiques entre gamètes mâles et femelles, impactant ni la spéciation ni l'hybridation de nos espèces, mais également qu'elle n'a pas d'effet majeur sur le sex-ratio.

Plusieurs situations avec un biais de sex-ratio ont pourtant été détectées dans cette étude. La première situation, celle de *J. praeheirsuta*, consiste en un sex-ratio secondaire et tertiaire équilibrés.

La seconde consiste en un biais de sex-ratio chez les adultes mais pas à la naissance (*J. albifrons* et *J. forsmani*). Et la dernière consiste en un biais fort envers les femelles chez *J. ischiosetosa*. Très peu de mortalité de juvéniles a eu lieu en laboratoire, et une mortalité plus forte chez les mâles ne pourrait pas expliquer le fort biais envers les femelles chez *J. ischiosetosa*, ce qui signifie que le sex-ratio à la naissance est réellement biaisé chez cette espèce. Un raisonnement similaire pour les autres espèces amène à la conclusion que le sex-ratio secondaire n'est pas biaisé chez les trois autres espèces européennes du complexe *Jaera albifrons*. L'effet potentiel d'une mauvaise détermination des femelles (e.g. femelles *J. albifrons* prélevées sur des algues et donc identifiées comme *J. praehirsuta*) n'a donc aucun impact dans cette étude. Les femelles n'ayant pas une mortalité supérieure à celle des mâles, la règle de Haldane ne s'applique donc pas ici. Aucune trace de barrière post-zygotique sexe-spécifique n'a donc été détectée au cours de ce chapitre.

En conclusion, la présence de *Wolbachia*, trouvée chez certains individus *J. albifrons* et *J. praehirsuta* si la prévalence s'avère plus élevée que trouvée dans cette étude, n'aurait pas d'effet de féminisation au sein de ces deux espèces. En revanche, cela n'empêche pas la possible infection de *J. ischiosetosa* par une souche de *Wolbachia* féminisante, affectant le sex-ratio de cette espèce. Une priorité serait donc de ré-analyser ces individus avec le protocole de ré-amplification par amorces spécifiques. D'autres symbiotes pourraient également être impliqués dans le biais de sex-ratio chez *J. ischiosetosa*, telles que les *Rickettsia*, trouvés chez certains individus du complexe *Jaera albifrons* par M. Wenzel (communication personnelle), et une analyse à envisager serait de traiter les individus de cette population à l'aide d'un large spectre de traitement antibiotiques, afin de voir si il y a un rétablissement du sex-ratio à la génération suivante.

Une autre possibilité est l'influence de l'environnement sur le sex-ratio. Solignac (1976) avait ainsi montré qu'un biais de sex-ratio pour les femelles était commun dans la nature en hiver et au printemps, suggérant que les mâles toléraient moins bien les basses températures. Il insistait également sur le fait que les femelles sont plus facilement détectables dans la nature. C'est un fait que nous avons contrôlé en redoublant d'attention lors des échantillonnages qui ont été réalisés spécifiquement pour cette étude. De plus, nous avons gardé tous les individus en laboratoire dans des armoires thermostatées à 17°C pour ne pas biaiser le sex-ratio par de trop basses températures. Cependant, nos échantillonnages ont été effectués fin mars et de nouveaux échantillonnages de ces populations devraient être faits en été pour vérifier ce biais a priori dû aux saisons. Aucun biais possible en revanche ne pouvait être induit par l'environnement en laboratoire, tous les individus étant soumis aux mêmes paramètres de contrôle, et nous avons pourtant trouvé des individus *J. ischiosetosa* avec un biais à la naissance.

Piertney et Carvalho (1996) avaient également trouvé un biais de sex-ratio allant jusqu'à 12: 1 en faveur des femelles sur une trentaine de populations de *J. albifrons*, biais qu'ils n'avaient pas

retrouvé chez les juvéniles élevés en laboratoire. Leur conclusion était que le sex-ratio de ces espèces ne serait pas inné mais influencé par des facteurs écologiques. Dans notre cas, seuls quelques sites ont été échantillonnés, ce qui n'est pas suffisant pour détecter un quelconque facteur écologique biaisant le sex-ratio.

Bien d'autres facteurs peuvent cependant induire un biais de sex-ratio, tels que des interactions épistatiques. Par exemple, chez une espèce d'isopode marin *Paracerceis sculpta*, il y a une interaction génétique entre deux facteurs, un gène autosomal et un facteur extrachromosomal, qui, lorsqu'ils sont associés avec un locus lié au phénotype peuvent causer un biais de sex-ratio (Shuster & Sassaman, 1997).

Finalement, il pourrait y avoir un biais du à la présence de distorceurs de sex-ratio, *i.e.* éléments génétiques augmentant leur probabilité de transmission en biaisant la ségrégation mendélienne durant la méiose (Buckler et al., 1999). Les biais de sex-ratio primaires ayant lieu chez des systèmes hétérogamétiques pourraient être ainsi dus à une dérive méiotique altérant le sexe de l'hôte, les chromosomes sexuels étant particulièrement vulnérables à la dérive méiotique (Jaenike, 2001, Beukeboom & Perrin, 2014, p. 118). Afin de vérifier cette hypothèse il serait nécessaire de refaire des croisements chez *J. ischiosetosa* afin d'observer si les individus de seconde génération ne produisent que des femelles (Hardy, 2002).

Chapitre 5

**HISTOIRE DÉMOGRAPHIQUE SOUS-
JACENTE À LA DIVERGENCE ENTRE
J. ALBIFRONS ET *J. PRAEHIRSUTA***

Comprendre le contexte biogéographique de la différenciation entre deux espèces par la reconstitution de l'histoire de la divergence est une étape nécessaire dans l'étude de la spéciation et la recherche des barrières à la reproduction. Dans ce chapitre, par analyses de spectres joints de fréquences alléliques, différents scénarios de divergence seront testés afin de déterminer quel scénario semble expliquer le mieux la différenciation entre *Jaera albifrons* et *J. praeirsuta*, en Bretagne et en Normandie. Ces résultats ont été obtenus en collaboration avec Pierre-Alexandre Gagnaire, et les scripts utilisés sont visibles en annexe 3.

1. Introduction

Le contexte biogéographique (cf. Introduction Générale) dans lequel deux espèces ont divergé est un aspect important (et très débattu) de la spéciation. Ce contexte peut être déterminé si l'on parvient à reconstituer l'histoire de la divergence entre espèces en retraçant les processus migratoires et démographiques qui ont pu avoir lieu, avec interruptions ou reprises de flux géniques entre espèces divergentes. Le manque de données historiques peut aujourd'hui être compensé grâce au développement récent des méthodes de séquençage qui permettent de travailler sur une grande quantité de données polymorphes (SNPs, Baird et al., 2008) et qui ouvrent la possibilité d'inférer des histoires passées grâce à des méthodes d'analyse démo-génétiques (Gutenkunst et al., 2009, Excoffier et al., 2013, Roux & Pannell, 2015). En effet, les événements démographiques tels que la croissance, le déclin ou les flux migratoires entre populations, ainsi que les événements d'introgession ou de tri incomplet des lignées influencent l'histoire évolutive des populations (Nadachowska-Brzyska et al., 2013), laissant des empreintes sur le polymorphisme du génome et pouvant être identifiées dans une certaine mesure (Sousa & Hey, 2013).

Différents scénarios de divergence tels que l'isolement strict en allopatrie ou l'isolement avec migration ont été décrits dans la littérature, mais en pratique, des événements démographiques plus complexes avec des périodes alternées de divergence et de remise en contact ont pu avoir lieu lors de la spéciation, ne permettant pas de supposer les conditions initiales de la divergence à partir de la distribution géographique actuelle des espèces (Hewitt, 2011, Bierne et al., 2011). Par exemple, l'étude de Rougemont et al. (2015) sur la différenciation génétique entre lamproie de Planer (*Lampetra planeri*) et lamproie fluviatile (*L. fluviatilis*) illustre bien ce problème : l'observation seule des différents degrés de différenciation génétique entre plusieurs paires de populations ne permet pas de déterminer si la structure génétique inter-spécifique est due à une divergence actuellement en cours avec une réduction progressive des flux de gènes ou à un contact secondaire récent.

Le contexte géographique de la spéciation est particulièrement débattu dans le cadre de la différenciation entre écotypes, *i.e.* populations adaptées à des habitats différents. Ces espèces

proviennent-elles de spéciations écologiques parallèles répétées dans l'espace ou bien ont-elles une origine unique qui a été suivie par une colonisation des habitats par deux (ou plusieurs) espèces déjà différenciées ?

La spéciation parallèle peut se produire à partir de la sélection de variant déjà présents dans le fond génétique ('standing genetic variation') ou des mutations avantageuses *de novo* apparaissant indépendamment et à plusieurs reprises dans différentes populations (Welch & Jiggins, 2014), et étant impliqués dans un processus d'adaptation à des environnements contrastés, permettant l'apparition d'écotypes. Lorsque des traits ayant évolué à plusieurs reprises dans différents réplicas déterminent l'isolement reproductif, l'établissement de ces mutations est alors généralement attribué à l'impact de la sélection naturelle. (Schluter & Nagel, 1995, Johannesson, 2001, Soria-Carrasco et al., 2014). Bien que la différenciation entre écotypes puisse être due à des adaptations parallèles à des environnements différents comme chez le gastéropode *Littorina saxatilis* (see Butlin et al., 2014), Bierne et al. (2013) proposent que la divergence parallèle pourrait également résulter d'une différenciation partagée lors de la spéciation puis maintenue lors du redéploiement spatial des écotypes. Cependant, Ravinet et al. (2016) ont montré dans le cas particulier des littorines un faible niveau de partage génomique interspécifique et suggèrent que la divergence ne serait donc pas due à une divergence génomique partagée mais bien à des adaptations locales parallèles.

En opposition à l'hypothèse de spéciation parallèle, la spéciation avec contact secondaire implique un évènement de divergence unique, suivi par un ou plusieurs évènements de contact entre écotypes (ou espèces divergentes) avec des degrés variables de différenciation génétique dus à une introgression génomique différentielle et dont les résultantes sont des barrières semi-perméables au flux de gènes (Barton & Hewitt, 1985, Harrison & Larson, 2016). Les périodes glaciaires ont notamment entraîné l'installation de beaucoup d'espèces dans des zones de refuge, et auraient ainsi permis la reprise de flux de gènes entre espèces lors de nouvelles rencontres durant les périodes interglaciaires (Bernatchez & Wilson, 1998, Hewitt, 2001, April et al., 2013) suite par exemple au chevauchement des habitats permettant l'apparition de zones hybrides (April et al., 2013).

Les progrès récents dans l'acquisition et le traitement de données génomiques permettent désormais des tests explicites de modèles alternatifs et complexes de divergence en utilisant des approches de coalescence ou de diffusion (Gutenkunst et al., 2009, Roux & Pannell, 2015). L'approche par inférences basées sur le spectre joint de fréquences des variant génétiques, implémentée dans le programme $\delta\alpha\delta i$ (Gutenkunst et al., 2009), permet ainsi de détecter différentes histoires évolutives probables, des scénarios d'allopatrie avec ou sans flux de gènes aux scénarios de contacts secondaires (Tine et al., 2014, Le Moan et al., 2016, Rougemont et al., 2016a).

Utiliser des modèles simples, sans prendre en compte la variation du paramètre de flux génique (Nm) le long du génome peut donner des résultats trompeurs (Roux et al., 2014). En effet, une partie du génome lié à des locus sous sélection, et dont la proportion fluctuera en fonction de la semi-perméabilité du génome, tendra à avoir un taux de migration réduit par rapport au reste du génome. Quelques études commencent donc à prendre en compte l'effet de variations d'intensité de la dérive le long du génome en permettant des réductions locales de la taille effective de la population (Rougeux et al., 2016, Sousa & Hey, 2013, Roux et al., 2016), ainsi que l'effet de migrations (flux génique interspécifique) variables le long du génome, (Roux & Pannell, 2015, Tine et al., 2014, Rougeux et al., 2016, Le Moan et al., 2016, Sousa & Hey, 2013) pour aborder l'impact des barrières génétiques réduisant le taux de migration efficace aux locus neutres liés à des locus sous sélection (Barton & Bengtsson, 1986, Feder & Nosil, 2010).

L'étude de la spéciation entre paires de populations divergentes répliquées occupant différents habitats et dont les schémas de différenciation génomique sont hétérogènes entre les répliques (Gagnaire et al., 2013, Rougemont et al., 2016a, Jones et al., 2012, Deagle et al., 2012, Westram et al., 2014) est intéressante pour estimer et comparer les processus de divergence au sein de ces répliques, afin d'identifier une origine commune où des spéciations parallèles. Suite à l'étude de la différenciation génétique entre lamproie de Planer (*Lampetra planeri*) et lamproie fluviatile (*L. fluviatilis*) abordée plus haut, une étude portant sur l'histoire démographique a été réalisée, distinguant des scénarios de contacts secondaires entre paires d'espèces géographiquement connectées et des scénarios de contacts anciens entre paires d'espèces plus isolées.

La paire d'espèces *Jaera albifrons* et *Jaera praehirsuta* représente un très bon modèle pour étudier l'histoire démographique sous-jacente à la divergence entre ces deux espèces à travers deux répliques indépendants. En effet, nous avons montré préalablement à cette étude que ces deux espèces, retrouvées en sympatrie dans deux zones géographiques distinctes, *i.e.* Bretagne et Normandie, montrent pourtant une forte hétérogénéité de différenciation génomique entre les répliques (Ribardière et al., 2017, cf. également chapitre 2). Cette étude permet donc une analyse intéressante basée sur la comparaison de deux zones de sympatrie avec potentiellement des histoires démographiques différentes entre les espèces étudiées, *i.e.* contact secondaire en Normandie et schéma de type allopatrique, sans flux de gènes actuel en Bretagne.

Dans cette étude, une analyse de spectres joints de fréquences alléliques (AFS) dont les données polymorphiques ont été recueillies sur un grand nombre de locus obtenus par séquençage RAD a été effectuée puis comparée à des représentations simplifiées de différents scénarios évolutifs intégrant des effets ayant pu influencer les processus démographiques tels que la variation du taux de migration interspécifique et de la taille efficace le long du génome. Ces scénarios proposent ainsi une

histoire évolutive possible des populations étudiées en supposant leur histoire démographique à partir de critères évolutifs donnés et pourront être comparées afin de définir s'il s'agit d'une origine commune ou d'une spéciation parallèle, de l'ancienneté de la divergence, et si l'état actuel des flux de gènes est dû à un isolement strict, à un isolement avec migration, à une migration ancestrale ou à un contact secondaire.

2. Matériel et méthodes

2.1 Echantillonnage

L'estimation des histoires démographiques interspécifiques entre *J. albifrons* et *J. praehirsuta* en Bretagne, zone d'isolement reproductif, et en Normandie, zone d'hybridation, a été réalisée par l'obtention pour chacune des régions d'un spectre joint des fréquences alléliques.

En Bretagne, 89 individus mâles comprenant 50 *J. albifrons* et 39 *J. praehirsuta* provenant des sites Bouillenou, St-Jean et Lingoz (Figure 5.1, sites 1, 2 et 3) ont été utilisés, alors qu'en Normandie le jeu final est composé de 78 individus comprenant 40 *J. albifrons* et 38 *J. praehirsuta*, provenant des sites Grandcamp et Ste-Honorine (Figure 5.1, sites 7 et 8).

Ces spectres joints peuvent être orientés par l'ajout d'individus utilisés comme groupe éloigné pour distinguer les allèles dérivés des allèles ancestraux. Nous avons donc décidé d'utiliser l'espèce la plus divergente du complexe *J. albifrons* des côtes européennes (cf. chapitre 2, et avons ainsi retenus 9 individus de l'espèce *J. ischiosetosa* comme outgroup dans cette étude (voir échantillonnage chapitre 2, Tableau 2.1 et Figure 2.1).

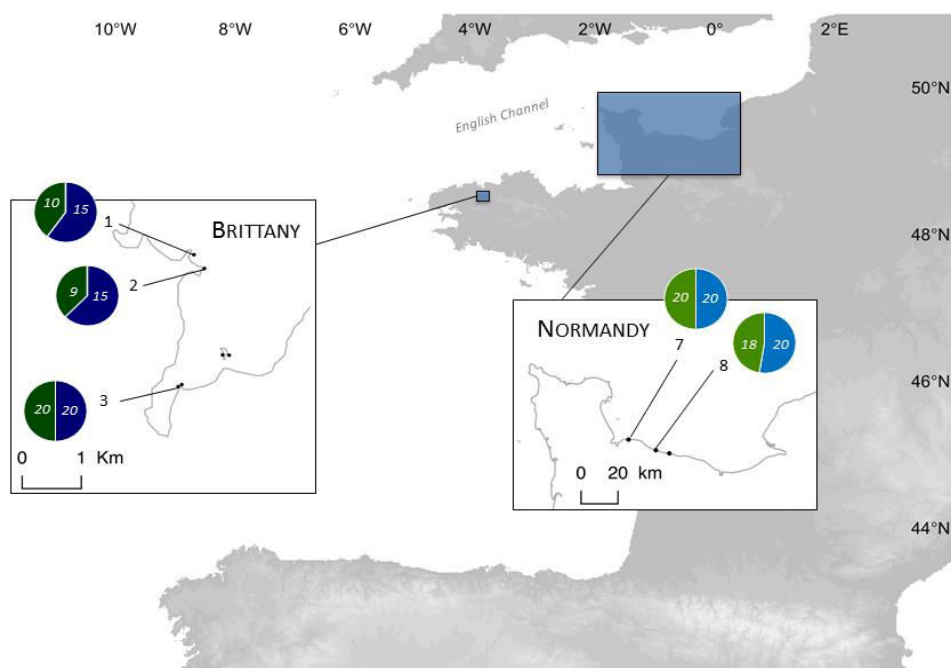


Figure 5.1 : Sites d'échantillonnage en Bretagne et Normandie des individus utilisés pour les inférences démographiques. Les individus *J. albifrons* sont représentés en bleu et les individus *J. praeahirsuta* sont représentés en vert.

Dans ce chapitre le terme population sera utilisé pour désigner une espèce (*J. albifrons*, *J. praeahirsuta* ou *J. ischiosetosa*) dans une région (Bretagne ou Normandie) tous sites confondus.

2.2 Génotypage

L'extraction de l'ADN génomique total de chaque individu a été réalisée sur colonne avec le kit d'extraction Qiagen Tissue Kit, puis un séquençage ddRAD (Peterson et al., 2012) utilisant les enzymes Pst1 et Mse1 suivant le protocole de Brelford et al. (2016) a été réalisé (voir annexe 1), puis un séquençage HS2500 v4 single read sous-traité à Eurofins genomics. Le traitement des données a ensuite été exécuté à l'aide du logiciel Stacks (Catchen et al., 2013), et le pipeline d'analyses suivi a été détaillé Figure 5.2. N'ayant pas de génome de référence disponible, les SNPs (Single Nucleotides Polymorphism) ont été retenus à l'aide d'une approche *de novo*. Nous avons fixé les contraintes suivantes pour retenir un allèle : une profondeur de séquençage minimale m de 3 reads (*i.e.* empilement de 3 séquences identiques créant une pile (*stack*), soit $3X$), une distance maximale M de 3 nucléotides entre deux piles pour pouvoir les assembler en un même locus et un nombre maximum de différences entre individus n fixé à 5 nucléotides pour considérer que deux piles de séquences sont homologues lors de la construction du catalogue.

Après l'étape *populations* de la suite d'analyses (Figure 5.2), 260 145 locus ont été gardés, à raison d'un SNP pris aléatoirement par locus (Tag) afin de limiter l'impact du déséquilibre de liaison. Par la suite les analyses ont été faites séparément en Bretagne et en Normandie mais suivent le même schéma général.

Dans les deux populations de Bretagne (*i.e.* *J. albifrons* et *J. prae-hirsuta*), deux filtres ont été appliqués sous *Vcftools* v0.1.14 (Danecek et al., 2011): 30% de données manquantes maximum ont été autorisées, *i.e.* présence d'un locus chez au moins 70% des individus de la population ; et l'équilibre à Hardy-Weinberg (HWE) a été testé à chaque site, ceux ayant une p-value < 0.01 n'étant pas considérés à l'HWE sont alors retirés des analyses. Concernant les deux populations de Normandie, *Vcftools* a été utilisé pour n'autoriser que 30% de données manquantes maximum, l'équilibre à Hardy-Weinberg n'ayant pas été testé dû aux échanges inter-populationnels avec flux de gènes hétérogène le long du génome qui ont lieu dans cette région (chapitres 2 et 3).

La population de *J. ischiosetosa* a d'abord été analysée séparément, et les SNPs présents dans au moins 5 des 9 individus de la population ont été gardés. A la suite des filtres, et dans chaque région, seuls les locus communs aux populations de *J. albifrons*, de *J. prae-hirsuta* et de l'outgroup *J. ischiosetosa* ont été gardés et réunis en un même jeu de données, soit 5 572 SNPs en Bretagne et 6 623 SNPs en Normandie. Par la suite, seuls les SNPs polymorphes dans au moins une des deux populations et monomorphes chez l'outgroup ont été définitivement gardés, soit 1 913 SNPs en Bretagne, et 1 082 SNPs en Normandie. Parmi ces locus, 380 sont communs aux 2 régions soit 20 à 35% des locus utilisés pour les inférences démographiques.

2.3 Histoire démographique : données

Pour pouvoir obtenir un spectre de fréquences alléliques observées il a d'abord été nécessaire de déterminer le nombre minimum d'observations à un locus donné souhaité.

Sachant qu'un locus doit être génotypé dans au moins 70% de chaque population, et que dans le cas de la Bretagne la plus petite population est celle des *J. prae-hirsuta* avec 39 individus, le côté du spectre doit faire au maximum : $39 \times 70 / 100 = 27.3$, soit en terme de nombre de copies de gènes : $27.3 \times 2 = 54.6$ arrondi à 50 classes possibles. Le même calcul a été fait en Normandie, et le spectre maximal est également contenu dans un cadre de côté 50.

Or, comme expliqué ci-dessus, le nombre de copies utilisé pour faire le spectre peut être diminué pour en améliorer sa résolution, l'important étant de trouver le bon compromis entre résolution et informations obtenues. Dans notre cas, après plusieurs essais le choix de la taille du spectre déplié s'est donc porté à 20x20 que ce soit en Bretagne ou en Normandie.

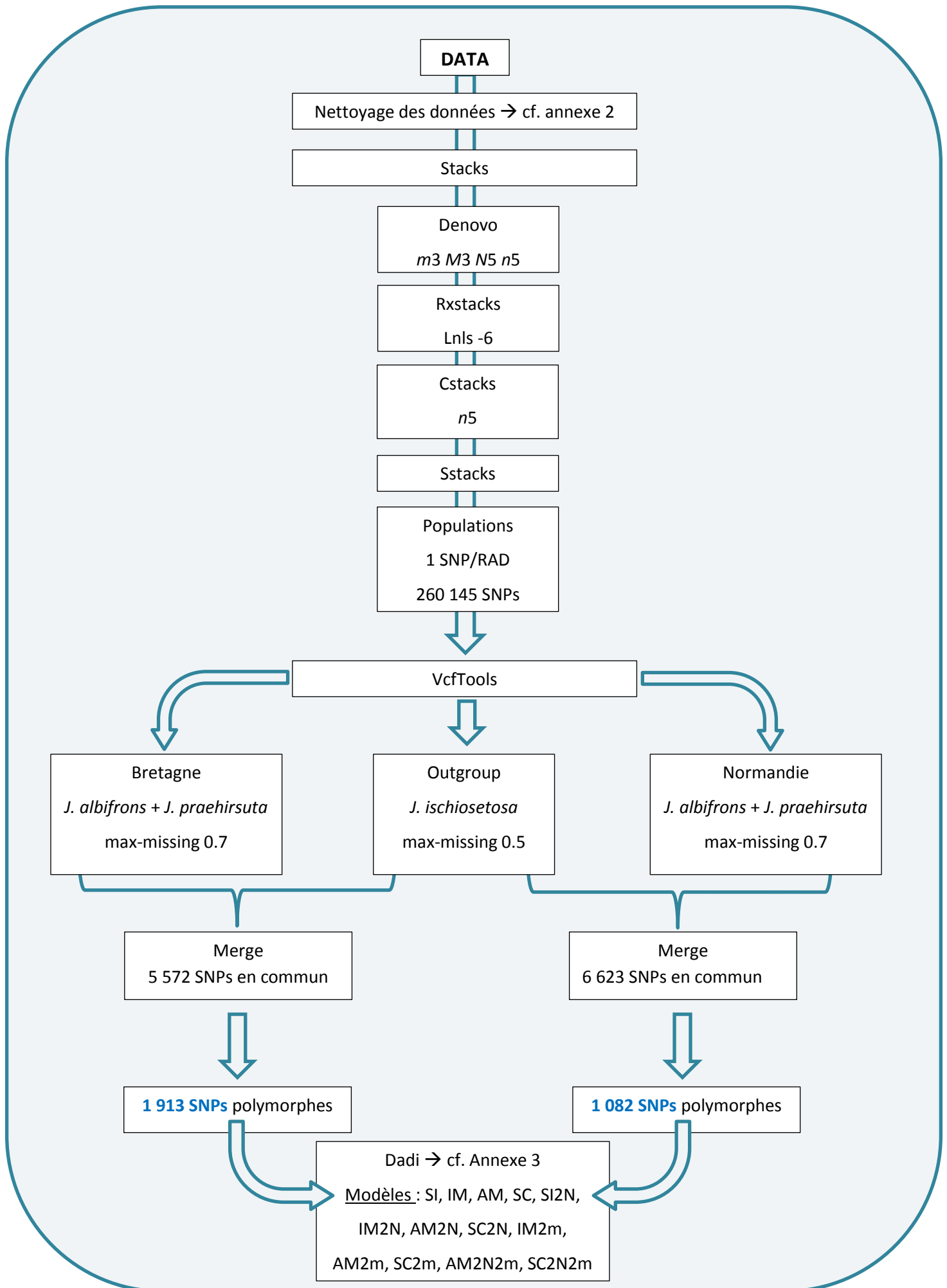


Figure 5.2 : Pipeline de traitement des données avant utilisation de *daði*

Un spectre se lira de la manière suivante : en ordonnée, la population 1, en abscisse la population 2, chaque carré étant défini par le nombre de locus dont l'allèle i correspondant à une mutation dérivée est présent en x_i copie(s) dans la population 1 et y_i copie(s) dans la population 2. Ce nombre de locus est visualisé par une couleur sur une l'échelle allant de 0.1 à 1000 (de l'orange/jaune au violet/rouge), et les zones blanches correspondant à une absence de locus. Des nombres inférieurs à 1 sont permis ici uniquement car le spectre a été réduit et que les valeurs sont donc moyennées sur l'ensemble des individus.

2.4 Histoire démographique : modèles

Le spectre des données observées doit par la suite être comparé aux spectres de données obtenus par approche pseudo analytique pour chaque modèle démographique approximé. La vraisemblance d'un modèle dépendant du nombre de paramètres utilisés, les modèles doivent donc être comparés à partir du critère d'Akaike (AIC) qui pénalise la vraisemblance du modèle en fonction du nombre de paramètres : $AIC=2k-2\ln(L)$ où k = nombre de paramètres, $\ln(L)$ = logarithme (naturel ou base 10) de la valeur de vraisemblance. Le modèle ayant le plus faible AIC sera désigné comme ayant le scénario démo-génétique le plus vraisemblable.

La méthode utilisée est celle implémentée dans Tine *et al.*, 2014 puis améliorée dans Rougeux *et al.*, 2016, et prend en compte l'hétérogénéité du flux génique et l'hétérogénéité de la taille efficace locale le long du génome. Treize scénarios de divergence (Figure 5.3) avec et sans flux de gènes entre espèces et avec et sans ces deux sources d'hétérogénéité ont donc été testés dans cette étude. La vraisemblance de ces 13 modèles, (décrits ci-dessous), a ainsi été statistiquement analysée afin d'estimer le scénario expliquant au mieux l'histoire de la divergence entre *Jaera albifrons* et *Jaera praeheirsuta*.

Les quatre scénarios de divergence à partir desquels sont basés tous les modèles utilisés dans cette étude sont :

- i) le modèle d'isolement strict (SI) entre deux populations divergentes de tailles effectives N_1 et N_2 provenant d'une même population ancestrale de taille N_{ref} (Wakeley & Hey, 1997), *i.e.* un scénario de spéciation allopatrique.

Paramètres de ce modèle :

n_1 : coefficient permettant de calculer la taille efficace de la population 1 ($N_1 = \frac{n_1}{N_{ref}}$)

n_2 : coefficient permettant de calculer la taille efficace de la population 2

t_s : coefficient permettant de calculer l'échelle de temps depuis la divergence ($T_s = t_s \times 2N_{ref}$)

O : probabilité de bonne orientation du spectre

ii) le modèle d'isolement avec migration continue durant la divergence (**IM**)

Paramètres supplémentaires par rapport au modèle précédent :

M_{12} : coefficient permettant de calculer le taux de migration efficace neutre de la population 2 vers la population 1 ($m_{12} = \frac{M_{12}}{2N_{ref}}$)

M_{21} : coefficient permettant de calculer le taux de migration efficace neutre de la population 1 vers la population 2

iii) le modèle dit de migration ancienne (**AM**), se composant d'une divergence entre deux espèces avec flux génique suivi d'un isolement allopatrique avec arrêt des échanges génétiques.

Paramètres supplémentaires :

t_{am} : coefficient permettant de calculer le temps depuis l'arrêt de la migration ancestrale jusqu'à aujourd'hui.

iv) Le modèle de contact secondaire (**SC**), décrivant un isolement allopatrique suivi d'une reprise du flux génique.

Paramètres supplémentaires :

t_{sc} : coefficient permettant de calculer le temps depuis le contact secondaire jusqu'à aujourd'hui.

A ces 4 modèles peuvent être ajoutés des paramètres permettant de prendre en compte de possibles variations d'intensité de dérive génétique et de flux de gènes entre locus le long du génome.

La première source d'hétérogénéité que l'on peut prendre en compte est une variation de la taille efficace locale due à des effets de sélection d'arrière-plan. Deux types de locus, affectés ou non par ces effets indirects, peuvent être définis. En permettant une telle hétérogénéité, quatre nouveaux modèles ont été construits : **SI2N**, **IM2N**, **AM2N**, **SC2N**.

Paramètres supplémentaires :

Hrf : Facteur Hill-Robertsonien permettant de prendre en compte la variation de la taille efficace de la population dans le génome en raison des effets hill-robertsonien (hill et robertson 1966). Ce facteur est appliqué à la taille efficace des locus neutres ($n'_1 = Hrf \times n_1$) et ($n'_2 = Hrf \times n_2$) permettant ainsi de quantifier une réduction locale de la taille efficace due à la sélection aux sites liés, en raison du « background effect » (charlesworth et al., 1993) et d'une sélection positive (smith et haigh, 1974), et représentant une proportion Q du génome.

Une autre source d'hétérogénéité concerne la migration. Jusque-là le flux génique a été considéré comme homogène le long du génome, *i.e.* le paramètre de migration est le même pour tous les locus pour une même direction (m_{12} ou m_{21}). Cependant le taux de migration peut être variable dans le

génomique, donc des paramètres de migration différents peuvent être appliqués à certains locus (m'_{12} et m'_{21}). Ainsi 3 modèles avec migration hétérogène ont été construits : **IM2m**, **AM2m** et **SC2m**.

Paramètres supplémentaires :

m'_{12} et m'_{21} : coefficients permettant de calculer le taux de migration efficace réduite respectivement de la population 2 vers la population 1 ($m'_{12} = \frac{M'_{12}}{2N_{ref}}$), et de la population 1 vers la population 2 ($m'_{21} = \frac{M'_{21}}{2N_{ref}}$), qui correspondent au taux de migration des locus neutres éloignés des barrières d'isolement, ou des locus liés à des mutations avantageuses.

P représente la proportion de locus échangés avec un taux de migration efficace neutre et $1 - P$ est par conséquent la proportion de locus ayant un taux de migration efficace réduit.

Enfin pour les modèles AM et SC ayant deux phases distinctes démographiques, une avec flux génétique et une sans, il est possible d'utiliser la période sans flux génétique pour séparer les effets de la migration et dérive et ainsi d'estimer conjointement ces deux forces évolutives. Ces deux modèles, appelés **AM2N2m** et **SC2N2m** prennent en compte tous les paramètres précédemment cités.

Afin de vérifier la consistance des résultats, chacun de ces 13 modèles a été simulé 20 fois (20 runs) et seul le meilleur run (plus petit AIC, donc modèle ayant la meilleure valeur explicative) de chaque modèle a été analysé en détail.

Les intervalles de confiance des paramètres estimés pour le meilleur des 13 modèles ainsi que pour les autres modèles dont la valeur explicative n'est pas trop éloignée du meilleur modèle ($\Delta AIC < 10$) ont été calculés en réalisant un bootstrap de type conventionnel (1000 réplicas) avec remplacement des SNPs échantillonnés. Les intervalles de confiance à 95% de chaque paramètre ont été obtenus en multipliant les valeurs d'écart-type de la distribution des données rééchantillonnées par 1.96. Les valeurs absolues de taux de migration (symboles), de tailles efficaces de populations (symboles), ou encore la datation de la divergence (symbole) peuvent théoriquement être estimées en calculant la taille de la population ancestrale N_{ref} comme suit :

$N_{ref} = \frac{\theta}{4L\mu}$, avec θ la diversité génétique globale (Gutenkunst et al., 2009), et $L = \frac{zy^{110}}{x}$, avec z = le nombre de SNPs retenus pour les inférences démographiques et x le nombre de SNPs détectés après *Populations* à partir d'un nombre y de RAD tags à 110 pb (Rougeux et al., 2016).

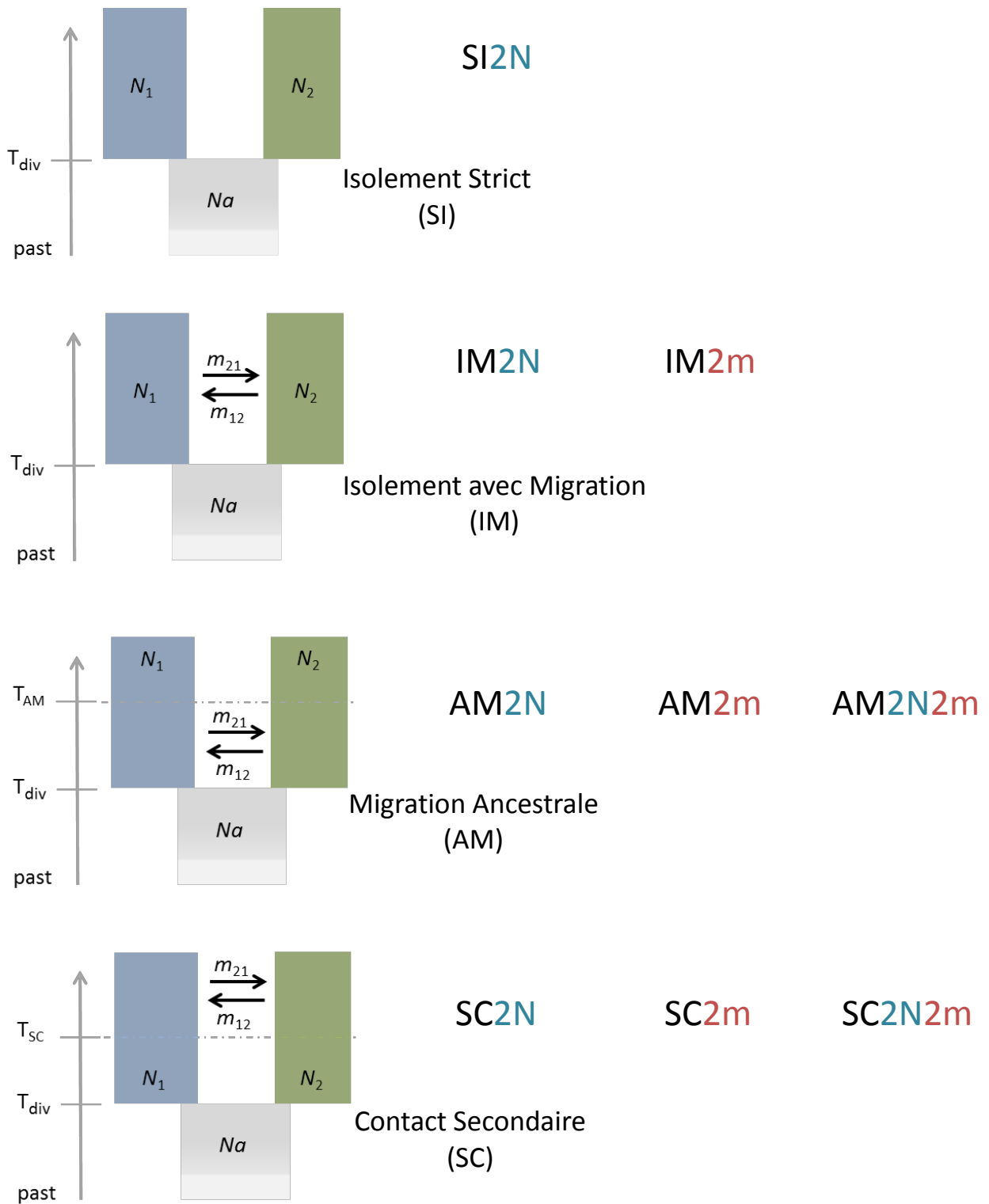


Figure 5.3 : Modèles démographiques simulés, prenant en compte l'hétérogénéité de la taille efficace le long du génome (2N) et/ou l'hétérogénéité du taux de migration entre régions du génome (2m), d'après Rougeux, 2016.

3. Résultats

3.1 Analyse de la diversité génétique au sein de la Bretagne et de la Normandie

Lorsque l'on analyse la distribution du polymorphisme utilisé pour inférer l'histoire de la divergence entre *J. albifrons* et *J. praeheirsuta* dans les deux régions étudiées (Figure 5.4), on s'aperçoit que le nombre total de SNPs polymorphes est presque deux fois plus grand en Bretagne (1913 locus) qu'en Normandie (1082) mais que le polymorphisme n'est pas distribué de la même façon entre espèces au sein de ces deux régions. Il y a notamment plus de polymorphisme partagé (indiqué en rouge dans la Figure 5.4) en Normandie, tandis qu'une très grande partie du polymorphisme Breton est propre à l'une ou l'autre espèce (1541 locus sur 1913). Les différences de polymorphisme entre espèces (ici en termes de nombre absolu de SNPs polymorphes) ne sont pas fortement affectées par les tailles d'échantillons car ces tailles sont relativement proches (50 *J. albifrons* et 39 *J. praeheirsuta* en Bretagne, 40 et 38 en Normandie, sachant qu'un seuil maximum de 30% de données manquantes par espèce et par région a été fixé pour qu'un locus soit retenu dans l'analyse). Les patrons obtenus avec un jeu de données réduit aux locus communs aux deux régions ($n=380$) sont très similaires (données non représentées).

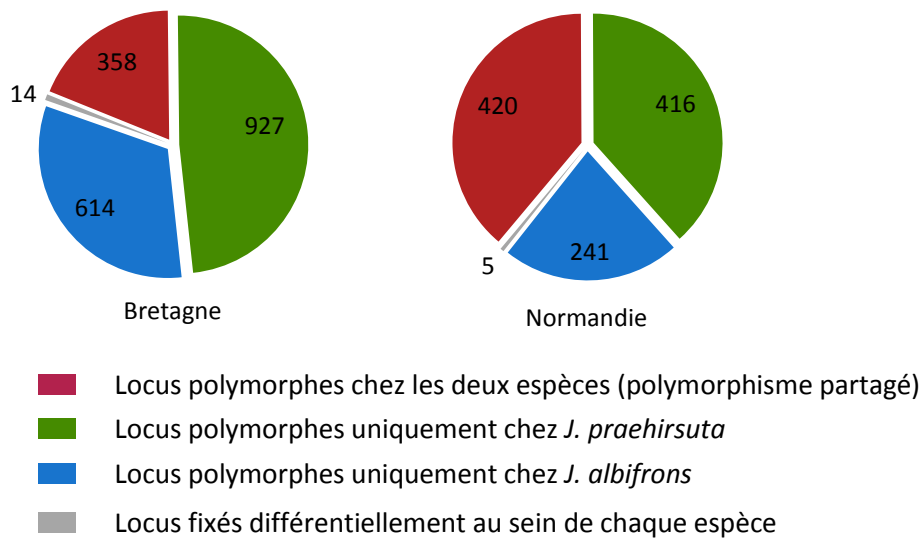


Figure 5.4 : Distribution du polymorphisme utilisé pour inférer l'histoire de la divergence entre *J. albifrons* et *J. praeheirsuta* dans les deux régions étudiées (au total 1913 locus en Bretagne et 1082 en Normandie).

En Normandie, bien que le nombre de locus polymorphes chez au moins une des deux espèces soit presque deux fois plus petit qu'en Bretagne (1913 en Bretagne et 1082 en Normandie), la diversité génétique totale, et la diversité génétique au sein de *J. albifrons* et *J. praeheirsuta* y sont presque 2 fois plus grandes (Tableau 5.1, Figures 5.5 et 5.6). De plus, au sein de chaque région, *J. praeheirsuta* présente une plus grande diversité génétique que *J. albifrons*. Ce même patron est observé lorsque l'on réduit le jeu de données aux 380 locus polymorphes communs aux deux régions.

Tableau 5.1. Distribution de la diversité génétique mesurée avec les SNPs utilisés pour inférer l'histoire de la divergence entre *J. albifrons* et *J. praeheirsuta* au sein des régions Bretagne et Normandie, avec indiqués le nombre de locus utilisés pour calculer la diversité génétique totale (H_T) et la diversité intra-populationnelle (H_S). Toutes les valeurs relatives à la différenciation génétique interspécifique (F_{ST}) sont hautement significatives.

	Bretagne	Normandie	
	Tous sites	Tous sites	Ste-Honorine Grandcamp
nb locus	1913	1082	1082
H_T	0.079	0.129	0.124
$H_{S, albifrons}$	0.056	0.096	0.094
$H_{S, praeheirsuta}$	0.07	0.13	0.121
F_{ST}	0.355	0.226	0.230 ^a

a) cette valeur de F_{ST} entre espèces au sein d'un site est basée sur 859 locus polymorphes

b) cette valeur de F_{ST} entre espèces au sein d'un site est basée sur 781 locus polymorphes

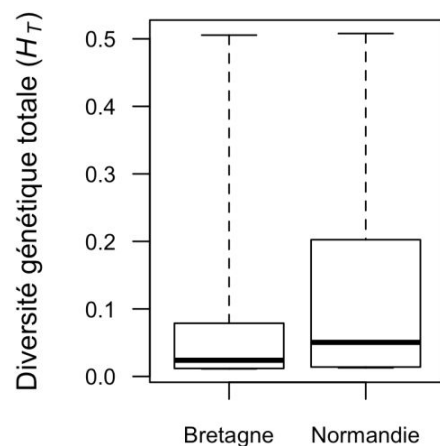


Figure 5.5 : Distribution de la diversité génétique totale H_T au sein des régions Bretagne et Normandie.

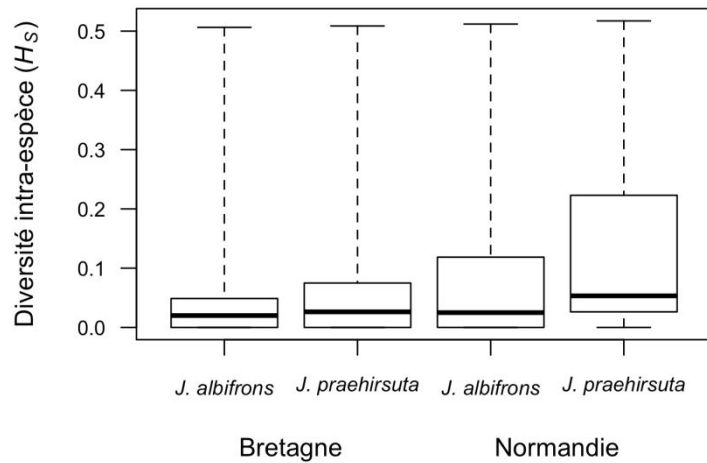


Figure 5.6 : Distribution de la diversité génétique intra-espèce H_S chez *J. albifrons* et *J. praehirsuta* au sein de la Bretagne et de la Normandie.

La différenciation génétique entre les deux espèces semble plus forte en Bretagne qu'en Normandie (F_{ST} respectif de 35% et 23%, Tableau 5.1). La distribution des F_{ST} locus par locus en Bretagne et en Normandie ne montre pas de décalage évident (Figure 5.7A), si ce n'est pour les locus à F_{ST} faible. En effet, en Bretagne 25% des locus ont un $F_{ST} \leq 0$ alors qu'en Normandie cette proportion est égale à 35%. On note également en Bretagne une plus grande proportion de locus ayant un F_{ST} très proche de 1. La distribution de la différenciation entre espèces à partir des locus partagés entre la Bretagne et la Normandie montre des résultats plus tranchés (Figure 5.7B).

La corrélation entre les F_{ST} mesurés sur les locus en commun en Bretagne et en Normandie ($n=380$; Figure 5.8) permet d'observer qu'aucun locus ayant un F_{ST} proche de 0 en Bretagne n'est fortement différencié ($F_{ST} > 0.8$) en Normandie, alors que certains locus fortement différenciés en Bretagne ($F_{ST} > 0.8$) ne montrent pas de différenciation interspécifique en Normandie (F_{ST} proche de 0).

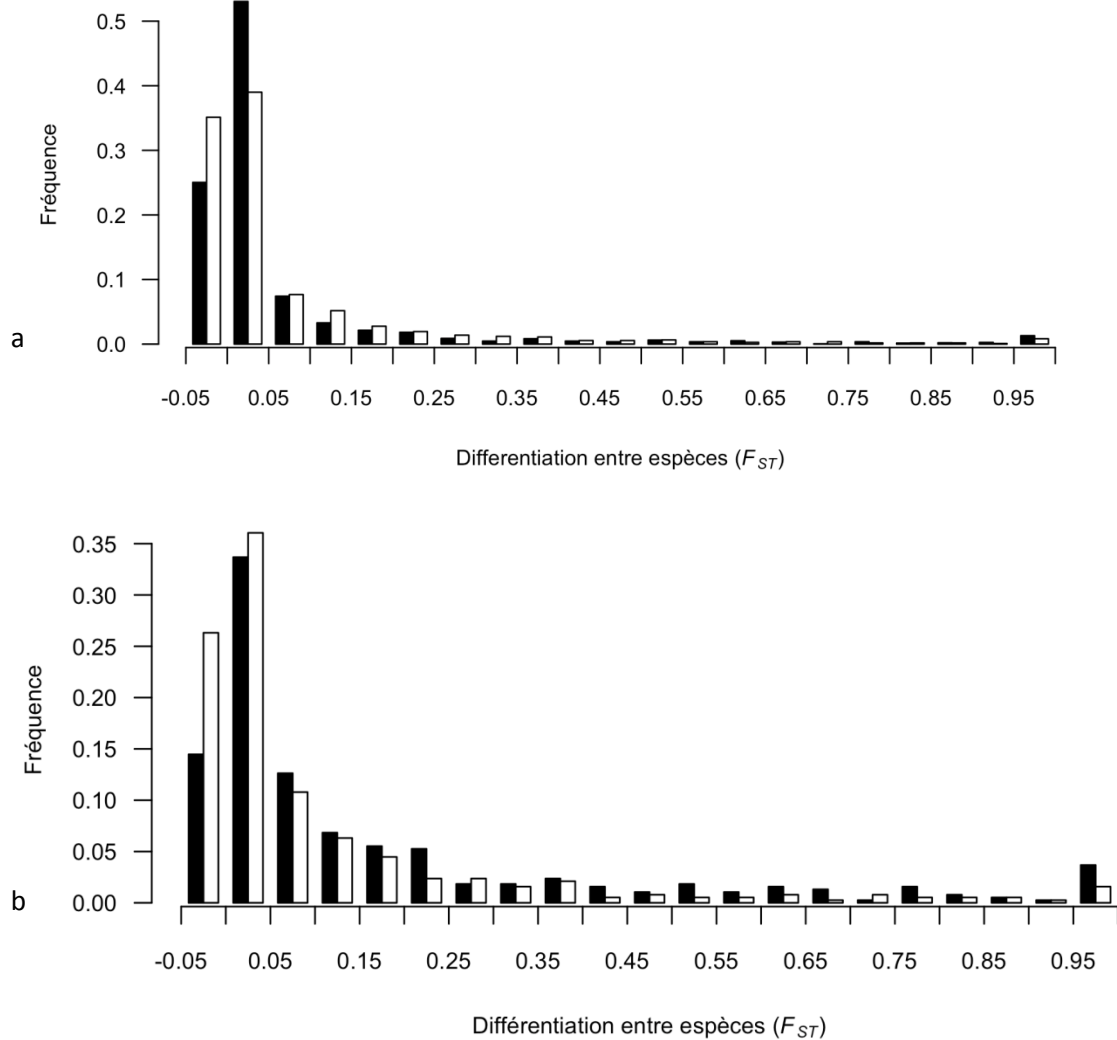


Figure 5.7 : a) Distribution de la différenciation génétique F_{ST} mesurée à chaque locus entre *J. albifrons* et *J. praehirsuta* en Bretagne (en noir, $n= 1913$ SNPs) et en Normandie (en blanc, $n= 1082$ SNPs). L'axe des abscisses est défini par des intervalles de F_{ST} semi-ouverts à gauche (*e.g.* le premier intervalle est $]-0.05;0]$). b) Distribution de la différenciation génétique F_{ST} mesurée à chaque locus en commun entre les deux régions ($n=380$) entre *J. albifrons* et *J. praehirsuta* en Bretagne (en noir) et en Normandie (en blanc).

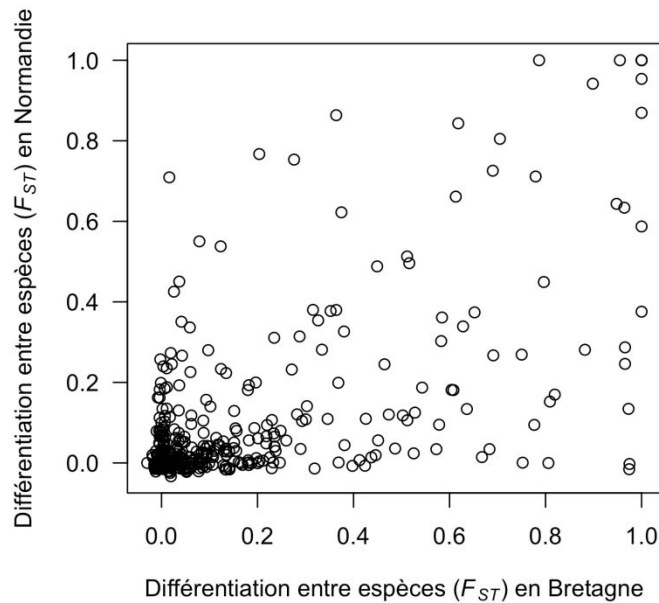


Figure 5.8 : Distribution de la différenciation génétique mesurée à chaque locus ($n=380$) entre *J. albifrons* et *J. praehirsuta* en Bretagne et en Normandie.

3.2 Histoire démographique : données observées

En Bretagne (Figure 5.9), nous pouvons observer que le polymorphisme est principalement distribué sur le cadre du spectre et très peu sur la diagonale, montrant un grand nombre de locus monomorphes chez une espèce alors qu'ils sont polymorphes chez l'autre, et relativement moins de locus dont les allèles sont en fréquences proches entre espèces.

Au contraire, en Normandie (Figure 5.10) les locus se trouvent principalement sur la diagonale, montrant un polymorphisme largement partagé. Ainsi on observe moins de locus dont les allèles sont préférentiellement fixés chez une espèce et variant chez l'autre. De plus, lorsque l'on regarde le cadre du spectre il semble qu'il y a plus de locus ayant des allèles en fréquences intermédiaires chez *J. praehirsuta* mais absents chez *J. albifrons* que de locus ayant des allèles en fréquence intermédiaire chez *J. albifrons* mais absents chez *J. praehirsuta*. Enfin lorsqu'on s'intéresse aux allèles dérivés représentés sur le cadre en haut et à droite, on observe que la majorité des locus ayant des allèles fixés chez *J. albifrons* sont polymorphes chez *J. praehirsuta*.

3.3 Histoire démographique : inférences

Treize scénarios ont été modélisés et comparés à nos données observées. Dans les deux régions, le modèle expliquant le mieux nos données, *i.e.* ayant le meilleur AIC, est le contact secondaire avec dans les deux cas une migration hétérogène le long du génome (SC2m), plus une hétérogénéité de la taille efficace le long du génome en Bretagne uniquement (SC2m2N). L'adéquation entre modèles et données, quantifiée par les calculs d'AIC, peut-être visualisée par la très grande similitude apparente entre les spectres théoriques et observés (Figures 5.9 et 5.10).

En Bretagne, le modèle SC2N2m représente le meilleur scénario avec un Δ AIC de 6 avec le modèle suivant (modèle d'isolement par migration avec hétérogénéité de la taille efficace le long du génome IM2N ; Tableau 5.2), les autres scénarios présentant un Δ AIC de plus de 10 avec le meilleur modèle. En Normandie le modèle SC2m semble expliquer le mieux nos données, avec un Δ AIC de plus de 2 avec le second meilleur modèle. Trois autres modèles pourraient cependant expliquer correctement nos données en Normandie et sont mentionnées ici car ayant tous un Δ AIC < 10 au meilleur modèle : la migration ancestrale et l'isolement avec migration, tous les deux avec un taux de migration hétérogène (AM2m et IM2m), et le contact secondaire avec migration et taille efficace hétérogène (SC2m2N). Bien que données à titre informatif pour l'ensemble des modèles ayant un Δ AIC inférieur à 10 dans le Tableau 5.2, les valeurs des paramètres de chacun des modèles ne seront analysées que pour le modèle ayant le meilleur AIC dans chaque région (*i.e.* le modèle de contact secondaire). En Bretagne, le modèle SC2m bien qu'ayant un Δ AIC de plus de 10 avec le modèle SC2N2m (Tableau 5.2), a été rajouté afin de permettre une comparaison inter-régions basée sur le même modèle (cf. discussion). Les intervalles de confiance donnés suggèrent une relativement bonne estimation des valeurs obtenues. On observe que dans le cas de la Normandie, l'intervalle de confiance sur la durée du contact secondaire n'exclue pas 0, ce qui équivaudrait à un isolement avec migration. Cependant, la distribution des résidus du modèle IM2m (Figure 5.11) confirme que le modèle du contact secondaire explique mieux nos données.

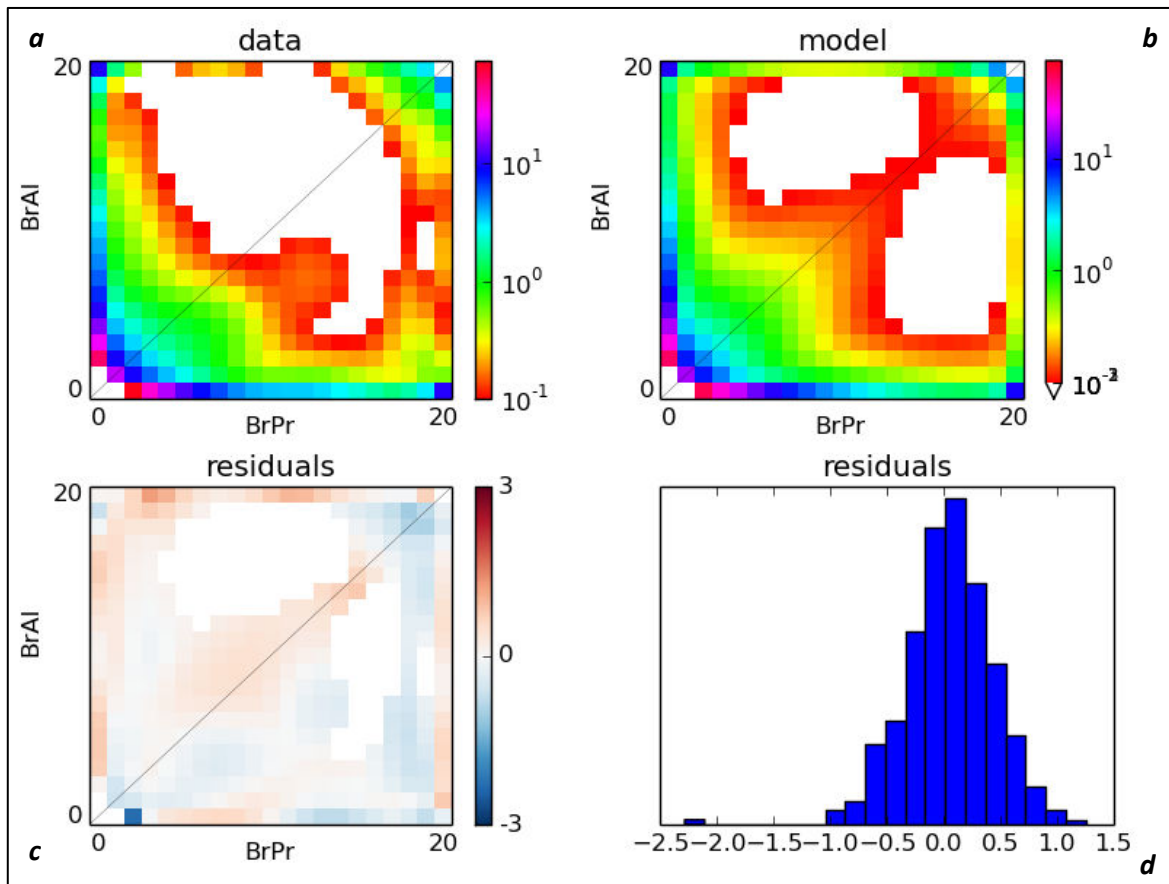


Figure 5.9 : Comparaison des spectres joints des fréquences alléliques observées (data, *a*), du meilleur modèle obtenu (model, *b*), et de la distribution des résidus (*c* et *d*), en Bretagne (scénario SC2m2N, 1973 locus). Le placement dans le spectre correspond aux fréquences alléliques dans les populations *J. albifrons* (en ordonnée) et *J. prae-hirsuta* (en abscisse), et l'échelle de couleur représente le nombre de SNPs ayant ces fréquences alléliques. Les carrés blancs représentent quant à eux une absence de SNPs ayant ces fréquences alléliques. Les résidus (residuals) correspondent à la différence entre les valeurs observées et les valeurs modélisées.

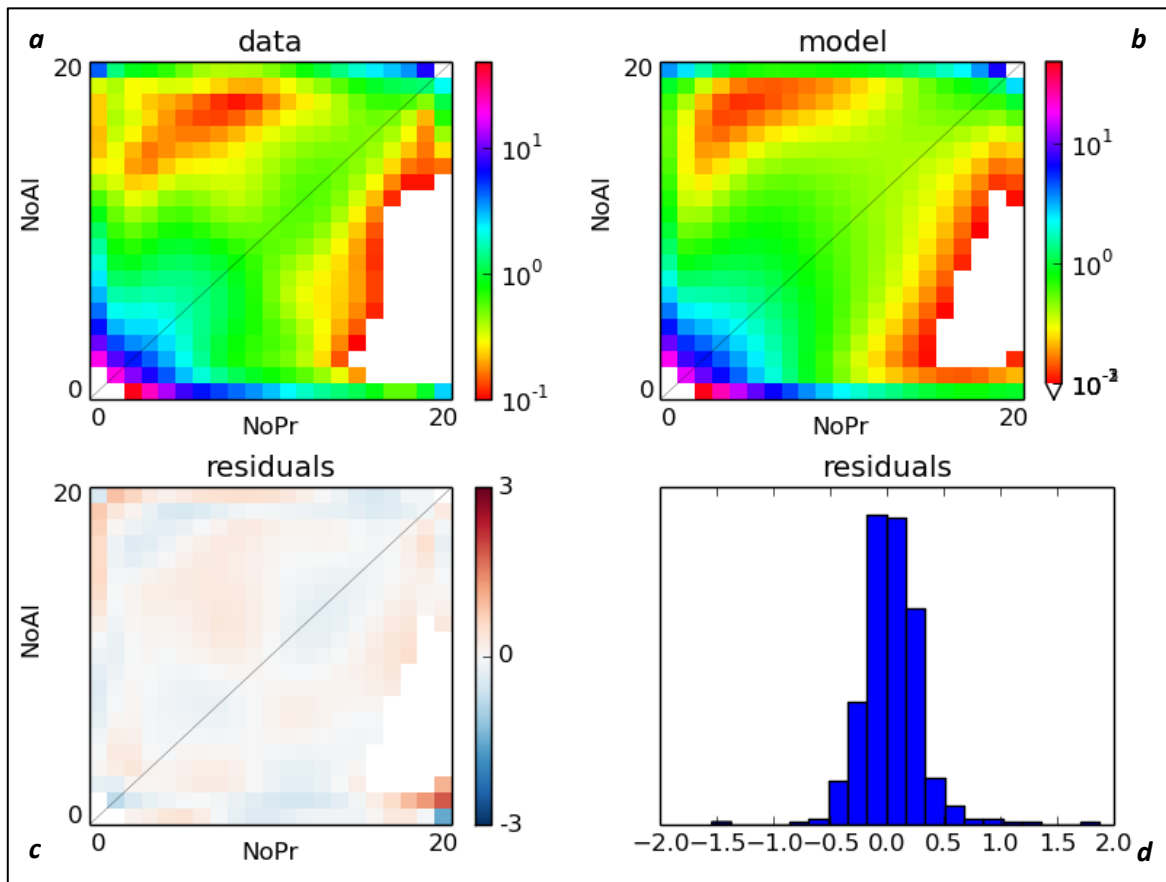


Figure 5.10 : Comparaison des spectres joints des fréquences alléliques observées (data, *a*), du meilleur modèle obtenu (model, *b*), et de la distribution des résidus (*c* et *d*), en Normandie (scénario SC2m, 1082 locus). Le placement dans le spectre correspond aux fréquences alléliques dans les populations *J. albifrons* (en ordonnée) et *J. praeheirsuta* (en abscisse), et l'échelle de couleur représente le nombre de SNPs ayant ces fréquences alléliques. Les carrés blancs représentent quant à eux une absence de SNPs ayant ces fréquences alléliques. Les résidus (residuals) correspondent à la différence entre les valeurs observées et les valeurs modélisées.

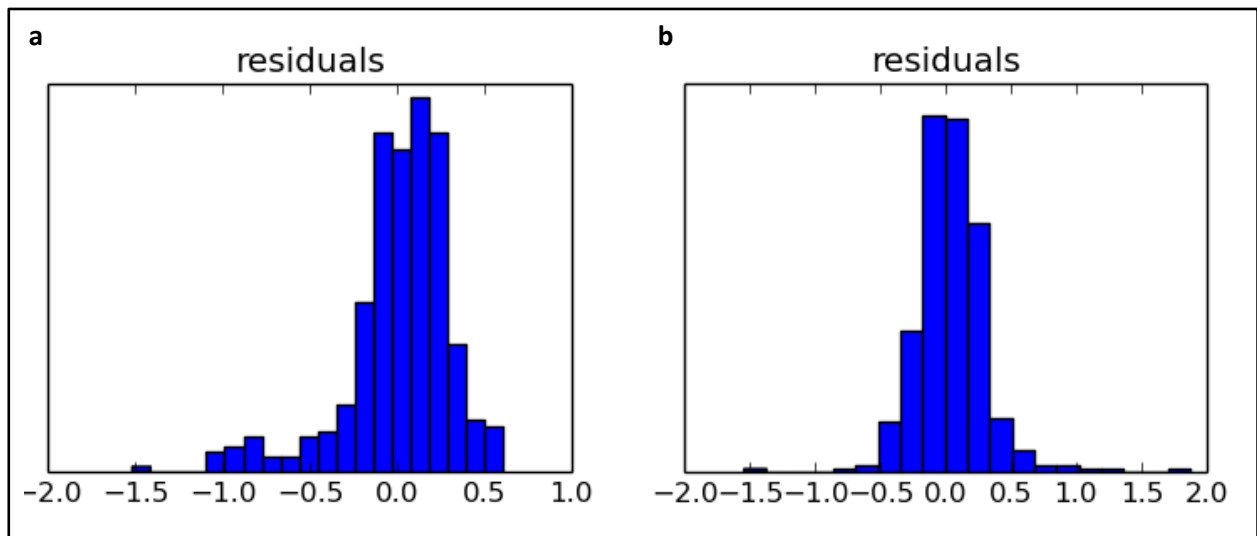


Figure 5.11 : Comparaison des résidus concernant a) le modèle IM2m et b) le modèle SC2m en Normandie

Pour chacun des deux modèles retenus, *i.e.* SC2N2m en Bretagne et SC2m en Normandie, le paramètre O (Tableau 5.2) permet d'estimer que le spectre est bien orienté à 97 % pour la Bretagne et à 96% pour la Normandie (*i.e.* les états ancestraux et dérivés ont été correctement identifiés).

L'indice θ estimé à partir des données observées est un facteur d'ajustement permettant une comparaison directe entre les spectres joints construits à partir de données réelles vs théoriques obtenus sous différents modèles (La taille efficace de la population ancestrale est arbitrairement fixée à 1 dans les modèles puis réajustée par la suite à l'aide du scalaire θ). Etant donné que la valeur de θ est plus de 6 fois plus grande en Normandie ($\theta_N = 103.64$) qu'en Bretagne ($\theta_B = 16.23$), toute comparaison directe entre régions faisant intervenir ce paramètre est assez difficile à interpréter. En particulier, θ entrant dans le calcul de la taille efficace de la population ancestrale ($N_{ref} = \frac{\theta}{4L\mu}$) et donc dans ceux des différents taux de migration ($M = m \times 2N_{ref}$) et tailles efficaces de populations ($N = \frac{n}{N_{ref}}$), nous avons décidé de ne pas comparer les valeurs absolues de ces paramètres entre régions mais de comparer uniquement les paramètres qui ne dépendent pas directement de θ .

C'est donc un contact secondaire qui semble avoir eu lieu dans les deux régions. En Bretagne comme en Normandie, le temps passé en allopatrie (T_{div}) représente plus de 80% du temps depuis la divergence. En effet, la durée relative du contact secondaire ($\frac{T_{SC}}{T_S + T_{SC}}$) semble relativement proche en Bretagne et Normandie, respectivement 17% vs 11%.

Dans les deux régions, ce contact secondaire est associé à une migration hétérogène le long du génome, certains locus étant donc soumis à un taux de migration efficace neutre et d'autres à un

taux de migration efficace réduit. Les locus soumis à un taux de migration efficace neutre représentent 37.6% du génome en Bretagne et 67.4% du génome en Normandie.

Dans les deux régions le taux de migration efficace neutre semble plus important de *J. praehirsuta* vers *J. albifrons* ($\frac{m_{12}}{m_{21}} = \frac{M_{12}}{2N_{ref}} \cdot \frac{2N_{ref}}{M_{21}} = \frac{M_{12}}{M_{21}} = 1.9$ en Bretagne et 1.7 en Normandie). Le taux de migration efficace réduit s'applique à une proportion $1-P$ de locus dont la migration interspécifique est réduite par rapport au reste du génome, et semble légèrement plus important de *J. albifrons* vers *J. praehirsuta* ($\frac{m'_{21}}{m'_{12}} = \frac{M'_{21}}{2N_{ref}} \cdot \frac{2N_{ref}}{M'_{12}} = \frac{M'_{21}}{M'_{12}} = 1.2$ en Bretagne et 1.4 en Normandie).

L'introgession à ces locus est ralentie d'un facteur moyen $\frac{1}{2} \left(\frac{m_{12}}{m'_{12}} + \frac{m_{21}}{m'_{21}} \right) = \frac{1}{2} \left(\frac{M_{12}}{M'_{12}} + \frac{M_{21}}{M'_{21}} \right) = 24.18$ en Bretagne et 18.23 en Normandie. Cependant, ce "ralentissement" semble plus fort dans le sens *J. praehirsuta* vers *J. albifrons* ($\frac{m_{12}}{m'_{12}} = \frac{M_{12}}{2N_{ref}} \cdot \frac{2N_{ref}}{M'_{12}} = \frac{M_{12}}{M'_{12}} = 33.5$ en Bretagne et 25.6 en Normandie) que dans le sens opposé ($\frac{m_{21}}{m'_{21}} = 14.8$ en Bretagne et 10.8 en Normandie)(Figure 5.12).

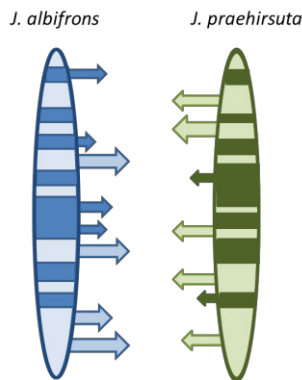


Figure 5.12 : Représentation schématique de l'asymétrie de migration le long du génome : exemple de la Normandie, où 62% du génome, représenté par les zones foncées, va migrer 15 fois moins que le reste du génome de *J. albifrons* vers *J. praehirsuta*, et 33 fois moins de *J. praehirsuta* vers *J. albifrons*.

Le produit Nm (prenant en compte les proportions du génome impacté par un taux de migration neutre ou réduit) permet d'obtenir une estimation du flux de gènes effectif moyen que l'on peut comparer entre régions. Cette estimation est réalisée de la manière suivante :

$$\begin{aligned} Nm &= N_i \times (P \times m_{ij} + (1 - P) \times m'_{ij}) \\ &= n_i \times N_{ref} \times \left(P \times \frac{M_{ij}}{N_{ref}} + (1 - P) \times \frac{M'_{ij}}{2N_{ref}} \right) \\ &= \frac{1}{2} \times n_i \times (P \times M_{ij} + (1 - P) \times M'_{ij}) \end{aligned}$$

Le flux de gènes efficace moyen de *J. praehirsuta* vers *J. albifrons* est estimé à 1.24 migrants/génération et celui de *J. albifrons* vers *J. praehirsuta* est quant à lui estimé à 3.17 migrants/génération en Normandie.

Ce calcul n'est interprétable qu'en Normandie car en Bretagne le meilleur modèle SC2m2N prend en compte l'hétérogénéité de la taille efficace locale le long du génome. Ainsi, 50% du génome (Q) à une taille efficace réduite le long du génome de 13% (Hrf), soit une taille efficace réduite $N'_1=0.73$ pour 50% des locus chez *J. albifrons* et $N'_2=0.831$ pour 50% des locus chez *J. praeheirsuta*.

Or, le produit Nm calculé ci-dessus ne prend pas en compte l'hétérogénéité de la taille efficace locale le long du génome, et la comparaison est alors difficile entre les deux régions. L'intégration de la réduction locale de la taille efficace dans le calcul du flux de gènes efficace est donc nécessaire en Bretagne et se traduit par la formule $Nm = \left[\frac{1}{4} \times n_i \times (P \times M_{ij} + (1 - P) \times M'_{ij}) \right] + \left[\frac{1}{4} \times n_i \times 0.13 \times (P \times M_{ij} + (1 - P) \times M'_{ij}) \right]$. Le flux de gènes efficace moyen en Bretagne est estimé à 1.63 migrants/génération de *J. praeheirsuta* vers *J. albifrons* et à 1.03 migrants/génération de *J. albifrons* vers *J. praeheirsuta*.

Dans les deux régions, la taille efficace de la population *J. praeheirsuta* (n_2) est plus grande que celle de la population *J. albifrons* (n_1), cependant en Bretagne la taille efficace de la population *J. albifrons* est 1.5 fois plus petite que celle de *J. praeheirsuta* alors qu'en Normandie elle est plus de 4 fois plus petite.

A titre indicatif, il est possible de faire une estimation de ces tailles efficaces de populations à partir de la taille de la population ancestrale.

Ainsi, en Bretagne, $L = 72707$ donc $N_{ref} = 558$ avec un taux de mutation $\mu = 10^{-8}$ (mais le même calcul donne entre 55.8 et 5580 avec un taux de mutation variant de 10^{-7} à 10^{-9}).

En Normandie, $L = 72085$ donc $N_{ref} = 3594$ avec un taux de mutation $\mu = 10^{-8}$ (et là-encore le même calcul donne entre 359.4 et 35944 avec un taux de mutation variant de 10^{-7} à 10^{-9}).

Le taux de mutation n'étant pas connu, aucune supposition n'a été faite à partir de ces valeurs.

Cette étude est basée sur l'analyse du modèle semblant le plus probable entre les modèles prédictifs testés.

Tableau 5.2. Pour les deux régions, les valeurs des paramètres pour le modèle ayant le critère d'information d'Akaike (AIC) le plus faible (*i.e.* le meilleur modèle, en gras), et les modèles ayant un DeltaAIC <10 à ce modèle (en plus du modèle SC2m en Bretagne, utilisé comme élément de comparaison à la Normandie) sont décrites.

	<i>AIC</i>	θ	<i>n1</i>	<i>n2</i>	<i>Hrf</i>	<i>M12</i>	<i>M21</i>	<i>M'12</i>	<i>M'21</i>	<i>ts</i>	<i>tam/tsc</i>	<i>P</i>	<i>Q</i>	<i>O</i>
Bretagne														
SC2N2m	569.182	16.229	5.509	6.248	0.133	2.653	1.392	0.079	0.094	4.282	0.870	0.376	0.499	0.972
			[4.516;6.502]	[4.972;7.524]	[0;1]	[0;5.83]	[0;5.227]	[0;0.951]	[0;1.081]	[3.418;5.147]	[0.349;1.391]	[0;0.798]	[0;1.485]	[0.959;0.986]
IM2N	575.161	38.005	7.862	7.664	0.1	0.124	0.138	-	-	4.973	-	-	0.478	0.976
			[6.652;9.072]	[5.424;9.904]	[0;0.678]	[0;1.537]	[0;2.172]	-	-	[3.768;6.178]	-	-	[0;0.997]	[0.964;0.988]
SC2m	583.836	22.963	4.543	6.644		2.601	1.039	0.058	0.053	7.403	0.266	0.390		0.964
			[3.863;5.223]	[6.022;7.265]		[1.922;3.28]	[0.265;1.814]	[0;0.736]	[0;0.956]	[6.696;8.116]	[0;0.847]	[0.222;0.558]		[0.949;0.978]
Normandie														
SC2m	684.505	103.642	0.363	1.551	-	9.930	5.811	0.387	0.536	0.784	0.094	0.674	-	0.963
			[0;0.95]	[1.228;1.875]	-	[8.574;11.286]	[4.886;6.737]	[0;1.337]	[0;1.138]	[0.266;1.303]	[0;0.873]	[0.388;0.961]	-	[0.943;0.984]
AM2m	687.023	68.555	0.521	2.319	-	5.120	2.166	0.179	0.125	2.626	0.000	0.723	-	0.959
			[0;2.176]	[0.874;3.765]	-	[3.106;7.135]	[0.585;3.747]	[0;2.097]	[0;1.61]	[0.307;4.944]	[0;44.074]	[0.524;0.923]	-	[0.931;0.986]
IM2m	689.936	88.961	0.462	1.712	-	3.963	1.951	0.000	0.131	1.599	-	0.811	-	0.963
			[0;1.267]	[0.848;2.576]	-	[3.031;4.895]	[0.594;3.307]	[0;2.377]	[0;1.597]	[0;3.270]	-	[0.692;0.929]	-	[0.937;0.989]
SC2N2m	690.705	18.759	1.165	5.576	0.128	2.517	1.299	0.063	0.061	4.638	0.966	0.609	0.493	0.944
			[1.085;1.245]	[3.995;7.158]	[0;0.674]	[1.515;3.52]	[0;4.624]	[0;2.216]	[0;2.131]	[2.6;6.676]	[0.946;0.986]	[0.055;1]	[0;2.244]	[0.899;0.989]

*Le tableau contient dans l'ordre, la valeur de l'AIC associé au modèle, l'estimation de la diversité génétique θ de la population ancestrale, la taille efficace pondérée de la population *J. albifrons* (*n1*) et *J. praeheirsuta* (*n2*), la valeur du facteur Hill-Robertsonien (*Hrf*) correspondant à l'estimation de la réduction de la taille efficace dans des zones du génome étant soumises aux effets du balayage sélectif, les taux de migration neutre pondérés de *J. praeheirsuta* dans *J. albifrons* (*M12*) et inversement (*M21*) ainsi que les taux de migration réduit pondérés correspondant aux régions génomiques sous sélection de *J. praeheirsuta* dans *J. albifrons* (*M'12*) et inversement (*M'21*), l'estimation de la durée depuis la divergence des deux espèces (*ts*) et le temps depuis que la migration est arrêtée dans le cas de la migration ancestrale (*tam*) ou a recommencé après un contact secondaire (*tsc*). Enfin, la proportion du génome avec une taille efficace neutre (*P*), la proportion du génome avec une taille effective réduite due à une sélection aux sites liés *Q* et la bonne orientation des SNPs (ancestraux ou dérivés; *O*) sont également donnés.

4. Discussion

Retracer l'histoire de la divergence entre deux espèces, en distinguant les processus démographiques et sélectifs sous-jacents, est un réel défi dans la compréhension de la spéciation. Cette étude qui porte sur une paire d'espèces, *J. albifrons* et *J. praehirsuta*, montrant une différenciation génétique interspécifique hétérogène entre deux zones de sympatrie, apporte un soutien supplémentaire quant à la prise en compte de modèles de contact secondaire face aux hypothèses de spéciation parallèle.

Divergence unique suivie de contacts secondaires

Parmi les 13 modèles inférés basés sur des modèles classiques d'isolement strict, de migration ancienne, d'isolement avec migration et de contact secondaire, on peut affirmer que les modèles détectés SC2N2m en Bretagne (*i.e.* contact secondaire avec taille efficace et migration hétérogène le long du génome) et SC2m en Normandie (contact secondaire avec migration hétérogène le long du génome) expliquent mieux l'histoire démographique entre *J. albifrons* et *J. praehirsuta* que des modèles simples alternatifs. Ce résultat a été obtenu malgré la difficulté potentielle de discriminer un modèle d'un autre. En effet, selon Rougemont et al. (2016b), certains modèles comme l'isolement avec migration et le contact secondaire peuvent être confondus si le contact secondaire a eu lieu rapidement après la divergence, et que le signal historique de la divergence a disparu.

Dans notre cas, on peut noter dans un premier temps que les spectres sont bien orientés, *i.e.* les états ancestraux et dérivés ont été correctement identifiés (cf. paramètre O, Tableau 5.2), ce qui confirme le choix correct comme outgroup de l'espèce *Jaera ischiosetosa*. Dans les deux régions, le paramètre d'Akaike (AIC) vient également soutenir l'hypothèse d'un contact secondaire, et le graphique des résidus non seulement confirme que le modèle de contact secondaire est meilleur que le modèle invoquant de la migration avec flux de gènes continu mais suggère de plus qu'un scénario de type SC explique très bien les données observées. Le même travail a été réalisé sans groupe externe c'est-à-dire sans la possibilité de détecter les états ancestraux spectres, *i.e.* repliés, sur un nombre de SNPs plus grand, mais avec moins de modèles démographiques possibles (cf. annexe 4). Dans tous les cas, et malgré des jeux de données et des filtres différents, c'est un modèle de type contact secondaire qui a été détecté, relativement récent, puisque en Bretagne comme en Normandie le temps passé en allopatrie représente plus de 80% du temps depuis la divergence.

Dans cette étude, deux zones de sympatrie sont représentées, dont l'une avec une adaptation différente à l'habitat. Pour rappel, en Bretagne, les individus ont été trouvés sur des habitats différents, *J. albifrons* sous les pierres et *J. praehirsuta* sur des algues, alors qu'en Normandie

tous les individus (indépendamment de l'espèce) partagent un même habitat et ont été trouvés sous les pierres. Or, dans les deux cas, il y a eu séparation allopatrique suivi par un contact secondaire, ce qui vient renforcer l'idée qu'il ne s'agit pas d'une spéciation parallèle initiée par l'adaptation locale (voir aussi Le Moan et al., 2016). Contrairement au faible niveau de partage génomique entre écotypes de types *Littorina saxatilis* (Ravinet et al., 2016) et à la différenciation génétique intraspécifique inter-écotypes de cichlidés du genre *Pundamilia* plus grande que la différenciation génétique interspécifique intra-écotype (résultat qui semblerait être dû à des adaptations locales parallèles, Meier et al., 2017), 20 à 35% des locus utilisés dans cette étude sont des locus partagés entre régions. Les locus impliqués dans l'isolement reproductif et partagés entre répliques dateraient donc de la divergence allopatrique initiale (Bierne et al., 2013, Welch & Jiggins, 2014).

Des barrières semi-perméables

En Bretagne, on observe que les échanges entre espèces sont nettement plus réduits qu'en Normandie (séparation d'habitat et isolement sexuel), et on peut confirmer que les différences de H_s entre régions et la diminution du F_{ST} en Normandie sont liées à des flux de gènes entre espèces et non pas dues à un problème d'échantillonnage. En effet, les valeurs au sein de chaque site normand sont équivalentes, et le F_{ST} spatial entre sites normands est proche du F_{ST} spatial entre sites bretons (Ribardière et al., 2017).

En Normandie, la plus grande proportion du génome est soumise à une migration neutre, et l'introgression différentielle, observée dans les spectres joints, vient confirmer la conclusion du chapitre 1 réalisée avec des marqueurs microsatellites sur la présence de barrières semi-perméables. Concernant l'introgression asymétrique interspécifique, le calcul $\frac{m_{12}}{m_{21}}$ ne prend pas en compte la taille de la population ($N_1 > N_2$), ce qui vient biaiser les calculs ne donnant pas de résultats cohérents par rapport aux études précédentes, à l'inverse du calcul Nm qui montre bien l'introgression plus forte de *J. albifrons* vers *J. prae-hirsuta*.

En Bretagne, la plus grande proportion du génome est soumise à une migration réduite, et ce flux de gènes hétérogène était inattendu au vue des précédents résultats obtenus avec des marqueurs microsatellites (chapitre 1). Cette semi-perméabilité s'accompagne, de plus, d'effets sur l'ensemble du génome de sélection indirecte dans le cas de la Bretagne (effets Hill-Robertson au sens large). En effet, la moitié du génome à une taille efficace réduite de 13% ce qui augmente l'effet de la dérive avec la fixation d'allèles différenciés et/ou délétères entre populations, et par conséquent la mise en place d'un isolement reproductif (Excoffier & Ray, 2008), ce qui explique donc qu'on retrouve plus de locus polymorphes mais privés en Bretagne qu'en Normandie (Figure 5.4). Le calcul du flux génique Nm tel que réalisé en Normandie ne

prenait pas en compte l'hétérogénéité de la taille efficace le long du génome, et lorsque l'on calcule Nm en incluant le paramètre Hrf , le flux de gènes efficace moyen est plus faible qu'en Normandie, résultat attendu au vu des spectres joints des fréquences alléliques observées, des croisements expérimentaux réalisés (chapitre 3), et des résultats de différenciation génétiques obtenus précédemment (chapitre 2). Le flux de gènes bien qu'existant est faible ce qui i) peut restreindre les échanges interspécifiques, et permettre le maintien de forts isolements reproductifs tels que la présence de caractères sexuels secondaires, les préférences sexuelles, ou encore l'adaptation à l'habitat, soit ii) être au contraire expliqué par le renforcement de la sélection par l'adaptation locale avec comme résultante une diminution de la migration efficace.

Spéciation

Un scénario possible serait que lors de la phase d'isolement, il y a eu une évolution indépendante des fréquences alléliques entraînées par sélection ou dérive, permettant la mise en place d'incompatibilités de type Dobzhansky-Muller (iDM), générant des barrières endogènes au flux de gènes, ou la mise en place d'un isolement sexuel. Lors de la remise en contact le génome subit une homogénéisation progressive des fréquences alléliques, pouvant être bloquée aux locus d'incompatibilités, créant ainsi des flux de gènes hétérogènes, détectés dans cette étude pour chaque réplica.

En Normandie, seulement 33 % du génome est soumis à une migration réduite, permettant de larges flux de gènes et expliquant ainsi le F_{ST} interspécifique faible en Normandie. En revanche, le faible isolement interspécifique ne permet pas de détecter une hétérogénéité de la taille efficace le long du génome. De plus, le taux de migration aux locus soumis à des effets sélectifs directs ou indirects est beaucoup plus fort de *J. albifrons* vers *J. praehirsuta* qu'inversement, expliquant l'introgression asymétrique visible dans le spectre des fréquences alléliques observées, ainsi que dans les résultats d'ACP (cf. chapitres 1 et 2). En Normandie, les iDM ne semblent pas imposer de barrières fortes, au vu des résultats de croisements expérimentaux (chapitre 3), et de par la présence d'hybrides et d'individus introgressés (chapitre 2). L'isolement sexuel semble en revanche avoir un rôle prépondérant dans l'établissement de flux de gènes hétérogènes le long du génome. Une autre hypothèse, non exclusive avec la présence d'un isolement sexuel ou d'iDM serait une introgression adaptative de l'espèce *J. praehirsuta* au vu de l'absence de différenciation d'habitat en Normandie. Ainsi, lors du contact secondaire, les individus *J. praehirsuta* se seraient retrouvés dans des environnements mal adaptés, avec peu d'endroits abrités le long des côtes normandes, et ils auraient tirés avantage de l'hybridation avec l'espèce *J. albifrons* pour s'adapter à un nouvel habitat, *i.e* sous les pierres

(Hedrick, 2013). Cette introgression adaptative aurait conduit à une homogénéisation du génome, tout en gardant un certain isolement reproductif (Clarkson et al., 2014) permettant ainsi le maintien des morphes parentaux.

Limites de l'étude : Valeurs absolues des paramètres inférés

Ces analyses pourraient être transformées de manière à donner une estimation de la taille efficace de la population ancestrale et ainsi pouvoir dater le contact secondaire ainsi que de pouvoir comparer plus directement les populations des deux régions les unes avec les autres. Pour cela, il faudrait connaître le taux de mutation, ce que nous avons décidé de ne pas supposer. Une autre difficulté est d'utiliser l'indice θ (Gutenkunst et al., 2009, Tine et al., 2014, Rougeux et al., 2016), qui est un facteur d'ajustement permettant une comparaison directe entre les spectres joints construits à partir de données réelles vs simulées sous différents modèles (θ est arbitrairement fixé à 1 dans les modèles). Dans cette étude nous avons donc décidé de s'affranchir de ce paramètre en réalisant une comparaison de modèle à partir du ratio des résultats obtenus. Enfin, dans cette étude, nous n'avons pris en compte que les variations dans le taux de migration et de la taille efficace pour prendre en compte les barrières génétiques et l'influence de la background selection, mais d'autres modèles incorporant des variations, *i.e* expansions ou réductions, temporelles dans les tailles des populations sont par exemple possibles, et ont été utilisées dans l'étude de Rougeux et al. (2016). Il reste donc un large champ de possible quant à l'utilisation des modèles d'inférences démographiques dans la compréhension des évènements de spéciation passés.

Conclusions et perspectives

Les résultats de ce chapitre montre clairement un événement de contact secondaire en Bretagne et en Normandie dont la durée relative paraît similaire. Ajouté au niveau de différenciation trouvé entre nos espèces (voir notamment l'ACP et l'arbre retraçant les divergences et migrations, Chapitre 2), il apparaît donc qu'il n'y ait eu qu'une seule étape de divergence suivi de contacts secondaires, tels que trouvés dans les études de Rougemont et al. (2016a) et Le Moan et al. (2016). Logiquement cela signifie n'avoir qu'une seule population ancestrale précédant l'évènement de divergence. Cependant, la taille efficace de la population ancestrale normande $\theta_N = 103.64$ semblerait être plus de 6 fois plus grande que celle de la population ancestrale bretonne $\theta_B = 16.23$. Il y a donc plusieurs suppositions : i) la taille ancestrale est différente entre les deux régions, ce qui ne correspondrait pas à un unique évènement de divergence ou ii) le modèle en Normandie n'arrive pas vraiment à distinguer les effets de la dérive (taille efficace ancestrale) des effets des flux de gènes (taux m depuis le

contact secondaire), et il surestimerait θ au détriment de m ce qui induirait une sous-estimation du taux de migration. Une possibilité est d'utiliser le modèle SC2N2m en Normandie, *i.e.* le même modèle que celui détecté en Bretagne, qui montre une information similaire entre les deux régions (cf. Tableau 5.2) si ce n'est la proportion du génome soumise à une migration neutre qui est beaucoup plus grande en Normandie qu'en Bretagne (respectivement 61% vs 38%). Le θ de ce modèle est similaire entre régions, ce qui est compatible avec l'hypothèse d'un unique événement de divergence.

Dans ce chapitre, il a donc été montré que l'histoire de la divergence entre *J. albifrons* et *J. praeirsuta* semble suivre un modèle de type contact secondaire, en Bretagne et en Normandie, avec dans les deux cas un flux de gènes hétérogène actuel le long du génome, qui semble bien plus restreint en Bretagne qu'en Normandie. En Bretagne, le contact secondaire n'était pas attendu, et le flux de gènes actuel ou récent est un résultat nouveau et intéressant, non suspecté par les résultats précédemment obtenus (chapitre 2), ou par les observations des morphes et leur distribution sur le terrain. En revanche en Normandie, le contact secondaire, le flux de gènes actuel et l'asymétrie d'introgression de *J. albifrons* vers *J. praeirsuta* détectés dans ce chapitre ne font que confirmer les constats des travaux réalisés auparavant. Les résultats obtenus sont donc cohérents i) entre les spectres joints obtenus avec données observées et données simulées ii) avec les résultats obtenus par $\delta a \delta i$ sans utilisation des allèles ancestraux (annexe 4) iii) avec la description de la différenciation génétique entre les deux espèces et dans les deux régions effectuées avec des marqueurs microsatellites (cf. chapitre 2). En plus d'apporter de nouveaux renseignements sur ce complexe d'espèces, les modèles inférés vont pouvoir être utilisés pour détecter l'enveloppe neutre attendue et identifier quels sont les locus soumis à migration et/ou taille efficace réduites dans le cadre d'une analyse de type scan génomique (chapitre 6).

Chapitre 6

**ARCHITECTURE GENOMIQUE DE
L'ISOLEMENT REPRODUCTIF**

L'objectif de ce chapitre est d'analyser l'isolement reproductif d'un point de vue génomique entre *Jaera albifrons* et *J. praehirsuta*, en décrivant la distribution de la différenciation entre espèces le long de leur génome. Cette description est basée sur la réalisation de cartes de liaison, afin de détecter d'éventuels réarrangements chromosomiques interspécifiques (voire intra-spécifiques), et d'un scan génomique permettant de mesurer la différenciation entre espèces et d'en observer la distribution sur les cartes génétiques réalisées.

Les résultats de ce chapitre ont été obtenus en collaboration avec Komlan Avia et Pierre-Alexandre Gagnaire.

1. Introduction

L'isolement reproductif peut résulter d'une accumulation de barrières pré- et/ou post-zygotiques (Barton & de Cara 2009), façonnant le paysage génomique conjointement à des processus externes tels que sélection divergente ou sélection d'arrière-plan sur des régions de polymorphisme ancestral partagé. Les régions dont la divergence n'est pas attendue sous un modèle neutre, sont appelées "îlots génomiques" (Turner et al., 2005, Harr, 2006, Nosil et al., 2009), et peuvent être détectées grâce à un scan génomique (observation de la différenciation génomique interspécifique le long du génome). Il est prédit que ces îlots génomiques de différenciation apparaissent dans le cadre d'un isolement reproductif interspécifique, alors que les flux de gènes (intervenant lors de contacts secondaires par exemple), vont agir comme supprimeurs de la différenciation génomique (Navarro & Barton, 2003, Wu, 2001, Turner et al., 2005, Feder et al., 2012).

Lorsque ces scans génomiques sont associés à des recherches d'expression de gènes ou de loci liés à des gènes codants pour des traits impliqués par exemple dans le comportement sexuel ou l'adaptation locale, cela peut mener à une compréhension plus globale du paysage de divergence génomique (Smadja et al., 2012).

Détecter les régions du génome sous sélection peut permettre de comprendre les patterns d'introgession dans des zones hybrides (Teeter et al., 2010). Les patterns de divergence peuvent être variables, de l'isolement strict à la spéciation hybride, et peuvent être affectés par le taux de recombinaison, les phénomènes de réarrangements (*e.g.* inversions) pouvant diminuer ce taux et dans certains cas participer à l'évolution de l'isolement reproductif (Navarro & Barton, 2003, Noor & Bennett, 2009, Roesti et al., 2012).

Depuis quelques années les outils de compréhension de la divergence entre organismes non modèles ont émergé (Metzker, 2010, Sousa & Hey, 2013, Seehausen et al., 2014, Ravinet et al., 2017). Avec la facilité d'obtention de larges jeux de données génomiques, de plus en plus de génomes de référence

sont disponibles, mais il est également possible de travailler à partir de cartes de liaison (Sun et al., 2017, Johnston et al., 2017).

Le sexe hétérogamétique chez les espèces du complexe *Jaera albifrons* est la femelle, dont le caryotype révèle un trivalent sexuel (Z, W1, W2, voir Staiger & Bocquet, 1954 et Bocquet, 1972). Les chromosomes sexuels étant souvent fortement impliqués dans la différenciation entre espèces (par exemple chez les gobe-mouches, Ellegren et al., 2012), il serait intéressant de vérifier si chez *Jaera sp.* une forte différenciation est discernable sur ces chromosomes. Les remaniements chromosomiques peuvent constituer une autre clef de la divergence et du maintien de la différenciation entre espèces (Rieseberg, 2001). Chez les espèces du complexe *Jaera albifrons*, la description de caryotypes réalisée dans les années 70 suggère que le polymorphisme chromosomique varie dans le nombre de chromosomes portés par chaque espèce (Lécher & Prunus, 1971). Ainsi l'étude de Bocquet (1972) a montré que dans certaines populations bretonnes les *J. ischiosetosa* avaient 12 paires d'autosomes et 1 trivalent sexuel, *J. praehirsuta* 10 paires d'autosomes et 1 trivalent sexuel, *J. forsmanni* 8 paires d'autosomes et 1 trivalent sexuel et *J. albifrons* 9 paires d'autosomes et 1 trivalent sexuel. Cependant, il a également été observé qu'une variabilité du nombre de chromosomes pouvait avoir lieu au sein d'une même espèce entre différentes populations, et parfois même au sein d'un même site. Ainsi l'espèce *J. albifrons* montre une variation entre populations de 9 à 11 paires de chromosomes entre et au sein de certains sites de Bretagne et Normandie (Lécher, 1968). L'étude la plus récente concernant de possibles réarrangements chez *Jaera sp.* est celle de Mifsud (2011). En créant des familles dont les individus sont issus de backcross réciproques entre *J. albifrons* et *J. ischiosetosa*, il a créé et comparé des cartes de liaison à l'aide de marqueurs AFLP. Ainsi pour la carte BcAlbs (croisement interspécifique femelle *J. albifrons* - mâle *J. ischiosetosa*, puis croisement de type backcross mâle hybride F1 - femelle *J. albifrons*) réalisée à partir de 207 juvéniles, 171 marqueurs AFLP ont été cartographiés sur 10 groupes de liaison. Quatre marqueurs liés au sexe ont été trouvés et cartographiés sur un même groupe de liaison, déterminé comme le chromosome sexuel. Concernant la carte BCIsch (croisement interspécifique femelle *J. ischiosetosa* - mâle *J. albifrons*, puis croisement de type backcross mâle hybride F1 - femelle *J. ischiosetosa*) réalisée à partir de 176 juvéniles, 99 marqueurs AFLP ont été cartographiés sur 13 groupes de liaison. Seul un marqueur lié au sexe a été identifié (sans que l'auteur puisse confirmer s'il s'agit du même groupe que le groupe lié au sexe cité ci-dessus). Neuf marqueurs partagés entre les deux cartes ont permis de détecter 7 groupes de liaison semblant homologues, dont 5 seraient impliqués dans des fusions Robertsoniennes. Les QTL pour 6 traits phénotypiques (nombre de soies et forme des différents segments des péréiopodes) ont été cartographiés sur 5 groupes de liaisons dont 3 montrant des réarrangements chromosomiques chez BcIsch, et sur 6 groupes de liaisons dont

4 forment des réarrangements chromosomiques chez BcAlbs. Enfin, 17 % de la variation phénotypique de l'un de ces 6 traits serait sur le chromosome sexuel présumé chez BcAlbs.

Face à une histoire complexe de différenciation interspécifique entre *J. albifrons* et *J. praeheirsuta* variable en fonction des sites étudiés (chapitres 1, 2, et 5), il est intéressant de rechercher si l'hétérogénéité de la divergence est corrélée à l'existence d'îlots de différenciation et si des réarrangements chromosomiques peuvent être liés à distribution génomique de la divergence. Un des avantages de ce modèle pour une telle étude est qu'il est possible de comparer deux réplicas de sympatrie avec différenciation génétique variable, et donc d'essayer d'inférer la nature des processus à l'origine de la différenciation génétique entre espèces. En Normandie, où la différenciation génétique interspécifique est faible, il devrait ainsi être possible d'identifier les régions génomiques impactées par les flux de gènes interspécifiques et l'introgression entre *Jaera albifrons* et *Jaera praeheirsuta*.

Pour cela, nous avons i) mesuré la différenciation génétique interspécifique en populations naturelles, puis ii) identifié les marqueurs ayant un niveau de différenciation incompatible avec un attendu neutre avec recombinaison libre (c'est-à-dire sans limitation des flux géniques et de l'introgression entre espèces par des effets sélectifs ou des contraintes à la recombinaison), iii) réalisé des cartes de liaison génétiques, et enfin nous avons iv) utilisé ces cartes pour détecter d'éventuels réarrangements inter- et intra-spécifiques et pour analyser la distribution de la différenciation génétique inter-spécifique le long du génome.

2. Matériel et méthodes

2.1 Cartes génétiques et polymorphisme chromosomique

Lors des croisements expérimentaux (voir chapitre 3), quatre familles susceptibles d'être utilisées pour créer des cartes génétiques ont été obtenues (Tableau 6.1) : deux familles *J. albifrons* respectivement constituées de 59 et 57 juvéniles, et 2 familles *J. praeheirsuta* de 70 et 62 juvéniles. Les deux familles *J. albifrons* sont d'origine Bretonne (*i.e.* les quatre parents sont eux-mêmes issus de portées produites au laboratoire par des femelles qui avaient été échantillonnées en Bretagne), tandis que les deux familles *J. praeheirsuta* sont d'origine Normande. En moyenne 6 portées successives ont été obtenues pour les familles *J. albifrons* (avec en moyenne 12 juvéniles par portées) et 4.5 portées (avec en moyenne 19 juvéniles par portée) ont été obtenues pour les familles *J. praeheirsuta*. Chaque juvénile de chaque portée avait alors été élevé pendant 50 jours avant d'être sexé et fixé à l'éthanol. Bien que l'obtention de ces familles représente un effort d'élevage conséquent (la survie et la fécondité des femelles adultes étant variables, un grand nombre de

croisements ont été réalisés et leur descendance élevée pour arriver finalement à obtenir ces quatre familles nombreuses), le nombre total de juvéniles par famille est, relativement à l'objectif d'estimer des taux de recombinaison méiotique, assez faible. Ces faibles nombres de juvéniles peuvent être en partie compensés par l'utilisation combinée de deux familles par espèce (ce travail sera réalisé ultérieurement, les résultats de ce chapitre étant basés sur les familles 1 et 3 seulement, cf. ci-dessous).

Tableau 6.1. Description des familles obtenues lors de croisements intraspécifiques contrôlés

Famille	espèce	Site d'origine (mère/père)	nb	nb	nb total
			males	femelles	juvéniles
1	<i>J. albifrons</i>	Lingoz/Lingoz (Bretagne)	35	24	59
2	<i>J. albifrons</i>	Lingoz/Lingoz (Bretagne)	34	23	57
3	<i>J. prae-hirsuta</i>	Sainte-Honorine/Grandcamp (Normandie)	31	39	70
4	<i>J. prae-hirsuta</i>	Sainte-Honorine/Grandcamp (Normandie)	21	41	62

Ces quatre familles (parents et descendants) ont été séquencées par ddRAD (Banque 6, janvier 2017, cf. annexe 1). Les différents logiciels et paramètres utilisés ont été récapitulés en Figure 6.1 et sont détaillés en annexe 2. Les paramètres utilisés dans *Stacks* (Catchen et al., 2013) ont été -m3 -M3 -N5 -n3, le catalogue a été créé sur les huit parents, puis les génotypes des juvéniles ont été assignés au catalogue et l'option permettant de visualiser les génotypes sur une database SQL en ligne a été activée. Enfin le programme *Genotypes* a été lancé sur chacune des 4 familles, avec une option -r permettant de ne garder les loci existants que chez un certain nombre de juvéniles (ici 30 juvéniles minimum séquencés pour garder un locus) et une correction appliquée sur les loci ayant peu de profondeur de couverture (option -c permettant de détecter de faux homozygotes par exemple). Le fichier de sortie consiste alors en une liste d'haplotypes (chaque haplotype correspondant à la combinaison d'allèles identifiés au sein d'un locus de type RAD, qui peut contenir plusieurs SNPs).

Un premier filtre a été réalisé sur les données ainsi obtenues afin de retirer les loci ayant un log likelihood supérieur à -30, ainsi que ceux absents chez plus de 10 juvéniles. Pour cela, les génotypes présentant des haplotypes suspects ont été regardés en détail sur la database SQL, qui permet de comparer facilement les génotypes des juvéniles à ceux des parents. Les haplotypes restants ont été réunis dans une liste des loci à garder, puis ces données ont été de nouveau analysées sur l'ensemble des juvéniles *via Genotypes*, afin de retenir les loci ayant une certaine profondeur de couverture minimale (ici 10 séquences nécessaires pour garder le locus).

A ce stade, quatre fichiers de 12 686 haplotypes présents chez le père et chez la mère ont été obtenus pour la Famille 1, 12 265 haplotypes pour la Famille 2, 11 568 haplotypes pour la Famille 3 et 9 698 haplotypes pour la Famille 4. Ces fichiers ont ensuite été utilisés avec le logiciel de calcul de cartes génétiques Joinmap 4 (Van Ooijen, 2006). Tout d'abord les individus ayant trop de données manquantes ont été retirés, soit un individu de la Famille 1 (une femelle), et deux individus de la Famille 4 (un mâle et une femelle) car présentant plus de 99 % de données manquantes. Le taux de données manquantes moyen pour les autres juvéniles était de 5 % (maximum à 40% de données manquantes) pour la famille F1 et de 12% (maximum de 42% de données manquantes) pour la famille F3. Ensuite les loci présentant des distorsions de ségrégation très significatives (p -value ≥ 0.0005 , équivalent à un X2 supérieur à 12) ont été retirés.

Après avoir créé les groupes de liaison en fonction du score LOD de test d'indépendance, l'assignation des groupes entre eux a été explorée grâce à la valeur SCL (paramètre « Strongest Cross Link ») permettant d'assigner par la suite un marqueur qui n'aurait pas été groupé dans un premier temps et de vérifier que les groupes avaient bien été définis. Les groupements suspects entre loci ont ensuite été tous retirés (retrait manuel progressif des loci les plus impliqués aux loci les moins impliqués dans ces groupements suspects). Une fois les groupes de liaison déterminés, les cartes génétiques ont été créées pour chacun des groupes, en utilisant l'algorithme du « Monte Carlo maximum likelihood » qui permet de produire rapidement des cartes parentales séparées.

Après avoir comparé les sorties des calculs avec options de base (« Chain length »: 1 000, « Initial acceptance probability »: 0.25, « Cooling control parameter »: 0.001, « Stop after # chains without improvement »: 10 000, « Length of burn-in chain »: 10 000, « Chain length per Monte Carlo EM cycle »: 1 000, fonction de Haldane) et options modifiées (« Chain length »: 10 000, « Initial acceptance probability »: 0.35, « Cooling control parameter »: 0.0005, « Stop after # chains without improvement »: 30 000, « Length of burn-in chain »: 30 000, « Chain length per Monte Carlo EM cycle »: 3 000, fonction de Haldane) pour chacune des cartes parentales de la Famille 1, il s'est avéré que cela n'avait pas d'impact sur les cartes génétiques obtenues, et les options de base ont donc été gardées pour les autres familles. Pour chacune des cartes parentales, 3 runs avec les options de base ont été réalisés pour chaque groupe de liaison afin de vérifier la stabilité des résultats obtenus.

Pour les analyses réalisées dans ce chapitre, seules les familles 1 et 3 ont été utilisées, soit une famille par espèce (celle ayant le plus de juvéniles) : la Famille 1 *J. albifrons* de Bretagne et la Famille 3 *J. praehirsuta* de Normandie.

Deux types de cartes ont été réalisées : 1) deux cartes de liaison parentales par famille dites "cartes de référence", obtenues par la méthode décrite ci-dessus mais uniquement à partir de loci n'ayant strictement aucune donnée manquante lors du premier filtre (nombre de loci finaux cartographiés chez la famille *J. albifrons* : $n_{\text{paternel}} = 383$, $n_{\text{maternel}} = 356$ et la famille *J. praehirsuta* :

$n_{\text{paternel}} = 395$, $n_{\text{maternel}} = 431$), et 2) deux cartes de liaison parentales par famille ("cartes totales"), *i.e.* retrait des loci absents chez plus de 10 juvéniles comme décrit dans la méthode ci-dessus (nombre de loci finaux cartographiés chez la famille *J. albifrons* : $n_{\text{paternel}} = 3\,909$, $n_{\text{maternel}} = 4\,195$ et la famille *J. praehirsuta* : $n_{\text{paternel}} = 3\,849$, $n_{\text{maternel}} = 3\,771$). On suppose que les cartes de référence produisent la meilleure estimation des placements relatifs des marqueurs le long des groupes de liaisons (GL), et ces placements ont donc été utilisés comme ancrage à l'ajout de marqueurs supplémentaires pour réaliser les cartes dites totales. A la fin de l'obtention des cartes totales, les noms des groupes de liaisons finaux ont également été attribués aux groupes de liaisons correspondant de la carte de référence pour plus de lisibilité lors des comparaisons de cartes. La représentation des cartes de liaisons dites de référence ont été obtenues sous R (Core Team, 2016) à l'aide du package *qtl* (Broman et al., 2003) et les différentes représentations des cartes de liaison dites totales ont été réalisées à l'aide du programme de cartes génétiques comparées *Vgsc2* (Xu et al., 2016). Ce programme permet la visualisation de la synténie et de la colinéarité au sein d'une famille de gènes (ici d'haplotypes placés sur des groupes de liaisons) en comparant les deux parents d'une même espèce ou d'espèces différentes.

Enfin, les loci montrant un patron de ségrégation lié au sexe ont été recherchés en analysant la distribution des génotypes parentaux et juvéniles. Cette analyse repose sur le fait que le sexe des juvéniles est connu. Ainsi, il est très efficace de contrôler si la ségrégation des allèles à un locus donné est compatible ou non avec une localisation autosomale ou liée au sexe. Tous les loci retenus pour les cartes finales et dont la ségrégation était de type "nn x np" (hétérozygote chez la mère), "ef x eg" (hétérozygote chez les deux parents, avec un allèle en commun), ou "ab x cd" (hétérozygote chez les deux parents, sans allèle commun) ont été analysés. Par exemple dans le cas "nn x np", un locus autosomal devrait être caractérisé par une proportion équivalente d'hétérozygotes np dans les juvéniles des deux sexes, tandis qu'un locus lié aux chromosomes sexuels devrait se traduire par une proportion nulle (allèle p sur le chromosome Z) ou totale (allèle p sur le chromosome W1 ou W2) d'hétérozygotes np chez les femelles. Le même raisonnement aurait pu être appliqué aux génotypes de type hk x hk, mais avec moins de précision (car les fréquences génotypiques attendues chez les juvéniles en fonction de leur sexe ne sont pas 0 ou 1). Plusieurs centaines de loci dont la ségrégation était strictement compatible avec une liaison au sexe ont ainsi été identifiés dans les deux familles utilisées pour réaliser les cartes génétiques (Tableau 6.2). Etant donné que notre objectif ici est l'identification du groupe de liaison correspondant aux chromosomes sexuels (et non l'identification de tous les loci possiblement liés au sexe), nous avons focalisé notre analyse sur les patrons de ségrégation les plus informatifs.

Tableau 6.2. Nombre de loci géotypés dans le cadre de la construction de cartes de liaison (famille 1 pour *J. albifrons* et famille 3 pour *J. praeheirsuta*) et dont la ségrégation est strictement compatible avec une liaison au sexe (et strictement incompatible avec une localisation autosomale).

Type	<i>J. albifrons</i>	<i>J. praeheirsuta</i>
nn x np	222	223
ef x eg	59	44
ab x cd	44	23

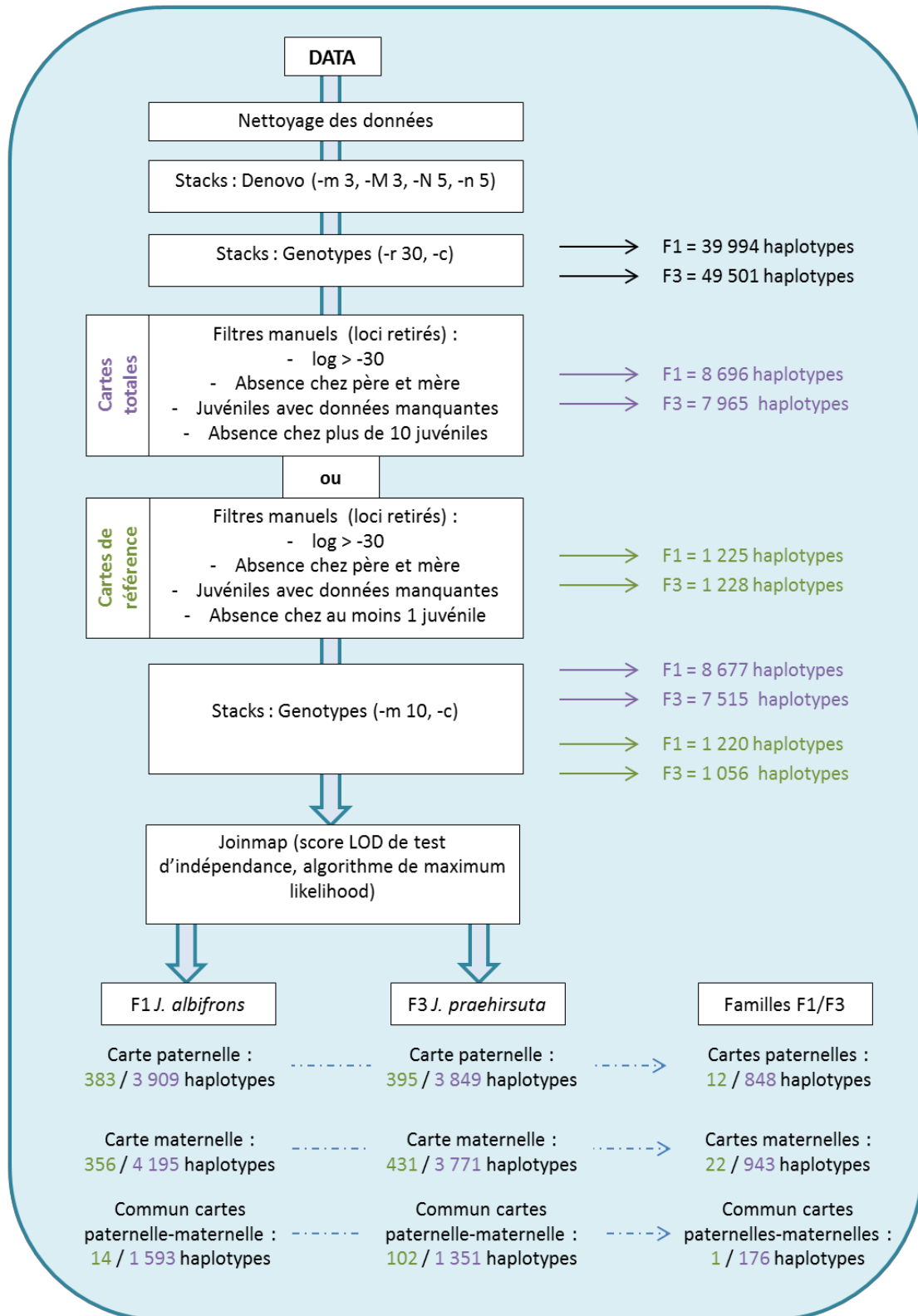


Figure 6.1 : Etapes des traitements des données dans le but d'obtenir des cartes génétiques de liaison. Résultats donnés pour la Famille 1 *Jaera albifrons* et la Famille 3 *J. praehirsuta* (Tableau 6.1).

2.2 Architecture génomique de l'isolement reproductif

La différenciation génétique entre espèces a été mesurée à partir des mêmes échantillons de populations naturelles que ceux qui avaient été utilisés pour l'inférence des histoires démographiques (chapitre 5), soit 50 *J. albifrons* et 39 *J. praeheirsuta* de Bretagne (Bouillennou, Saint-Jean, Lingoz, voir chapitre 2 – Figure 2.1 et Tableau 2.1, sites 8, 9 et 12) ainsi que 40 *J. albifrons* et 38 *J. praeheirsuta* de Normandie (Grandcamp et Sainte-Honorine, sites 30 et 32). Le protocole bio-informatique suivi a été le même que dans le chapitre 2 (voir aussi annexe 2). Un *denovo-map* (-m3 -M3 -N5) a d'abord été réalisé puis les fichiers *sstacks* obtenus ont été mappés sur le catalogue (-n 3) obtenu à partir des familles (décrit dans la première partie de ce chapitre) afin d'avoir le même nom entre les loci obtenus pour estimer la différenciation entre espèces et ceux utilisés dans les cartes génétiques. Le programme *rxstacks* a été utilisé pour faire quelques corrections (Inls -8, voir annexe 2), puis les programmes *cstacks* (-n 3) et *sstacks* ont été relancés avant de faire tourner *populations* (4 populations : *J. albifrons* et *J. praeheirsuta* de Bretagne et de Normandie) pour obtenir un fichier au format *vcf*. *VCFtools* a ensuite été utilisé pour éliminer les loci ayant été génotypés pour moins de 80% des individus contenus dans chaque population. Enfin, les 2 populations de Bretagne ont été réunies, ainsi que les deux de Normandie, et un *mac* (nombre de copies minimum d'un allèle donné) de 3 a été appliqué sur ces deux fichiers. Ainsi en Bretagne, le fichier final comporte 31 474 loci et en Normandie il comporte 21 749 loci, sur lesquels la différenciation entre espèces va pouvoir être estimée.

Le fichier *vcf* ainsi obtenu a ensuite été transformé sous R (Core Team, 2016) en fichier lisible par Genepop V4.6 (Rousset, 2008) avec les packages *vcfR* (Knaus & Grünwald, 2017) et *strataG* (Archer et al., 2017). Genepop a été utilisé pour calculer la différenciation interspécifique F_{ST} (*i.e.* différenciation entre espèces au sein de chaque région) et l'hétérozygotie attendue totale H_T (diversité génétique totale pour les deux espèces confondues, au sein de chaque région) à chaque locus.

Afin de détecter les loci "outliers" ou plus précisément les régions du génome ayant des flux de gènes réduits à cause d'effets sélectifs ou d'une diminution de la recombinaison, le scénario SC2m (contact secondaire avec taux de migration hétérogène le long du génome) détecté par $\delta a \delta i$ (chapitre 5) a été utilisé pour simuler 200 000 SNPs indépendants à l'aide du logiciel *ms* (Hudson, 2002) sous un modèle dit "neutre" (Beaumont & Nichols, 1996, Hudson, 2002). Ce modèle a été obtenu en se basant sur les valeurs obtenues par $\delta a \delta i$ correspondant aux régions du génome à "migration libre", ce qui représente en Bretagne seulement 40% du génome (pour rappel, θ : 22.963, n_1 : 4.543, n_2 : 6.644, M_{12} : 2.601, M_{21} : 1.039, M'_{12} : 0.058, M'_{21} : 0.053, t_s : 7.403, t_{sc} : 0.266, P : 0.390) et 70% du génome en Normandie (pour rappel, θ : 103.642, n_1 : 0.363, n_2 : 1.551, M_{12} : 9.930, M_{21} : 5.811, M'_{12} : 0.387, M'_{21} : 0.536, t_s : 0.784, t_{sc} : 0.094, P : 0.674) (cf. résultats du chapitre 5). Les F_{ST} et H_T

de ces données simulées ont également été obtenus à l'aide de Genepop V4.6 (Rousset, 2008). Notre approche ne correspond donc pas à une détection de loci à très forte (ou très faible) différenciation basée sur un modèle démographique classique (type modèle en îles à l'équilibre mutation-migration-dérive sans sélection), mais sur l'utilisation de simulations imitant la "partie neutre" du scénario démographique de contact secondaire inféré dans le cadre du chapitre 5. Nous nous attendons donc à ce que de nombreux loci ne suivent pas ce scénario neutre. Observer la distribution de ces loci le long de nos cartes génétiques est un des objectifs de ce chapitre.

Par la suite, les statistiques F_{ST} et H_T obtenues à partir des données observées ont été comparées à celles obtenues à partir des données simulées afin d'identifier les loci dont la diversité génétique et la différenciation ne semblent pas compatibles avec des processus neutres avec recombinaison libre. Ce travail a été effectué avec l'ensemble des loci génotypés (31 474 loci en Bretagne, 21 749 loci en Normandie) et a été répété avec un jeu de données réduit constitué des loci génotypés et polymorphes dans les deux régions (soit 13 336 loci communs et polymorphes dans les deux régions) afin d'évaluer la robustesse des résultats. Dans les deux cas les loci ont été considérés comme étant incompatibles avec des processus neutres avec recombinaison libre lorsque leur valeur de F_{ST} n'était strictement jamais obtenue par simulation sous notre modèle neutre (*i.e.* F_{ST} observé > F_{ST} simulé pour une valeur de H_T donnée). Le terme « outlier » sera un raccourci utilisé dans tout le chapitre pour désigner tous les loci pour lesquels le niveau de différenciation génétique interspécifique observé implique un processus qui accentue la divergence entre espèces (recombinaison contrainte, et processus sélectifs directs ou indirects).

Ainsi 3 709 loci « outliers » sur 31 474 loci totaux ont été identifiés en Bretagne, et 1 849/21 749 en Normandie. En ne considérant que les loci communs et polymorphes dans les deux régions, 1 733 loci « outliers » sur 13 336 loci ont été identifiés en Bretagne et 1 181/13 336 en Normandie. Au total, 478 loci communs aux deux régions et "outliers" dans les deux régions ont été détectés. Enfin, les cartes de liaison ont été utilisées pour localiser tous les loci dont la différenciation a été mesurée et qui ont pu être cartographiés (Manhattan plot obtenu sous R (Core Team, 2016)).

3. Résultats

3.1 Polymorphisme chromosomique

3.1.1 Cartes de liaison "de référence"

Onze groupes de liaison ont été identifiés chez les deux sexes chez *J. albifrons*, contre 10 chez le mâle et 12 chez la femelle *J. praehirsuta* avec 356 à 431 marqueurs par carte en fonction du parent et de la famille (Tableau 6.3). Cependant, malgré ces différences de nombres de groupes de liaison, le nombre de loci, la longueur totale de la carte et l'espace moyen entre marqueurs sont similaires

(Tableau 6.3). Les 14 (*J. albifrons*) et 102 (chez *J. praeheirsuta*) loci en communs entre cartes maternelles et paternelles au sein d'une même famille permettent ici de vérifier l'homologie entre sexes. Aucun pattern de réarrangement chromosomique n'est discernable à l'aide de ces cartes, que ce soit au sein d'une famille (ex. pour la famille *J. albifrons* : Figure 6.2) ou entre espèces (non montré), mais celles-ci ont servi de point d'ancrage pour faciliter la construction des cartes plus complètes.

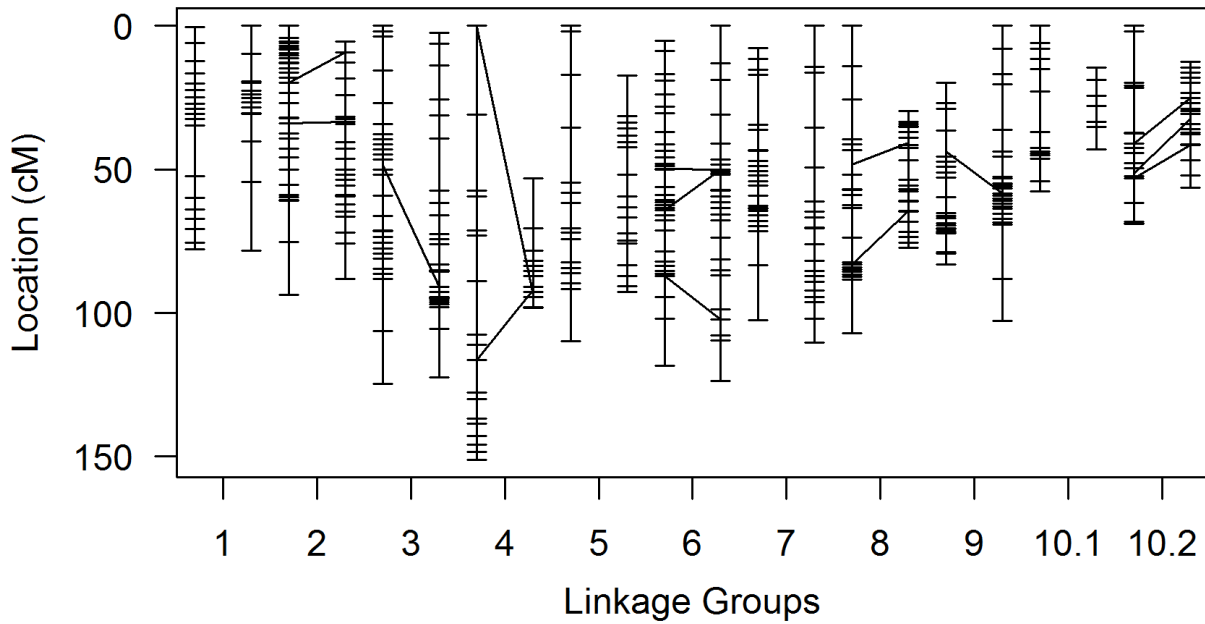


Figure 6.2 : Exemple de carte de référence pour la famille 1 *J. albifrons*. Les groupes de liaison identifiés chez le père (383 loci) et la mère (356 loci) sont représentés deux par deux (groupe paternel à gauche, groupe maternel à droite). Les 14 loci communs aux deux sexes sont reliés.

Tableau 6.3. Résumé des cartes génétiques de référence. Sont donnés pour chaque individu, le nombre de loci utilisés dans les cartes génétiques présentées, le nombre de Groupes de Liaison (GL), la longueur totale de la carte de liaison (en cM) et l'espace moyen entre les loci.

	nb loci	nb GL	longueur carte de liaison	espace moyen entre loci
<i>J. albifrons</i> - mâle	383	11	1061.5	2.9
<i>J. albifrons</i> - femelle	356	11	921.5	2.4
<i>J. praeheirsuta</i> - mâle	395	10	921.5	2.4
<i>J. praeheirsuta</i> - femelle	431	12	1035	2.5

3.1.2 Cartes de liaison "totales"

Au sein de *Jaera albifrons*, on retrouve à nouveau 11 groupes de liaison pour chacun des parents, à partir cette fois de 3 909 loci pour le père et 4 195 loci pour la mère, augmentant le nombre de marqueurs communs à 1 593, et permettant ainsi une comparaison inter-sexes plus détaillée. Étonnamment, une partie des marqueurs localisés au sein du GL 9 de la carte mâle se trouvent localisés sur le GL10.1 de la carte femelle (Figures 6.3 et 6.7).

Au sein de *J. praehirsuta*, on retrouve cette fois 11 GL chez le père (basés sur 3 849 loci), 12 GL chez la mère (basés sur 3 771 loci), et 1 351 marqueurs permettent de faire une comparaison inter-sexe. Là encore une différence de structure entre les deux sexes est visible : le GL 10 du père regroupe des loci, sans ordre particulier apparent (Figure 6.8), provenant des GL 10.1 et 10.2 chez la mère (Figure 6.4).

Le faible nombre de marqueurs en commun entre les quatre cartes ne permet d'analyse comparative de l'ensemble des espèces et sexes à la fois, cependant il est possible de comparer les cartes génétiques interspécifiques paternelles (848 loci en commun entre *J. albifrons* et *J. praehirsuta*) et maternelles (943 loci en commun). Dans l'ensemble on détecte 3 grandes différences entre espèces (Figures 6.5 et 6.6). Chez les individus *J. albifrons*, on ne retrouve qu'un grand GL 6, divisé en deux chez les individus mâle et femelle *J. praehirsuta*. Un réarrangement entre espèces entre les GL 7 et 8 (une partie du GL 7 de *J. albifrons* étant localisé chez le GL 8 de *J. praehirsuta* et inversement) est également visible. Enfin, on peut observer un réarrangement particulier entre les GL 9, 10.1 et 10.2 (ceux-là mêmes qui semblent indiquer des différences entre sexes au sein de chacune des espèces). En effet, chez la femelle *J. praehirsuta* on observe qu'une partie des marqueurs localisés sur le GL 10.1 sont trouvés sur le GL 9 chez la femelle *J. albifrons*. De même, le groupe 10 du mâle *J. praehirsuta* regroupe les GL 10.1 et 10.2 du mâle *J. albifrons*.

Une autre manière de visualiser cette comparaison interspécifique est de regarder la distribution des loci entre les groupes de liaisons (Figures 6.7 et 6.8). La division du GL 6 chez *J. praehirsuta* et les remaniements entre les GL 7 et 8 sont largement visibles, ainsi que les liens entre le GL 9 et le 10.1 chez les femelles, et ceux entre les GL 10.1 et 10.2 chez les mâles. De plus, il semble lorsque l'on regarde la distribution des loci d'un groupe de liaison à l'autre chez les femelles (Figure 6.9) que le GL2 montre une inversion des loci situés sur le centre du groupe de liaison et que le groupe de liaison 3 chez les mâles (Figure 6.10) montre lui une inversion sur l'une de ses extrémités. Il est difficile de distinguer d'autres éventuels patterns de par une lecture inversée de certains groupes de liaisons obligeant une rotation du sens de lecture (*i.e.* numérotation des marqueurs simplement inversée entre cartes, qui sont construites de manière indépendante entre sexes et entre espèces). Ça semble être le cas par exemple au niveau du GL 9 entre le mâle et la femelle *J. praehirsuta* (Figure 6.8).

Enfin, parmi les 615 marqueurs dont la ségrégation n'est compatible qu'avec une liaison au sexe (Tableau 6.2), 384 loci (sur l'ensemble des parents des deux Familles) ont été cartographiés dans les cartes finales, et tous sans exception ont été identifiés sur un seul groupe de liaison (GL 1) identique chez les deux sexes et les deux espèces.

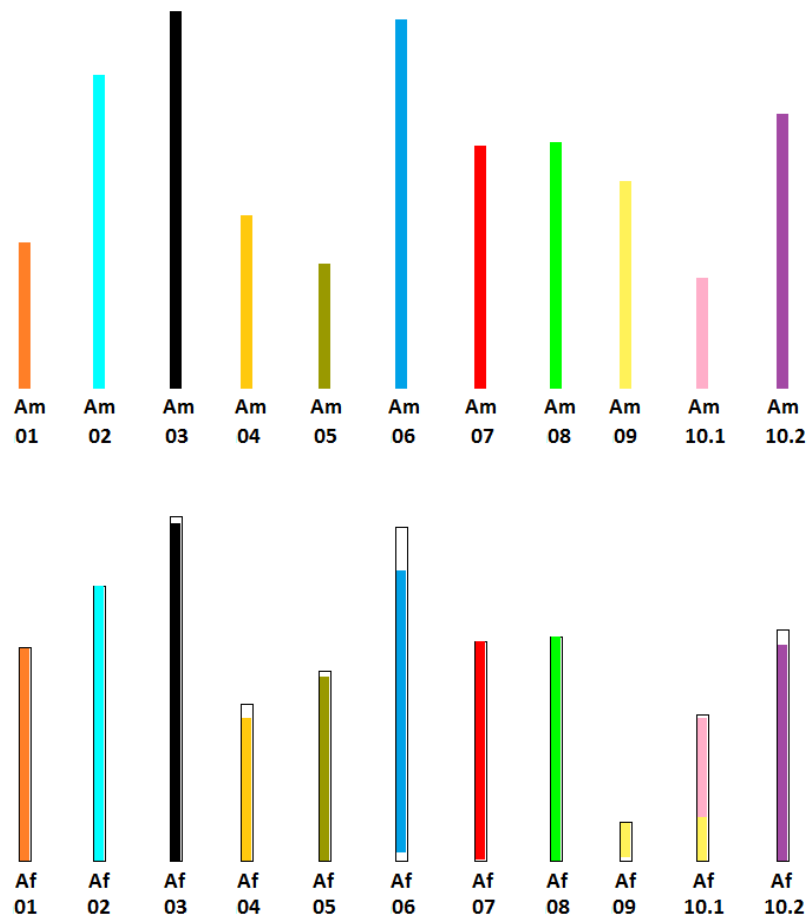


Figure 6.3 : Bar plot représentant les cartes génétiques chez *J. albifrons*, avec en haut le mâle (Am) et en bas la femelle (Af). Chaque groupe de liaison (dont la taille du cadre dépend du nombre de marqueurs) possède sa propre couleur, permettant de détecter d'éventuels réarrangements entre groupes.

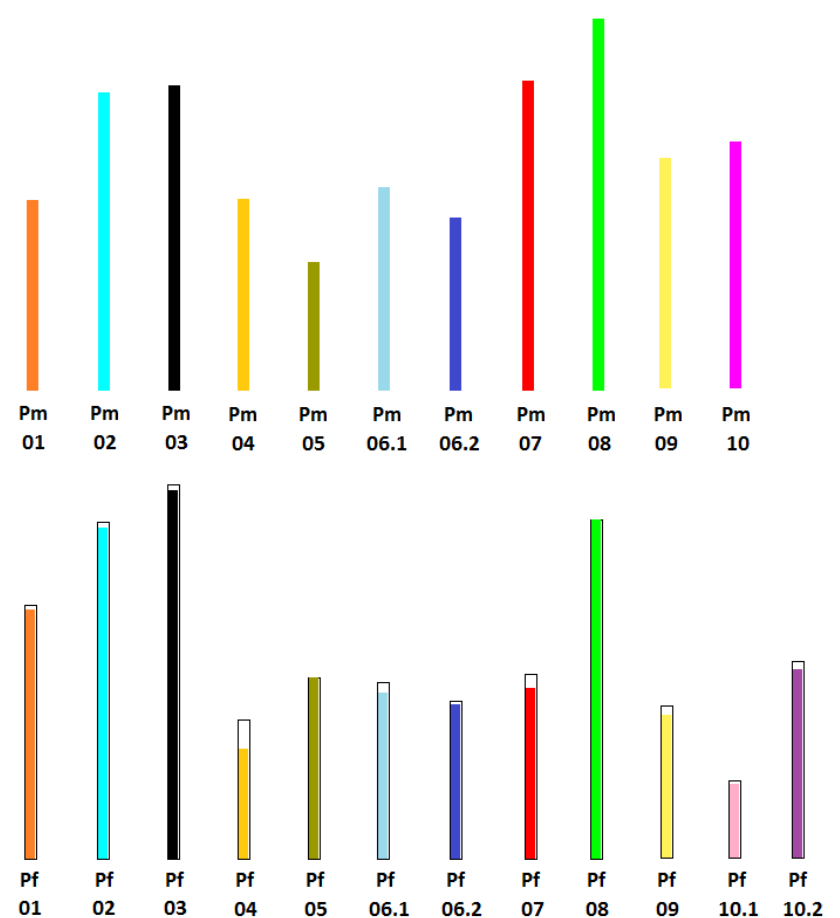


Figure 6.4 : Bar plot représentant les cartes génétiques chez *J. praeheirsuta*, avec en haut le mâle (Pm) et en bas la femelle (Pf). Chaque groupe de liaison (dont la taille du cadre dépend du nombre de marqueurs) possède sa propre couleur, permettant de détecter d'éventuels réarrangements entre groupes. Par exemple Pm 10 a une couleur intermédiaire à Pf 10.1 et Pf 10.2 car il possède des loci en communs avec ces deux groupes de liaison.

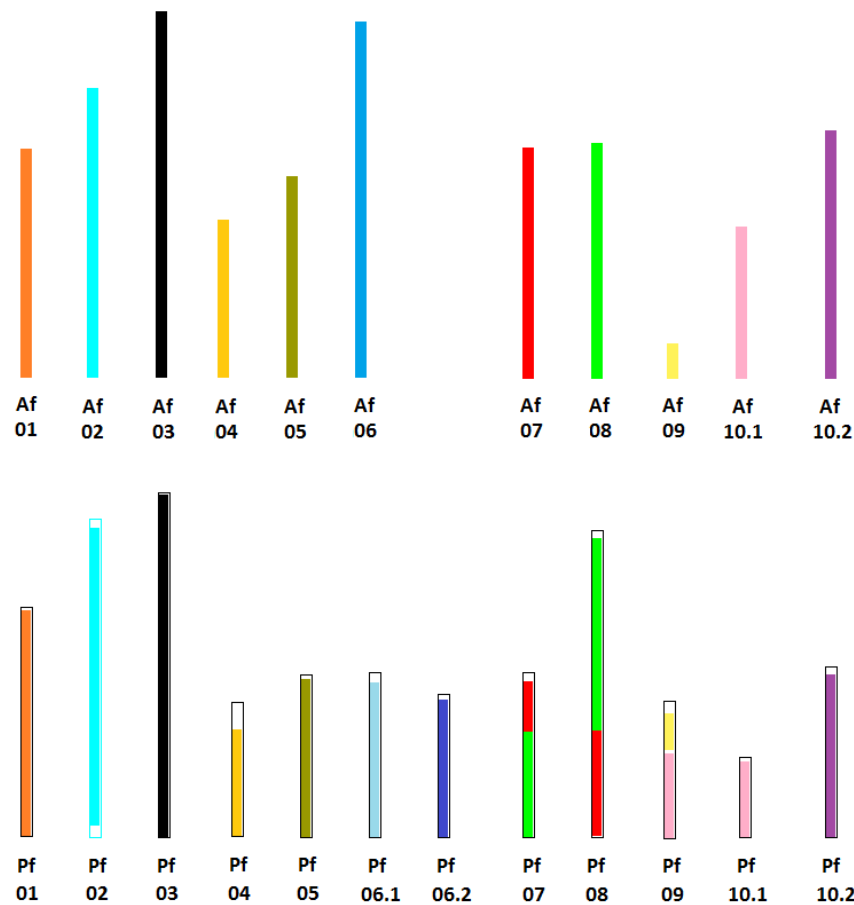


Figure 6.5 : Bar plot représentant les cartes génétiques chez les femelles *J. albifrons* (Af, en haut) et *J. praeirsuta* (Pf, en bas). Chaque groupe de liaison (dont la taille du cadre dépend du nombre de marqueurs) possède sa propre couleur, permettant de détecter d'éventuels réarrangements entre groupes. Par exemple Af 06 a une couleur intermédiaire à Pf 06.1 et Pf 06.2 car il possède des loci en commun avec ces deux groupes de liaison.

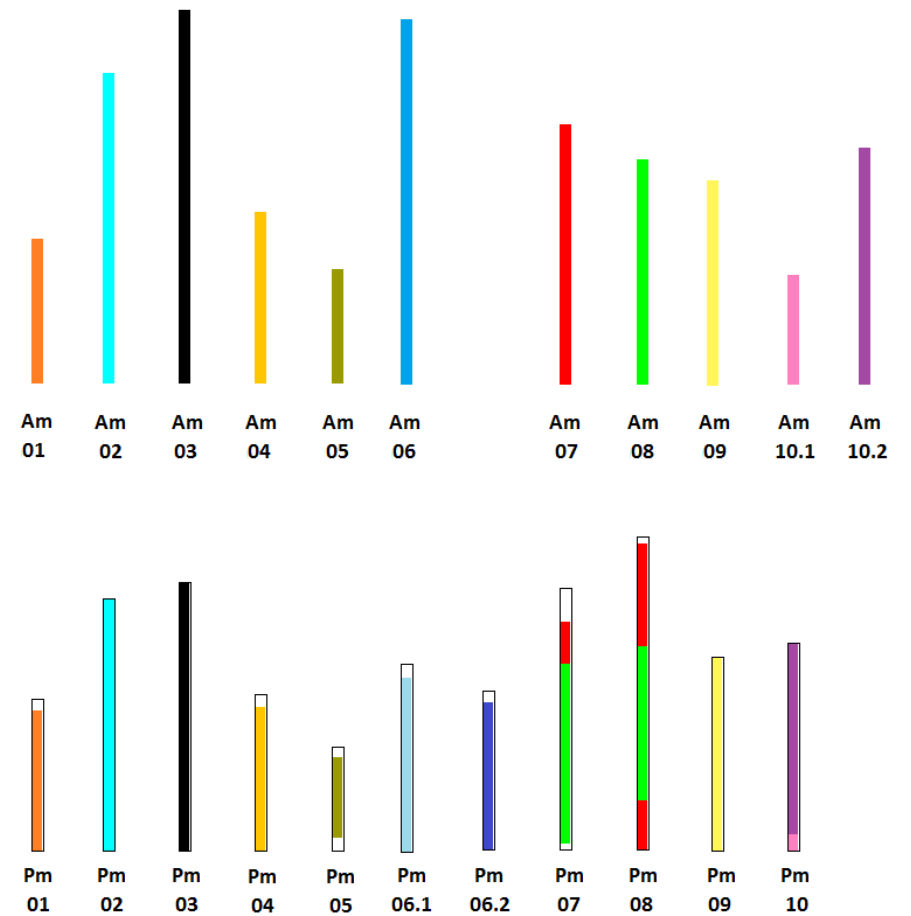


Figure 6.6 : Bar plot représentant les cartes génétiques chez les males *J. albifrons* (Am, en haut) et *J. praeirsuta* (Pm, en bas). Chaque groupe de liaison (dont la taille du cadre dépend du nombre de marqueurs) possède sa propre couleur, permettant de détecter d'éventuels réarrangements entre groupes.

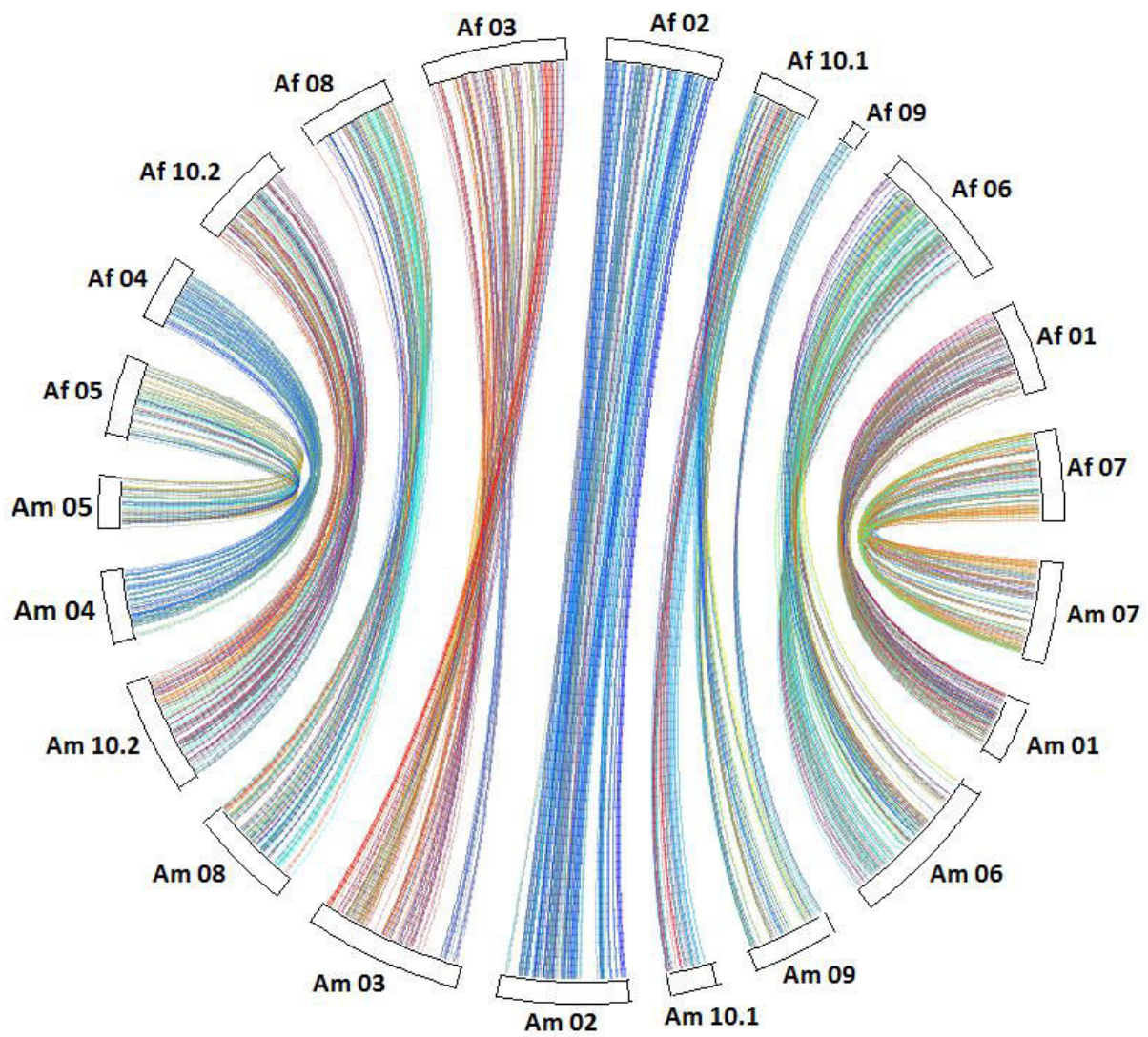


Figure 6.7 : Circle plot représentant les liens entre les marqueurs constituant les différents groupes de liaisons chez la femelle (Af, en haut) et le mâle (Am, en bas) *J. albifrons*.

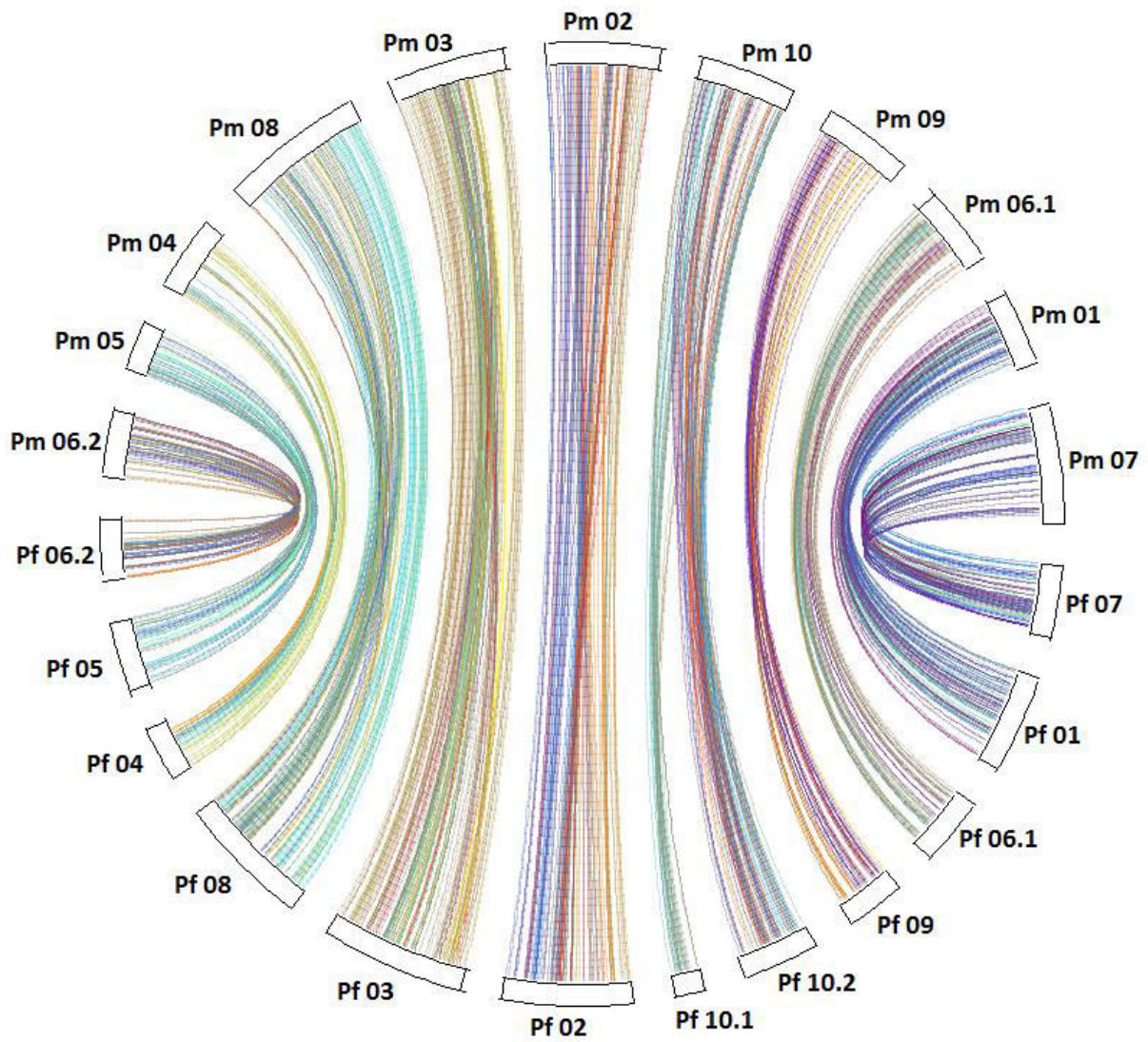


Figure 6.8 : Circle plot représentant les liens entre les marqueurs constituant les différents groupes de liaisons chez le mâle (Pm, en haut) et la femelle (Pf, en bas) *J. praeirsuta*.

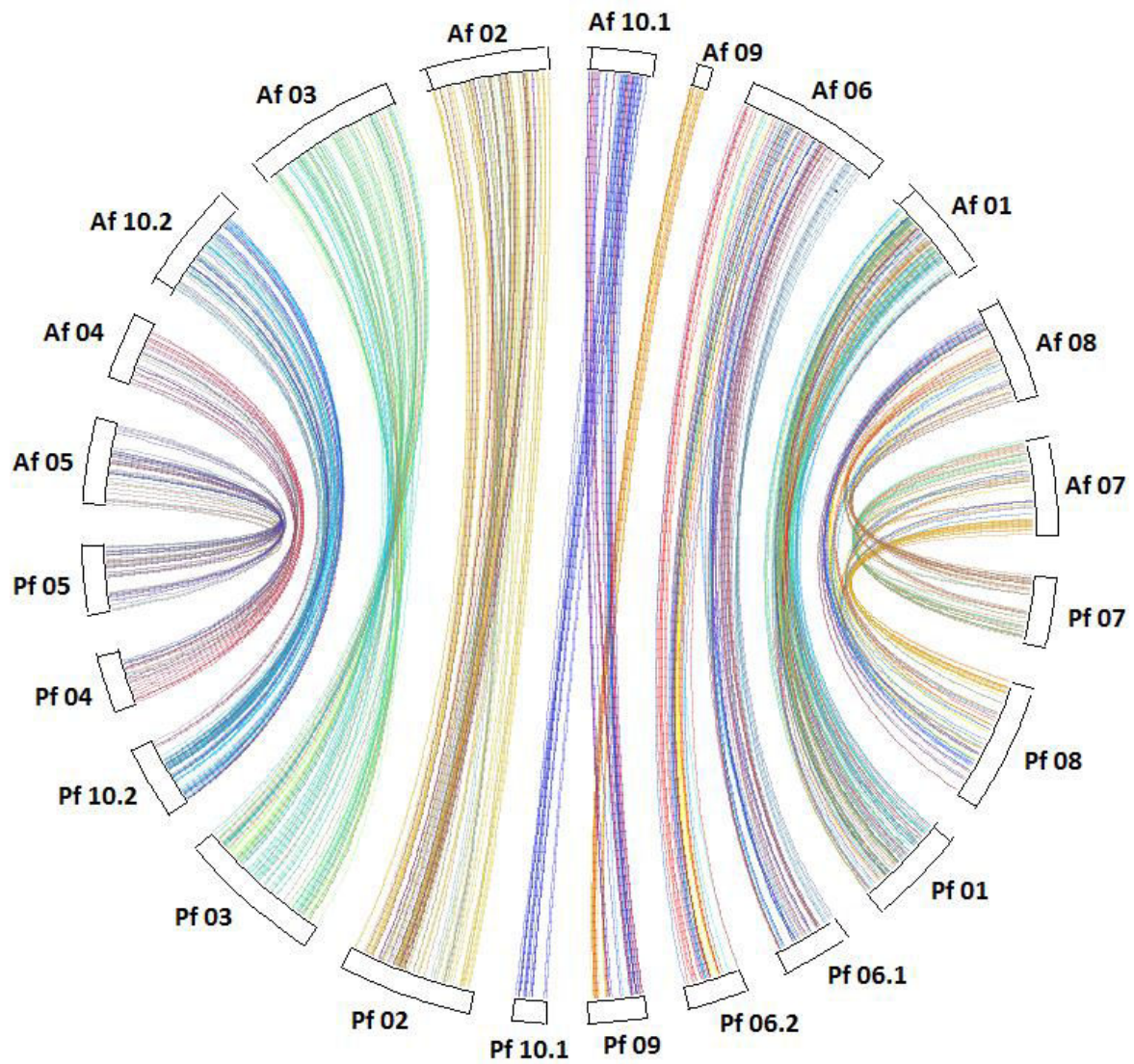


Figure 6.9 : Circle plot représentant les liens entre les marqueurs constituant les différents groupes de liaisons chez les femelles *J. albifrons* (Af, en haut) et *J. praehirsuta* (Pf, en bas).

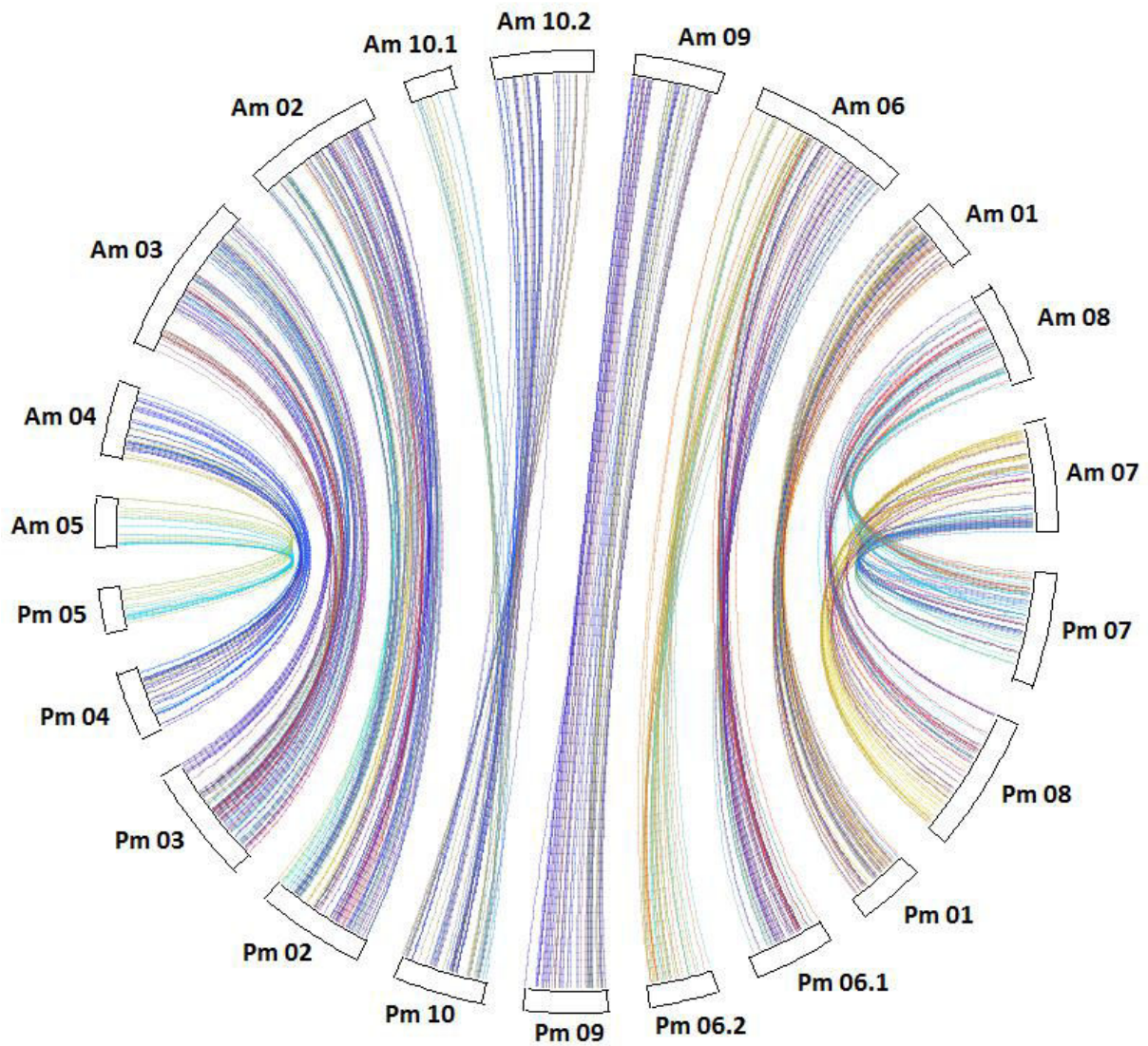


Figure 6.10 : Circle plot représentant les liens entre les marqueurs constituant les différents groupes de liaisons chez les mâles *J. albifrons* (Am, en haut) et *J. praehirsuta* (Pm, en bas).

3.2 Scans génomiques

3.2.1 Détection de loci soumis à migration réduite

Afin de détecter les loci pour lesquels le niveau de différenciation génétique interspécifique observé implique un processus qui accentue la divergence entre espèces, la distribution de la différenciation génétique (estimée via le calcul de F_{ST}) en fonction de la diversité génétique (estimée par H_T) a été construite à l'aide de 200 000 loci "neutres" simulés selon les scénarios inférés pour la Bretagne et la Normandie (Figures 6.11a et 6.11b). Cette distribution a ensuite été comparée aux données réelles observées en Bretagne (31 474 SNPs, Figure 6.12a) et en Normandie (21 749 SNPs, Figure 6.12b). Les loci dont les valeurs observées de F_{ST} et de H_T n'ont jamais été obtenues par simulation sous modèle neutre ont été considérés comme « outliers ». La différenciation le long du génome a été observée et un intérêt particulier a été porté à l'emplacement de ces loci « outliers ». 478 « outliers » sur 13 336 loci en communs dans les deux régions ont donc été détectés (Figure 6.13). Cette méthode permet ainsi de déceler les loci impliqués dans la divergence entre espèces, et de les distinguer des loci à fort F_{ST} pouvant être obtenus de manière neutre avec recombinaison libre.

3.2.2 Positionnement des loci sous sélection le long du génome

Le scan génomique obtenu pour la Bretagne à partir de la carte génétique femelle *J. albifrons* (Figure 6.12) permet d'observer une forte différenciation entre espèces répartie de manière assez homogène sur l'ensemble du génome (F_{ST} moyen = 0.137). On peut toutefois noter le groupe de liaison 1 qui présente une très forte différenciation (F_{ST} moyen = 0.175) et le groupe de liaison 9 qui, à l'inverse, semble montrer une plus faible différenciation entre espèces (F_{ST} moyen = 0.04) et une absence de marqueurs outliers.

Ces deux mêmes patrons sont observables dans la différenciation inter-spécifique mesurée en Normandie (représentée en utilisant la carte génétique femelle *J. prae-hirsuta*, Figure 6.15), avec au groupe de liaison 9 la partie de gauche faiblement différenciée correspondant au GL 9 de *J. albifrons* et la partie de droite fortement différenciée correspondant au GL 10.1 de *J. albifrons*. En revanche l'ensemble du génome est globalement moins différencié qu'en Bretagne (F_{ST} moyen = 0.114), avec 4 groupes de liaison dont la différenciation génétique est compatible avec un processus neutre (GL 2, 4, 6.2 et 10.2). Les groupes 3, 5 et 10.1 montrent une différenciation un peu plus forte avec quelques loci outliers, mais ce sont les groupes 1, 6.1, 7, 8 et 9 qui montrent la différenciation interspécifique la plus forte. Des résultats identiques ont été obtenus avec un jeu de données réduit aux loci communs et polymorphes en Bretagne et en Normandie (non montré).

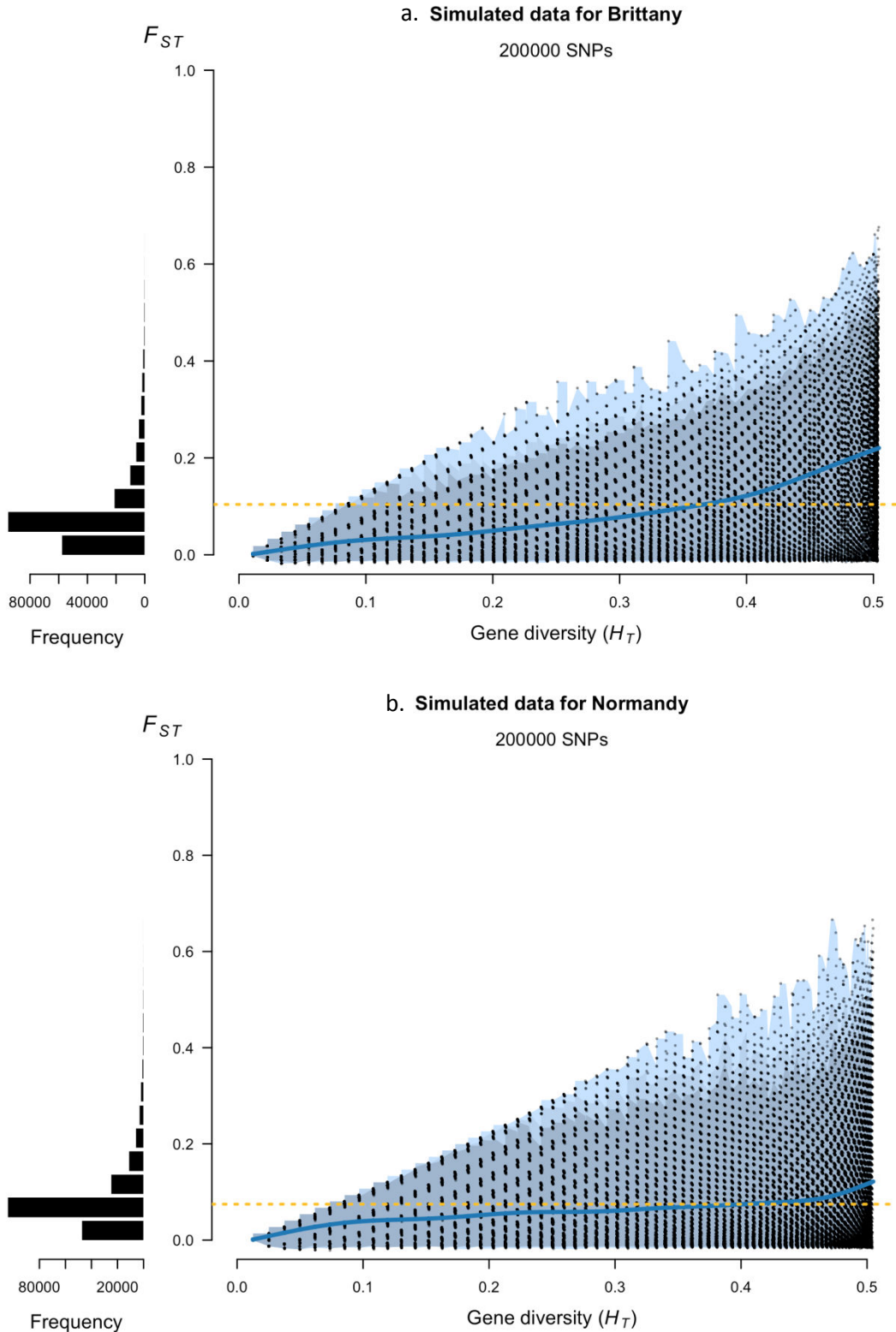


Figure 6.11 : Distribution de la différenciation génétique (F_{ST}) de 200 000 SNPs simulés sous modèle neutre avec recombinaison libre en Bretagne (a) et en Normandie (b) en fonction de la diversité totale (H_T). Les histogrammes à gauche représentent la distribution des valeurs de F_{ST} locus par locus. En jaune est tracé le F_{ST} moyen sur l'ensemble des loci, et en bleu un modèle non linéaire de F_{ST} en fonction de la diversité génique H_T (modèle additif généralisé GAM, utilisé ici pour illustrer la tendance d'évolution de F_{ST} en fonction de H_T). Le dégradé de bleus représente les enveloppes contenant 95 %, 99% et 100 % des données.

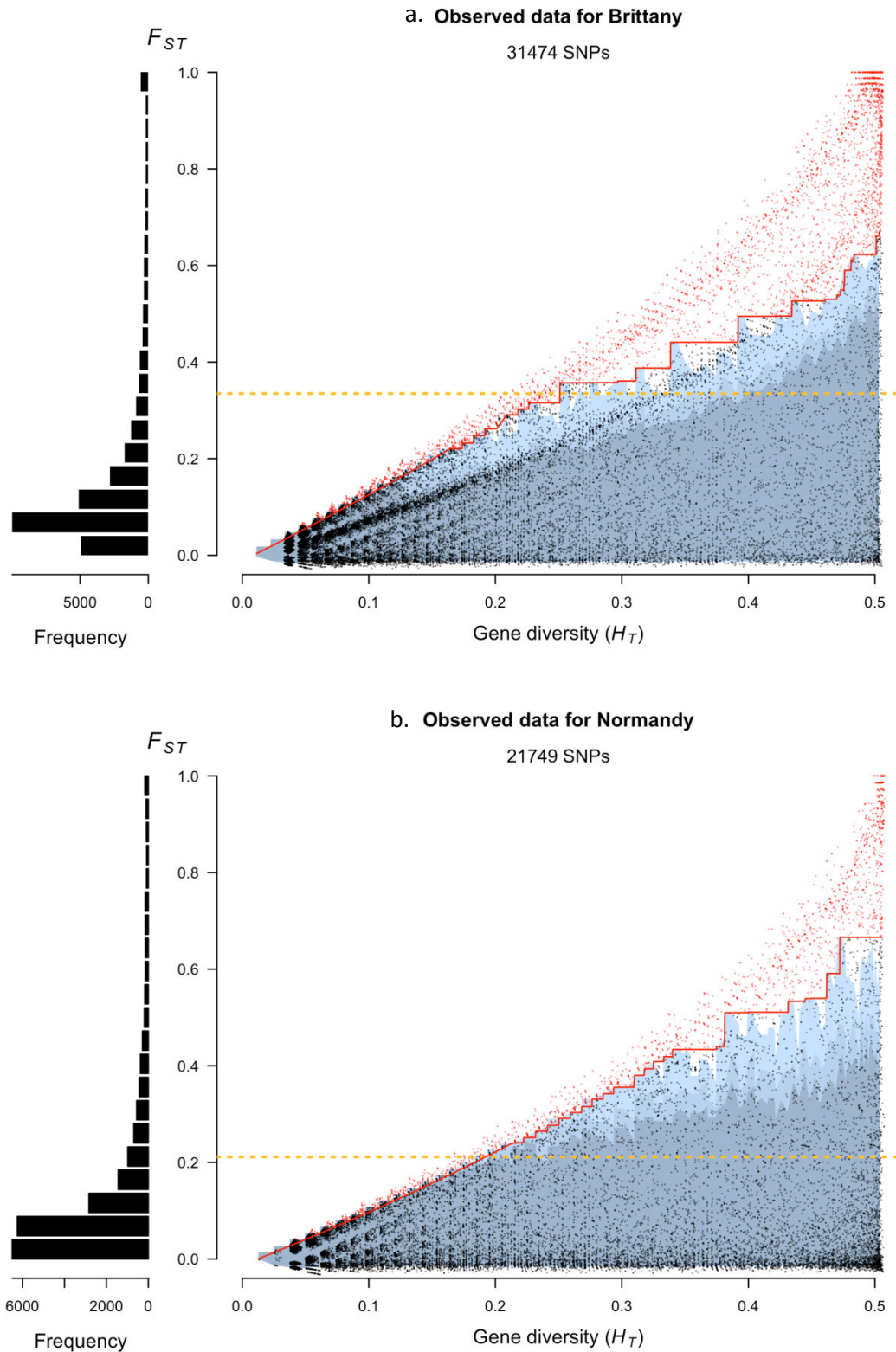


Figure 6.3 : Distribution de la différenciation génétique (F_{ST}) observée à 31 474 SNPs en Bretagne (a) et 21 749 SNPs en Normandie (b) en fonction de la diversité totale (H_T). Les histogrammes à gauche représentent la distribution des valeurs de F_{ST} locus par locus. En jaune est tracé le F_{ST} moyen sur l'ensemble des loci. Le dégradé de bleus représente les enveloppes contenant 95 %, 99% et 100 % des données obtenues par simulation (Figure 6.11). Les points rouges soulignent les valeurs de différenciation observées qui sont strictement supérieures aux maxima obtenus par simulation.

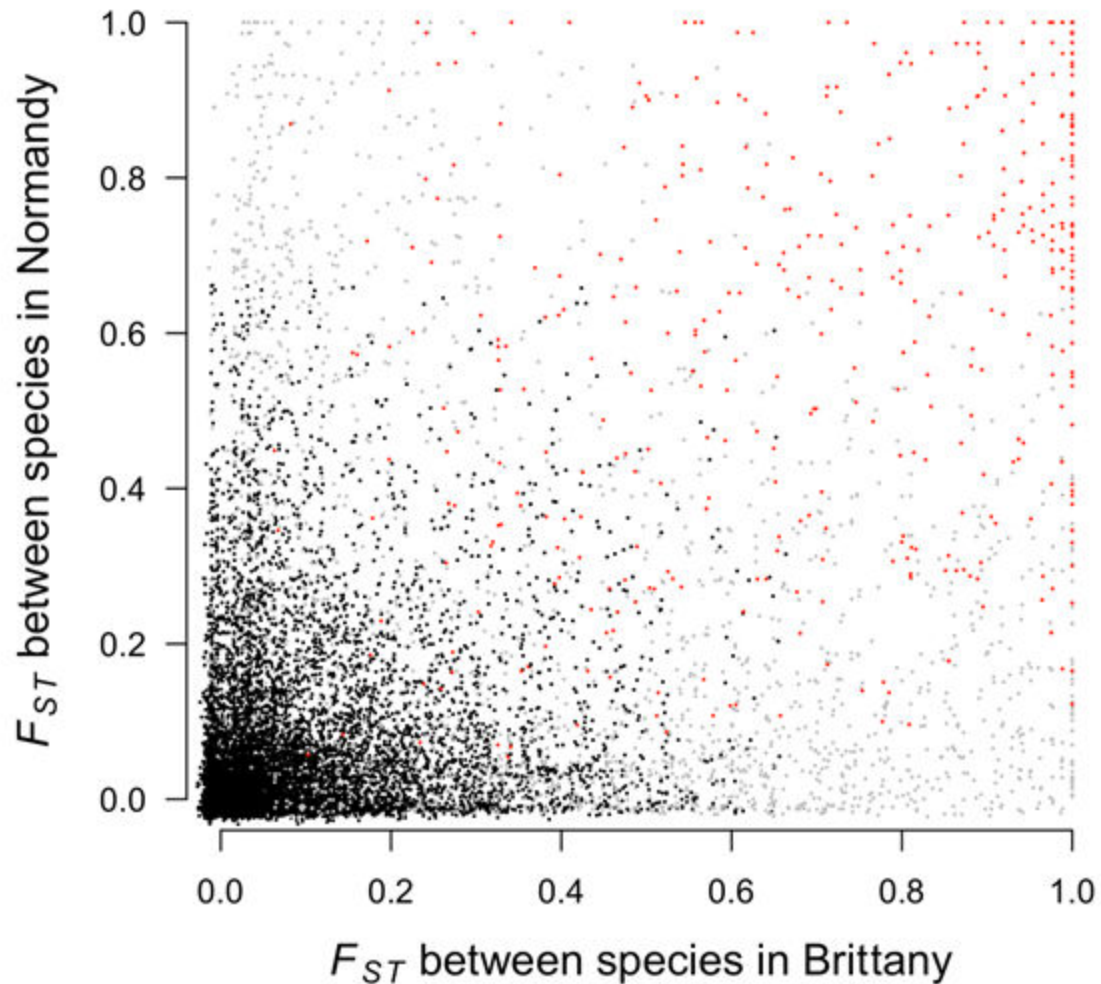


Figure 6.13 : Représentation de la différenciation génétique interspécifique (F_{ST}) en Bretagne et en Normandie, sur l'ensemble des 13 336 loci communs à ces deux régions. En noir sont représentés les marqueurs dont les valeurs de différenciation sont compatibles avec un modèle neutre avec recombinaison libre ($n = 10\,900$), en gris ceux « outliers » dans une seule région ($n = 1\,255$ en Bretagne et $n = 703$ en Normandie), et en rouge, ceux « outliers » dans les deux régions à la fois ($n=478$).

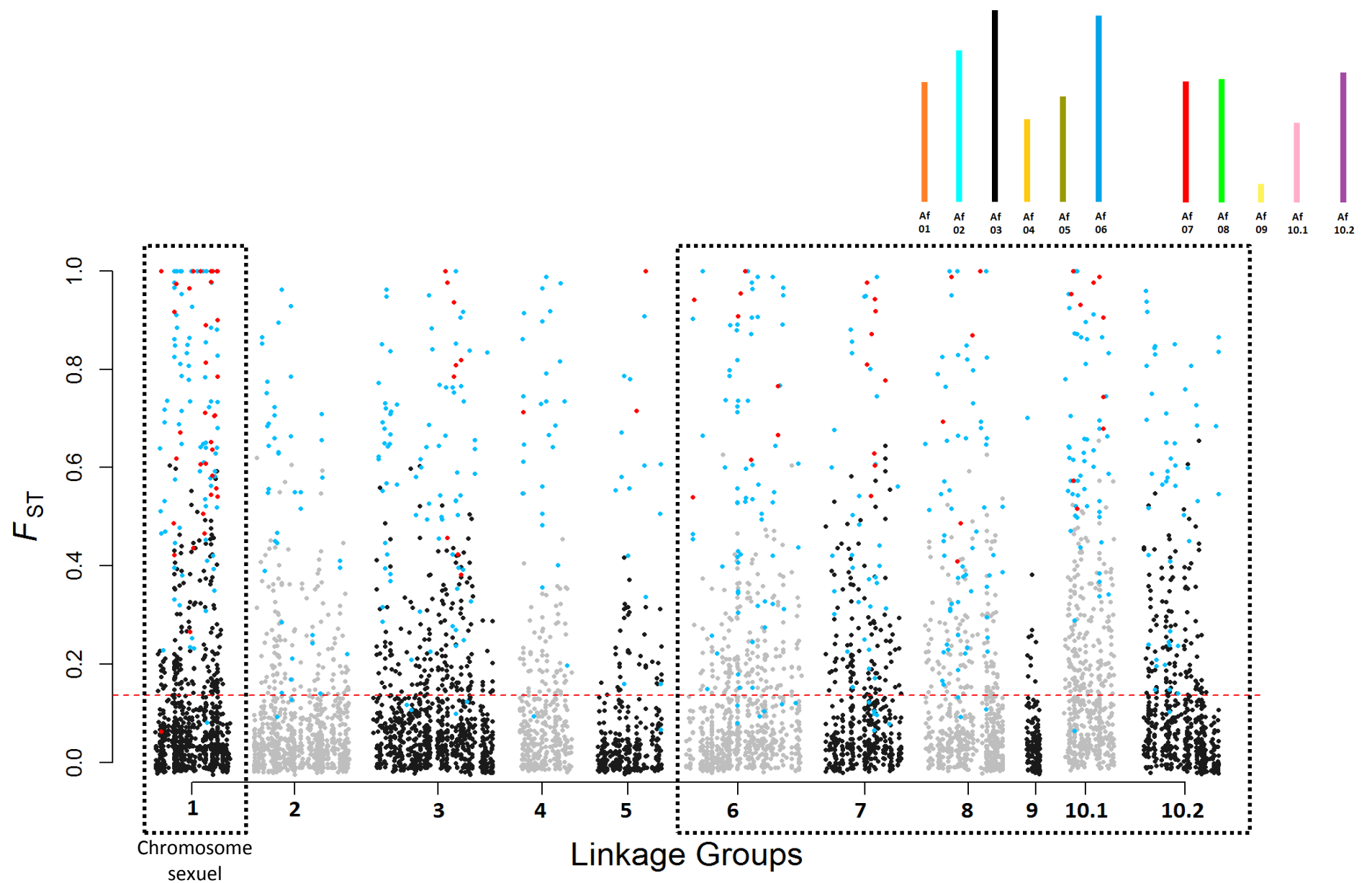


Figure 6.14 : Scan génomique de la différenciation entre espèces (F_{ST}) dans la région Bretagne basé sur la carte génétique de la femelle *J. albifrons*. Les marqueurs sont distribués en fonction de leur emplacement sur le groupe de liaison. Les points noirs représentent les loci dont les valeurs de F_{ST}/H_T sont compatibles avec un modèle neutre avec recombinaison libre (cf. Figure 6.12), ceux en bleus étant identifiés comme outliers en Bretagne et ceux en rouge outliers à la fois en Bretagne et en Normandie. La ligne rouge représente le F_{ST} moyen sur l'ensemble des loci. Sont encadrés les groupes de liaison fortement différenciés ou montrant des réarrangements chromosomiques (cadre de droite). En haut à droite se trouve un rappel du bar plot de la Figure 6.5 pour la femelle *J. albifrons*.

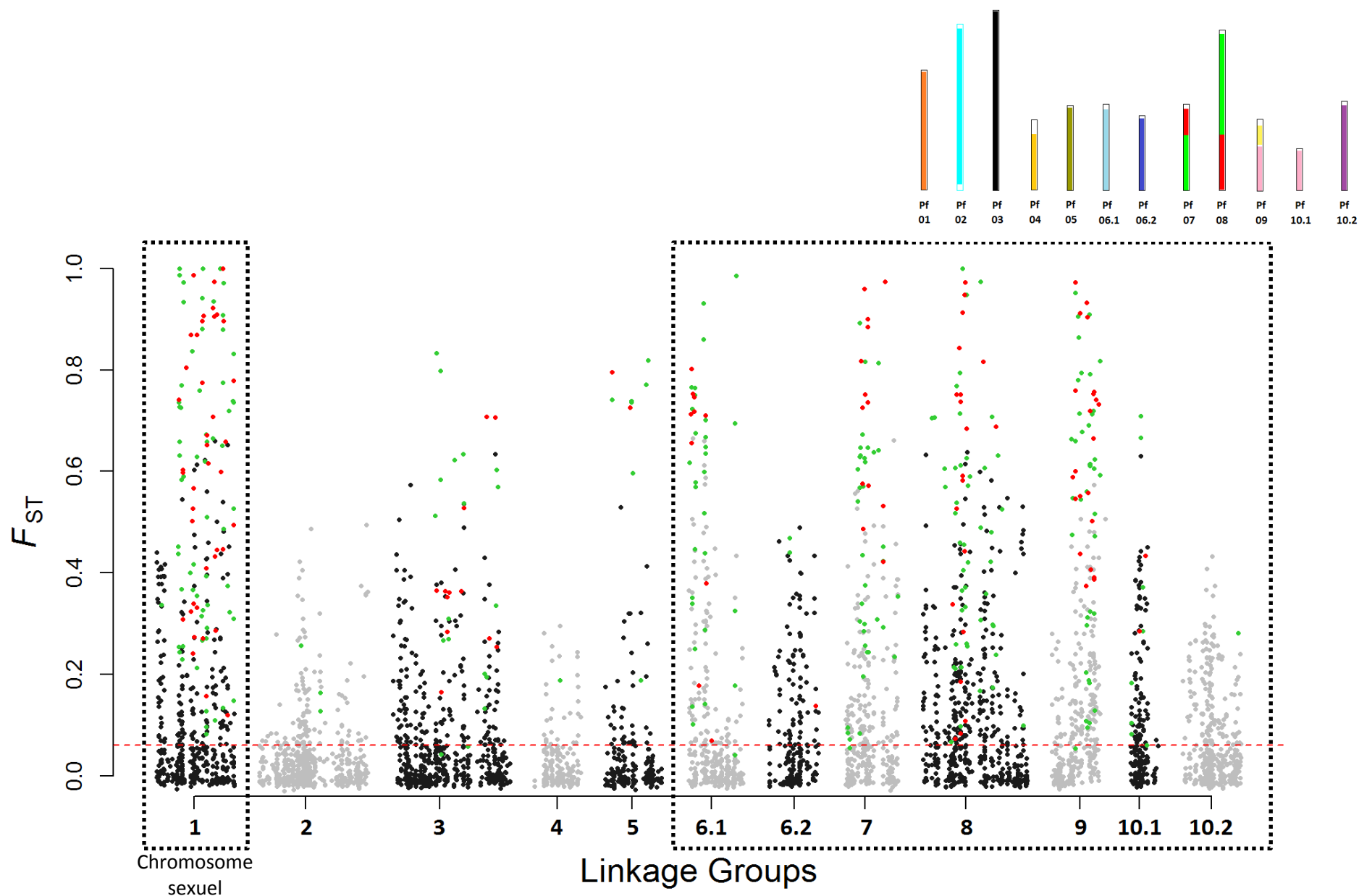


Figure 6.15 : Scan génomique de la différenciation entre espèces (F_{ST}) dans la région Normandie basé sur la carte génétique de la femelle *J. praehirsuta*. Les marqueurs sont distribués en fonction de leur emplacement sur le groupe de liaison. Les points noirs représentent les loci dont les valeurs de F_{ST}/H_T sont compatibles avec un modèle neutre avec recombinaison libre (cf. Figure 6.12), ceux en verts étant identifiés comme outliers en Normandie et ceux en rouge outliers à la fois en Bretagne et en Normandie. La ligne rouge représente le F_{ST} moyen sur l'ensemble des loci. Sont encadrés les groupes de liaison fortement différenciés ou montrant des réarrangements chromosomiques (cadre de droite). En haut à droite se trouve un rappel du bar plot de la Figure 6.5 pour la femelle *J. praehirsuta*.

4. Discussion

Ce chapitre avait deux objectifs principaux : i) réaliser des cartes de liaison à haute densité afin de comparer la structure chromosomique de nos deux espèces d'intérêt et ii) décrire la distribution génomique de la différenciation entre espèces et éventuellement identifier des régions du génome pouvant être liées à l'isolement reproductif entre espèces.

4.1 Structure des groupes de liaison

Au cours de ce chapitre un nombre variable de chromosomes a été détecté, soit 11 groupes de liaison chez les individus *J. albifrons* et le mâle *J. praehirsuta*, mais 12 GL chez la femelle *J. praehirsuta*. Ce nombre de chromosomes et la variabilité intra-espèce est compatible avec ce qui est connu dans la littérature puisque il a été décrit que chez toutes les espèces du complexe *J. albifrons* de nombreux remaniements pouvaient avoir lieu, même entre individus d'une même espèce et d'un même site (Prunus, 1968). Par ailleurs, l'identification des chromosomes sexuels confirme le système ZW inféré par Staiger and Bocquet (1954) d'après l'observation de caryotypes.

Les différences de structure des caryotypes au sein des espèces entre sexes peuvent donc soit être dues à des erreurs de cartographie liées au faible nombre de juvéniles, ou alors témoigner d'un polymorphisme intraspécifique réel (qui ne serait pas lié au sexe).

La limite de ce travail est le nombre de juvéniles obtenus dans chaque famille, qui détermine directement la précision des estimations de recombinaison méiotique et donc la position relative des loci le long des groupes de liaison. Le niveau de précision atteint avec nos cartes n'est pas clairement établi, puisque les différents *runs* montraient une forte consistance dans la détermination des loci sur les groupes de liaisons, mais quelques variations possibles dans l'emplacement de certains de ces loci le long du groupe de liaison. Ce travail est en train d'être poursuivi avec plus de données (familles 2 et 4) et une méthode alternative permettant notamment d'intégrer la possibilité d'erreurs de génotypage (Lep-MAP, Rastas et al., 2013).

Plusieurs indices témoignent en revanche d'une structure des marqueurs entre groupes (*i.e.* nombre et identification des GL) relativement robuste : les différents *runs* réalisés, l'identification sans aucune exception du groupe de liaison 1 comme chromosome sexuel, et également la distribution non aléatoire de la différenciation (cf. ci-dessous). Il est donc possible d'utiliser ces cartes pour détecter le polymorphisme chromosomique entre espèces et d'éventuels remaniements.

En effet, de nombreux remaniements chromosomiques ont ici été détectés par comparaison inter- et intra-espèces des cartes génétiques. Tout d'abord on peut observer un net

réarrangement entre les GL 7 et 8, de type translocation réciproque chez *J. praeheirsuta*. Ensuite, chez les individus *J. albifrons*, on ne retrouve qu'un grand GL 6, qui est divisé en deux chez les individus mâle et femelle *J. praeheirsuta* (GL 6.1 et 6.2), compatible avec un phénomène de fusion robertsonienne (ou fission). Enfin, un réarrangement particulier est également visible entre les groupes de liaison 9, 10.1 et 10.2, trois groupes impliquant des différences complexes à la fois au sein des espèces et entre espèces. Les GL 10.1 et 10.2, trouvés chez les individus *J. albifrons* et chez la femelle *J. praeheirsuta*, sont réunis en un unique groupe de liaison (GL 10) chez le mâle *J. praeheirsuta*, induisant une différence dans le nombre de groupes de liaison entre mâle et femelle *J. praeheirsuta*. Cela pourrait également correspondre à une fusion ou fission à l'instar du GL 6. Dans les deux cas cela peut être dû à i) une erreur d'assemblage des groupes de liaison dû à un manque de juvéniles ii) du polymorphisme robertsonien. Concernant le GL9, il semble y avoir eu translocation non réciproque telle qu'un morceau du GL 9 du mâle *J. albifrons* correspond au GL 10.1 de la femelle. Ce remaniement est également compatible avec la différence observée dans la structure de ces groupes entre femelles *J. albifrons* et *J. praeheirsuta*. Ces cartes de liaison vont permettre d'analyser la distribution de la différenciation génétique le long du génome, au moins en comparant la différenciation entre groupes de liaison à la lumière des remaniements discutés ci-dessus, si ce n'est à l'échelle fine de la position de chaque marqueur (pour laquelle l'imprécision est, peut-être, plus grande).

4.2 Distribution génomique de la différenciation génétique entre espèces

Bien que les simulations montrent que des F_{ST} très élevés (*e.g.* >0.6) peuvent être obtenus sous un modèle démographique ne faisant intervenir aucun effet sélectif et des flux de gènes forts entre espèces (notamment en Normandie, Figure 6.11), les données observées montrent qu'un grand nombre de loci ont des niveaux de différenciation interspécifiques encore plus élevés (points rouges dans les Figures 6.12a et 6.12b). Cela était attendu d'après les résultats du chapitre 5 et la méthode employée ici pour les simulations (basée sur les paramètres démographiques inférés pour la fraction du génome n'étant pas soumise à migration réduite ou à des effets Hill-Robertson). Il est donc possible d'identifier ces loci et d'essayer d'analyser leur distribution, afin de détecter un éventuel pattern de regroupement de ces loci le long du génome, voire d'identifier de possibles îlots de différenciation. L'objectif ici n'était pas de compter le nombre d'outliers (la méthode même, par construction, devrait en révéler un nombre qui correspond à la fraction du génome non prise en compte pour les simulations), mais bien d'identifier ces loci pour observer leur localisation.

Comme discuté plus haut, il n'est pas certain que la position relative des loci obtenue par analyse de recombinaison soit suffisamment précise pour une analyse fine des scans génomiques. Cependant, plusieurs résultats intéressants émergent de l'analyse de la distribution de la différenciation entre groupes de liaison.

Le premier résultat le plus intéressant de cette étude est le pattern général observé dans les deux régions : le groupe de liaison 1 se distingue par sa forte différenciation interspécifique et la présence de nombreux loci outliers, un patron d'autant plus remarquable en Normandie. Cela est intéressant puisqu'il s'agit du groupe de liaison correspondant aux chromosomes sexuels. Il est prédit que la sélection naturelle soit en effet plus efficace sur les chromosomes sexuels hétérogamétiques car une sélection positive agit directement sur les mutations bénéfiques. Ainsi, les chromosomes sexuels jouerait un large rôle dès les premières étapes de la divergence (Ellegren et al., 2012, Mank et al., 2007). Il est cependant difficile de dire si (et pourquoi) l'ensemble du chromosome résiste à l'introgession, bien qu'il s'agisse d'une possibilité.

Au-delà de ce groupe de liaison, on observe en Bretagne une différenciation forte et distribuée de manière homogène tout le long du génome (c'est-à-dire sur tous les groupes de liaisons), sauf sur le groupe de liaison 9, qui à l'inverse montre une très faible différenciation. Étant donné que nous avons utilisé la carte *J. albifrons* femelle comme référence pour visualiser le scan génomique Breton, ce groupe de liaison 9 correspond en fait à une petite fraction homologue d'un groupe de liaison plus grand et identique entre la carte paternelle *J. albifrons* et les deux cartes *J. praeheirsuta* (GL 9 sur les Figures 6.7 à 6.10). Seule la petite partie du GL9 qui semble commune à toutes les cartes montre une faible différenciation en Bretagne. De plus, cela reste vrai sur le scan génomique Normand : la partie commune du GL9 (liens jaunes sur la Figure 6.9) montre relativement peu de différenciation entre espèces en Normandie (début du GL9 sur la Figure 6.15). Peut-être cette petite section de chromosome est-elle le siège d'un ou plusieurs loci soumis à des effets sélectifs qui limitent la divergence entre espèces aussi bien en Bretagne qu'en Normandie. Il est en effet difficile d'imaginer quel processus neutre pourrait limiter la différenciation en une région précise du génome dans une population (Bretagne) où les flux de gènes entre espèces sont globalement faibles, et s'accompagnent d'une différenciation élevée sur le reste du génome.

La situation Normande dans son ensemble contraste fortement avec le scan génomique obtenu en Bretagne. La différenciation est globalement moins forte mais distribuée de manière très hétérogène en Normandie. En particulier, si certains groupes de liaison montrent un niveau de

différenciation assez élevée (GL 6.1, 7, 8, 9, et bien sûr le GL1) et comparable à la situation Bretonne, des groupes de liaison entiers montrent un niveau de différenciation beaucoup plus faible et une absence totale d'outliers (donc pas de trace d'effets sélectifs ou de recombinaison réduite impliqués dans la divergence entre espèces). Il s'agit des GL 2, 4, 6.2 et 10.2 au sein desquels les valeurs de différenciation mesurées sont parfaitement compatibles avec une dynamique neutre avec taux de migration interspécifique libre tel qu'inféré dans le chapitre 5. Il y a également quelques groupes de liaison (GL3, 5, 10.1) dont la situation semble être intermédiaire (quelques *outliers* sont positionnés sur ces GL mais en nombre bien moins important qu'en Bretagne).

Ceci nous amène au second résultat le plus marquant de cette étude : la variation dans la distribution de la différenciation entre GL est pour une bonne part (mais pas entièrement) corrélée aux remaniements décrits ci-dessus. En effet, les groupes les plus fortement différenciés dans les deux régions sont soit liés aux chromosomes sexuels, soit liés à des réarrangements chromosomiques (GL 6.1, 7, 8, 9). Certains réarrangements pouvant résulter en un arrêt de la recombinaison entre chromosomes homologues remaniés, il est plausible qu'une partie de la différenciation interspécifique observée (notamment en Normandie) résulte de flux géniques interspécifiques réduits à cause de ces freins à la recombinaison. Par exemple, chez les individus *J. albifrons*, on ne retrouve qu'un grand GL 6, qui est divisé en deux chez les individus mâle et femelle *J. praehirsuta* (GL 6.1 et 6.2), signe d'une fusion robertsonienne (ou d'une fission). Parmi ces deux groupes le GL 6.1 est fortement différencié alors que le GL 6.2 l'est faiblement, ce qui pourrait être signe d'une diminution de la recombinaison au groupe GL 6.1, pouvant être due à une inversion (Faria & Navarro, 2010, Rieseberg, 2001). De même, une inversion paracentrique semble être présente au niveau du GL 3 (Figure 6.10). Or, sur le scan génomique, une partie est fortement différenciée (droite) par rapport à l'autre partie (gauche). La partie fortement différenciée pourrait être due à l'inversion visible sur le Circle Plot, diminuant la recombinaison, et donc augmentant la différenciation (Rieseberg, 2001, Navarro & Barton, 2003, Dagilis & Kirkpatrick, 2016).

Ces remaniements n'expliquent cependant pas tout le paysage génomique de la différenciation (par exemple les GL 5, 6.2 et 10.2 ne suivent pas le schéma édicté ci-dessus), et, surtout, ce paysage peut résulter d'une combinaison de facteurs divers. L'interprétation n'en est donc pas triviale. Ainsi, si une forte sélection est en action dans de petites populations, les régions liées au locus sélectionné peuvent se différencier par effet de *hitchhiking* (Feder & Nosil, 2010, Via & West, 2008, Southcott & Kronforst, 2017), ou de sélection d'arrière-plan (*e.g.*, Stephan, 2010).

La présence d'îlots génomiques de forte différenciation interspécifique n'est pas nécessairement due à des barrières à la reproduction (Turner et al., 2005, Harr, 2006), et, de même, des zones à faibles F_{ST} induits par une ségrégation de la variation ancestrale peuvent être détectées même en absence complète de flux de gènes interspécifique (Noor & Bennett, 2009, Wolf & Ellegren, 2017).

Ce dernier effet peut cependant être exclu dans notre cas, et c'est là l'un des intérêts majeurs de la comparaison entre régions avec plus ou moins d'hybridation : les régions du génome qui sont faiblement différenciées en Normandie mais pas en Bretagne ne sauraient témoigner d'un polymorphisme ancestral partagé malgré une absence de flux de gènes. Il est bien plus probable que la faible différenciation dans ces régions soit due à l'action différentielle de l'introgession, c'est-à-dire aux flux de gènes interspécifiques. Ainsi ces scans génomiques semblent bien confirmer la perméabilité différentielle aux flux de gènes en Normandie (telle qu'esquissée au chapitre 1 avec une vingtaine de marqueurs). L'interprétation la plus parcimonieuse de nos résultats est que l'hybridation introgressive détectée en Normandie se traduit par une érosion de la différenciation interspécifique qui est efficace dans une certaine partie du génome (GL 2, 4, 6.2 et 10.2 : différenciation compatible avec des processus neutres et recombinaison libre) mais freinée, voir largement interdite, dans une autre partie du génome (et cette partie est très conséquente, comme suggéré dès le chapitre 1 par le fait que deux microsatellites sur 23 sont fortement différenciés). L'un des freins en jeu est, très probablement, la variation du niveau de recombinaison générée par la ségrégation au sein et entre espèces de chromosomes remaniés, mais il est difficile d'aller plus loin à ce stade et d'identifier dans le génome des effets sélectifs, directs ou indirects.

Cependant, l'ensemble de ce travail est basé sur l'identification spécifique d'individus mâles classés dans deux groupes sur la base de leurs traits sexuels secondaires : *Jaera albifrons* d'un côté, *Jaera praeheirsuta* de l'autre. Et cette classification est possible même en Normandie malgré l'introgession. Donc même si cela n'est pas obligatoire, il est très plausible que tout ou partie des gènes ayant un impact direct sur ces phénotypes soient localisés dans une ou plusieurs des régions que nous voyons résister à l'introgession. Certains de ces gènes pourraient être localisés sur les chromosomes sexuels par exemple, comme cela a été démontré chez d'autres espèces (Sæther et al., 2007, Kitano et al., 2009). Et il est certain que l'arrêt de la recombinaison engendré par un remaniement chromosomique pourrait faciliter l'évolution divergente ou au moins le maintien de la divergence entre morphotypes, tout comme il pourrait faciliter la co-évolution des traits mâles et préférences femelles (même si une liaison physique des gènes impliqués n'est pas indispensable pour qu'une telle co-évolution ait lieu).

Il est impossible actuellement de pouvoir interpréter plus en détails ces résultats qui nécessiteront des analyses supplémentaires, et notamment de type recherche de QTL. Cependant nos résultats semblent en partie similaires à ceux trouvés notamment par Mifsud (2011), *i.e.* présence de nombreux réarrangements chromosomiques interspécifiques, qui, dans le cas de la paire d'espèces *J. ischiosetosa* / *J. albifrons*, semblaient associés à une diminution de la fertilité hybride (Mifsud, 2011).

4.3 Perspectives

L'augmentation du nombre de loci entre les cartes de liaison de référence et totale a été une bonne stratégie puisque le nombre de groupes de liaison a pu ainsi être redéfini, mais l'utilisation des deux autres familles sera nécessaire pour créer des cartes consensus basées sur un plus grand nombre de juvéniles afin d'affiner les recherches qui ont été menées dans ce chapitre.

Nous avons néanmoins déjà montré que le chromosome présentant la plus forte différenciation est le chromosome sexuel. Plusieurs interrogations et perspectives s'ouvrent à nous. Ainsi nous pouvons nous demander l'impact de cette différenciation sur la divergence des espèces, et s'il y a une introgression différentielle des allèles liés au Z versus liés au W en zone de contact. La différenciation a été mesurée uniquement sur des mâles (F_{ST} calculé sur les mâles prélevés en populations naturelles et utilisés tout au long de cette thèse), mais des femelles ont également été prélevées mais non analysées au cours de cette thèse (voir chapitre 2) et il serait donc intéressant de regarder si la différenciation sur le chromosome W, est différente de celle sur le chromosome Z (Dunn & Bennett, 1967). Enfin, les gènes codant les caractères sexuels secondaires mâles et le choix des femelles sont-ils placés sur les chromosomes sexuels (*e.g.*, Flycatchers, Sæther et al., 2007), confirmant ainsi leur importance dans la divergence initiale?

Il est parfois difficile de distinguer les loci outliers liés à l'habitat de ceux qui sont impliqués dans des barrières endogènes telles que des incompatibilités post-zygotiques qui peuvent être couplées au même endroit dans le scan génomique (Bierne et al., 2011). Dans cette étude, la difficulté aurait été de ne se baser que sur le scan génomique de *J. albifrons*, prélevé en Bretagne sous les pierres sachant que les *J. prae-hirsuta* se situent dans le fucus, et qu'une partie des loci sous sélection peuvent être liés à l'habitat. Cependant, nous avons également pu créer un scan génomique basé sur une carte de liaison *J. prae-hirsuta* provenant de Normandie, *i.e.* individus localisés également sous les pierres. Il n'y a donc plus de différences d'habitat entre les deux espèces et les loci sous sélection fortement différenciés entre les espèces en

Normandie peuvent être, entre autres, signe d'incompatibilités endogènes. On peut donc supposer que certains loci sous sélection situés sur les GL fortement différenciés en Bretagne et pas en Normandie pourrait être liés à de l'adaptation locale. Afin de rechercher les zones soumises à l'adaptation locale, il faudrait faire les mêmes analyses avec des individus des Iles Scilly qui partagent le même habitat mais montrent une forte différenciation interspécifique (chapitre 2), et faire une recherche de QTL sur l'ensemble de ces scans génomiques.

Une recherche de QTL sur les traits sexuels secondaires (*e.g.*, Veltsos et al., 2015), *i.e.* soies sur les périopodes des mâles, pourrait également être rapidement réalisée puisque le phénotype de 135 mâles issus de backcross a été relevé (comptage du nombre de soies au microscope de chacun des sept périopodes du côté droit de l'animal), et que ces mâles ont tous été séquencés par la même méthode de RAD décrite tout au long de cette thèse, y compris de ce chapitre. Au vue du maintien des morphes parentaux en Normandie on pourrait s'attendre à ce que de tels caractères soient situés sur des îlots génomiques mais l'étude de Ruegg et al. (2014) montre que les loci associés au comportement migratoire des grives *Catharus ustulatus* sont situés en dehors des îlots génomiques. Une autre hypothèse serait que ces caractères soient situés sur le chromosome sexuel (Iyengar et al., 2002). Il est également possible que les traits sexuels soient polygéniques avec un grand nombre de gènes impliqués et qu'aucun lien évident n'existe entre leur localisation et l'architecture génomique décrite dans ce chapitre.

Une perspective supplémentaire serait l'étude de l'action des éléments transposables dans les réarrangements génomiques, tels que des insertions, des suppressions, des inversions et par conséquent leur action sur la spéciation en général (Rieseberg, 2001, Feder et al., 2014). Enfin, il semble important dans les prochaines études d'analyser les distorsions de ségrégations au sein des individus backcross obtenus lors de nos croisements afin de cartographier d'éventuelles incompatibilités génétiques (*e.g.*, Phadnis & Orr, 2009).

CONCLUSION & PERSPECTIVES

Tout au long de cette thèse, je me suis intéressée à comprendre l'influence combinée de barrières à la reproduction, de l'architecture génomique et de l'histoire évolutive sur la force de l'isolement reproductif entre deux espèces. Pour cela, j'ai utilisé comme modèles d'étude *Jaera albifrons* et *J. praeirsuta*, deux des cinq espèces d'isopodes marins appartenant au complexe *Jaera albifrons*, et j'ai i) découvert une variation de la différenciation interspécifique au sein de différentes zones de sympatrie (chapitres 1 et 2), ii) quantifié l'isolement reproductif dans deux de ces zones (chapitre 3) et iii) recherché si la bactérie *Wolbachia* pouvait avoir une implication sur cet isolement reproductif (chapitre 4). J'ai également iv) cherché à comprendre l'histoire démographique de la divergence entre *J. albifrons* et *J. praeirsuta* (chapitre 5) et v) commencé à décrire l'architecture génomique de l'isolement reproductif entre ces deux espèces (chapitre 6). La plupart de ces analyses ont été basées sur la comparaison de deux régions, l'une montrant un fort isolement reproductif entre ces espèces (Bretagne) et l'autre montrant une hybridation introgressive (Normandie), et ont été réalisées en grande majorité à l'aide de nombreux marqueurs nucléaires (SNPs). L'imbrication de ces différents résultats et les perspectives de recherche qu'ils ouvrent ont été réunis dans les paragraphes ci-dessous, sous forme de trois thèmes de recherche, suivis par une rapide conclusion générale.

Thème 1 : Evènements passés et flux de gènes actuels

L'étude empirique de la différenciation interspécifique et des barrières impliquées dans l'isolement reproductif nécessite une bonne connaissance générale des modèles biologiques choisis, de leur distribution géographique, et de leur histoire démographique (Ravinet et al., 2017, Tine et al., 2014).

Au cours de cette thèse, nous avons montré que *Jaera albifrons* et *Jaera praeirsuta* sont trouvés sur de multiples zones en sympatrie le long de la côte, en partageant ou non un même habitat (les deux espèces sous les pierres : zones de « syntopie » en Normandie, ou une espèce sous les pierres, et l'autre sur des algues à proximité immédiate des pierres : zones de « micro-allopatrie » en Bretagne). Cette configuration entre dans le cadre d'un sujet d'actualité dans l'étude de la spéciation : la possibilité de reconnaître des cas de spéciations parallèles (*i.e.*, évolution indépendante répétée du même mécanisme d'isolement reproductif, Schluter & Nagel, 1995), *e.g.*, *Littorina saxatilis* (Butlin et al., 2014) ou encore les cichlidés du genre *Pundamilia* (Meier et al., 2017), d'une spéciation unique, et si on est face à une spéciation avec flux de gènes comme chez les papillons du genre *Heliconius* (Martin et al., 2013), ou à un contact secondaire suivant une spéciation allopatrique, *e.g.* anchois *Engraulis encrasicolus* (Le

Moan et al., 2016). Ici, nous avons détecté qu'il s'agit d'un unique évènement de divergence qui a eu lieu entre les espèces *J. albifrons* et *J. praeheirsuta*, suivi de contacts secondaires ayant probablement eu lieu simultanément (voir chapitre 5) mais dont les conséquences s'avèrent être différentes. Ainsi nous avons montré que dans notre zone d'étude Bretonne il y a une forte différenciation génétique entre les deux espèces. Dans cette zone le flux de gènes interspécifique est faible et hétérogène le long du génome, ce qui est du à l'impact de processus sélectifs entraînant une limitation de l'introgession sur plus de la moitié du génome. Dans notre zone d'étude Normande en revanche, on détecte une faible différenciation interspécifique, beaucoup de flux de gènes entre espèces avec une migration hétérogène le long du génome, témoignant du processus d'hybridation introgressive. Ce flux de gènes est asymétrique et plus fort de *J. albifrons* vers *J. praeheirsuta* en partie en raison du taux de migration aux locus soumis à des effets sélectifs directs ou indirects qui est beaucoup plus fort de *J. albifrons* vers *J. praeheirsuta* qu'inversement. Entre ces deux régions, un seul autre site présentant une situation de syntopie a été trouvé. Dans ce site (Goury, site 28 sur la Figure 2.1 du chapitre 2) la différenciation interspécifique est intermédiaire à ce que nous avons observé en Bretagne et en Normandie, mais aucun morphotype intermédiaire n'a été détecté (mais sur seulement 8 individus mâles observés, ce qui est très insuffisant étant donné la fréquence des morphotypes intermédiaires, même dans les populations fortement introgressées).

Plusieurs hypothèses peuvent être imaginées pour tenter d'interpréter la variation du niveau d'introgession dans les différents sites étudiés. Une première hypothèse serait que la force de l'introgession suive un gradient géographique, d'inexistant à fort, d'Ouest en Est (Figure D.1), hypothèse soit i) basée sur la connectivité des sites et une éventuelle émigration d'individus introgressés provenant de Normandie vers l'Ouest avec homogénéisation du génome par échanges génétiques intraspécifiques levant ainsi les barrières à la reproduction actuellement en place (flux de gènes actuels), soit ii) liée à l'histoire de la colonisation et de la mise en place progressive des contacts secondaires le long des côtes (moins probable d'après les résultats du chapitre 5 et les datations qui semblent similaires). Une seconde hypothèse serait que les populations mixtes étudiées soient isolées les unes des autres et représentent donc autant de répliques de contact entre les deux espèces qui se traduisent par un niveau d'introgession très variable, de manière indépendante dans chaque site.

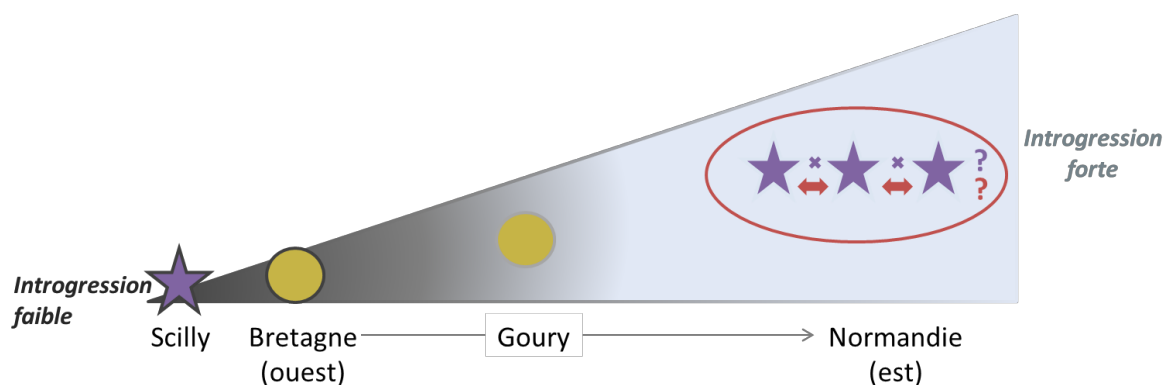


Figure D.1 : Schéma de l'introgession (absence en gris foncé, présence en gris clair) le long d'un gradient géographique d'ouest en est le long de la côte de la Manche. Les sites en sympatrie ne présentant pas d'hybridation sont représentés sous forme de ronds jaunes, ceux présentant de l'hybridation sont représentés sous forme d'étoiles violettes. En rouge est représentée une unique zone hybride avec échanges (flèches rouges) entre les différents sites introgressés, alors que les croix violettes représentent un isolement total entre les différents sites introgressés dans cette zone.

Ces suppositions amènent donc la question de la nature de la zone hybride en Normandie. S'agit-il d'une unique zone hybride où les populations sont connectées les unes aux autres malgré la distance (schématisée par une ellipse rouge, Figure D.1), ou de plusieurs zones hybrides indépendantes n'échangeant pas de migrants (représentées en violet, Figure D.1)? On peut même supposer un isolement par la distance tout le long de la côte, qui rendrait possible un pattern d'introgession suivant un gradient géographique. Cependant le long de cette zone peu de *J. prae-hirsuta* ont été détectés, et seuls des sites abritant uniquement *J. albifrons* sont trouvés plus ou moins régulièrement le long de la côte. Bien que n'excluant pas l'hypothèse que de telles populations uniquement composées de *J. prae-hirsuta* existent et n'aient pas été échantillonnées (e.g. zones à *Fucus vesiculosus* ou *Fucus serratus* ne présentant pas d'habitat de type galets à proximité immédiate), l'analyse d'isolement par la distance n'est pas envisageable actuellement pour cette espèce. Quoi qu'il en soit, le site d'hybridation éloigné (aux Iles Scilly), permet en revanche de supposer une apparition de l'hybridation à plusieurs reprises sur des sites non connectés (l'archipel des Scilly est situé à plus de 40Km de la côte la plus proche). Cette absence de *J. prae-hirsuta* le long de la côte en Normandie pose question également sur des conditions du maintien des deux morphotypes parentaux dans les populations à forte hybridation introgressive. Trouver en Normandie une population uniquement composée de *J. prae-hirsuta* pour savoir si il y a des échanges de migrants entre populations jouant un rôle dans

le maintien à long terme de cette espèce dans les populations fortement introgressées serait utile pour de futures études. Si les populations ne sont pas connectées, on suppose alors un niveau de différenciation interspécifique d'intensité similaire à la Bretagne. Alternativement, il est possible que dans certaines régions de telles populations pures de *Jaera praehirsuta* n'existent pas. Cela rejoint l'idée d'un gradient géographique d'introgression (Figure D.1) et souligne également la possibilité que les deux espèces *J. albifrons* et *J. praehirsuta* coexistent sur le long terme en absence d'immigration issue de populations parentales pures. Les conditions d'un tel maintien, discutées dès le chapitre 1, posent question. L'analyse de croisements expérimentaux (chapitre 3) apporte des éléments de réponse par le maintien d'un isolement sexuel fort bien qu'asymétrique, et le rôle limité d'une barrière de type post-zygotique. La nature et le rôle des barrières reproductives font l'objet de la section suivante.

Thème 2 : Barrières à la reproduction

Connaitre la nature et la force des barrières impliquées dans l'isolement reproductif en Bretagne et Normandie et détecter quelles sont les différences permettant aujourd'hui une hybridation introgressive dans une région et pas dans l'autre, alors qu'une même histoire (contact secondaire) a eu lieu dans les deux régions, est d'un réel intérêt dans l'étude de la spéciation. La difficulté réside en ce que les barrières impliquées dans l'isolement reproductif peuvent trouver leur origine lors d'une phase d'allopatricité et/ou avoir évolué après le contact secondaire et peuvent être différentes selon les histoires des sites étudiés.

L'absence d'isolement temporel et d'isolement mécanique avait déjà été démontrée par le passé entre les espèces du complexe *Jaera albifrons* (Jones & Naylor, 1971b, Bocquet, 1953). Dans cette étude, nous avons confirmé que la barrière comportementale semble être la principale barrière à la reproduction, soutenue par les barrières écologiques (en Bretagne) et post-zygotiques (dans les deux régions). En effet, une forte barrière comportementale a lieu lors de croisements interspécifiques (dans les deux régions) et, bien qu'elle reste forte en Normandie, elle y est incomplète, asymétrique et peut-être un peu affaiblie.

Les croisements hétérospécifiques ou de type backcross produisent globalement le même nombre d'individus à maturité que les croisements intraspécifiques (chapitre 3), signe d'une barrière post-zygotique inexistante à ces stades. Une diminution de la fitness des individus issus de backcross liée à leur capacité à trouver un partenaire pour la reproduction ne peut cependant pas être exclue.

Toutes les populations en sympatrie montrent une différenciation interspécifique variable selon les sites avec des isolements plus ou moins forts.

En Bretagne, le fort isolement comportemental est couplé avec un isolement écologique empêchant l'introgression. L'isolement écologique n'est pas total, ce qui se visualise sur le terrain par le chevauchement de niches entre les deux espèces, *i.e.* *J. praehirsuta* est trouvé sous les pierres et inversement (respectivement 0.04 % et 1.4% sur 6 sites bretons). La même souche de *Wolbachia* qu'en Normandie a été détectée, mais uniquement chez une femelle supposée *J. praehirsuta* car trouvée sur du fucus. Une incompatibilité cytoplasmique n'est donc pas strictement exclue puisque aucun hôte *J. albifrons* n'a été détecté en Bretagne. On peut donc supposer en milieu naturel des mâles *J. praehirsuta* infectés, et des femelles *J. albifrons* non infectées. Or, certaines souches de *Wolbachia* rendent les mâles infectés partiellement ou totalement incompatibles avec les femelles non infectées, *i.e.* aucun juvénile ne peut être issu d'un tel type de croisement (Merçot & Charlat, 2004), résultat que l'on obtient lors de croisements expérimentaux, puisqu'aucun croisement dans le sens femelle *J. albifrons*- mâle *J. praehirsuta* n'a donné de juvéniles (voir chapitre 3, Figure 3.2), et qui peut être schématisé tel que montré dans la Figure D.2.

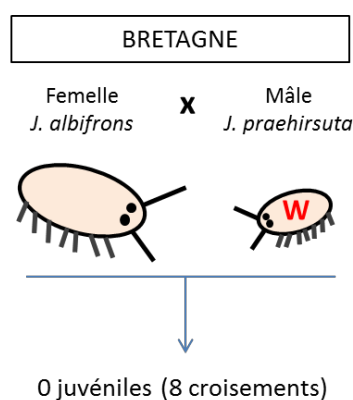


Figure D.2 : Croisement interspécifique en Bretagne dans l'hypothèse où un mâle *J. praehirsuta* serait infecté par *Wolbachia*, et la femelle *J. albifrons* non infectée. Sur 8 croisements effectués (chapitre 3), aucun n'a donné de juvénile ce qui est attendu sous cette hypothèse.

En Normandie, il existe également un isolement comportemental, semblant un acteur majeur du maintien des morphes parentaux puisqu'il n'y a pas d'isolement écologique visible et un très faible isolement post-zygotique. De plus, il ne semble pas y avoir d'incompatibilités cytoplasmiques dues à *Wolbachia* car la même souche a été trouvée chez les deux espèces.

Une hypothèse pourrait être que l'influence de la dynamique démographique liée à la variabilité des habitats impacte la possibilité de choix de partenaire et donc l'isolement comportemental.

On sait que l'habitat des sites Normands que nous avons étudiés sont beaucoup plus facilement remaniés par les tempêtes (*e.g.* variations saisonnières du niveau de recouvrement des habitats rocheux par le sable) que la plupart des sites Bretons auxquels nous nous sommes intéressés. Ainsi Solignac, seulement 2 ans après son premier passage à Luc-sur-mer, n'avait pas retrouvé les habitats propices à la présence de *Jaera sp.*, ce qui s'était caractérisé par une très forte réduction d'abondance à ce site (1 998 à 37 individus échantillonnés, Solignac, 1969b). On peut donc supposer qu'une telle variabilité dans le maintien de l'habitat ait pu diminuer la barrière comportementale dans la mesure où le choix de partenaire est fortement impacté par la possibilité qu'un tel choix ait lieu. En effet, il est possible que l'isolement fortuit d'individus d'une espèce au sein d'un groupe plus dense d'individus de l'autre espèce accepte des partenaires hétérospécifiques (testé en laboratoire, en fréquence rare, en situation de non-choix, chapitre 3). Lors du contact secondaire, il est possible que les individus *J. prae-hirsuta* se soient retrouvés dans des environnements dans lesquels ils sont mal adaptés, avec peu d'endroits abrités le long des côtes normandes, et aient tiré avantage de l'hybridation avec l'espèce *J. albifrons* sous la forme d'une adaptation à un nouvel habitat, *i.e.* sous les pierres (Hedrick, 2013). Cette introgression adaptative aurait conduit à l'homogénéisation du génome, tout en gardant un certain isolement reproductif (Clarkson et al., 2014) permettant ainsi le maintien des morphes parentaux, et est compatible avec l'introgression asymétrique détectée. Cette hypothèse repose cependant sur le fait que les espèces soient assez fortement inféodées à un type d'habitat donné, ce qui est loin d'être clairement établi. Si l'adaptation aux algues chez *J. prae-hirsuta* était ancestrale, alors elle aurait été maintenue dans la majorité des sites en sympatrie malgré le contact secondaire, excepté en Normandie, zone de forte dynamique des plages. En revanche, il est également possible que les espèces ne soient pas inféodées à un habitat, mais qu'elles profitent de la présence de plusieurs niches potentielles (Bretagne) ou alors vivent en syntopie lorsqu'une seule niche est exploitable (Normandie). Dans ce cas, des sites avec présence de *J. albifrons* et/ou *J. prae-hirsuta* indépendamment sous pierres et/ou sur algues devraient être détectés ce qui n'a pas encore été le cas.

Enfin, sur le site des Iles Scilly où un hybride a été détecté, il semble il y avoir d'autres barrières à la reproduction que l'isolement comportemental. En effet, il y a une absence d'isolement écologique (tout au moins les deux espèces sont trouvées dans le même macro-habitat, comme dans nos populations Normandes), et pourtant une forte différenciation génétique interspécifique est visible (comme dans nos populations Bretonnes). Plusieurs hypothèses peuvent être faites. 1 - La présence d'une barrière post-zygotique très forte, et/ou un isolement comportemental fort, voire un renforcement de l'isolement comportemental, absent au sein

des deux autres sites. 2 - Les deux espèces sont séparées d'une manière ou d'une autre le long de l'estran et utilisent des micro-niches différentes que nous n'avons pas identifiées. En effet, à la différence de la Normandie, la distribution individuelle précise des individus le long de l'estran n'a pas été analysée. Cependant, rien dans l'échantillonnage ne laisse supposer qu'une telle ségrégation existe. 3 - La densité est suffisamment forte et l'habitat moins variable que supposé en Normandie, ce qui permet le maintien d'un isolement sexuel plus fort. Des études supplémentaires seront nécessaires pour analyser la possibilité que de tels scénarios aient pu avoir lieu, et une perspective intéressante serait de réaliser un scan génomique à partir de la différenciation interspécifique trouvée à ce site et de comparer ce scan à ceux de Bretagne et de Normandie, ainsi que de quantifier l'isolement reproductif à ce site. Enfin, à Goury, aucune analyse n'a encore été effectuée, mais il semble y avoir des niveaux de flux de gènes intermédiaires à ceux observés en Bretagne et en Normandie, ce qui pourrait être, selon plusieurs études, optimal pour une divergence adaptative (Garant et al. 2007).

Thème 3 : Architecture génomique et forces évolutives en action

Le paysage génomique peut être façonné par des processus tels qu'une sélection divergente ou de la sélection d'arrière-plan sur des régions de polymorphisme ancestral partagé ou à faible recombinaison, visualisés sous forme de pics de différenciation qui vont tendre à être homogénéisés par le flux de gènes (Figure D.3).

Chez les gobe-mouches (espèces du genre *Ficedula*) une forte hétérogénéité de différenciation le long du génome est présente, mais les îlots de différenciation seraient liés davantage à la sélection d'arrière-fond et au balayage sélectif dans des zones à faible recombinaison qu'à la variation de flux de gènes interspécifiques (Burri et al., 2015).

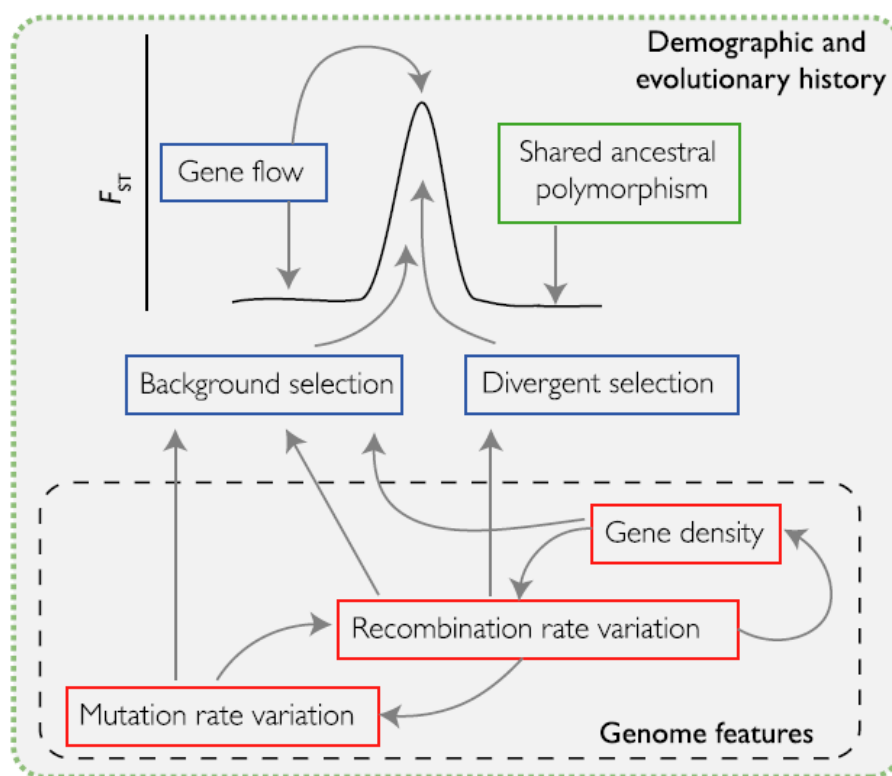


Figure D.3 : Processus pouvant façonner le paysage génomique. D’après Ravinet et al. (2017).

En ce qui concerne *Jaera albifrons* et *J. praeirsuta* en Bretagne, une semi-perméabilité du génome due à une sélection indirecte (effets Hill-Robertson au sens large) a été détectée. En effet, la moitié du génome a une taille efficace réduite de 13% ce qui augmente l’effet de la dérive avec la fixation d’allèles différenciés et/ou délétères entre populations, et par conséquent la mise en place de différences génétiques pouvant entraîner différentes formes d’isolement reproductif (Excoffier & Ray, 2008).

Nous avons montré au cours du chapitre 5, que le modèle démographique le plus probable pour les populations Bretonnes que nous avons étudiées est un contact secondaire, avec de faibles flux de gènes contemporains. En revanche, aucun flux de gènes n’a été observé au cours des chapitres 1 et 2. De fortes barrières se maintiennent donc en parallèle d’un flux de gènes résiduel actuel ou récemment stoppé. Ainsi, l’ensemble du génome montre une très forte différenciation.

En Normandie, le faible isolement interspécifique ne permet pas de détecter une hétérogénéité de la taille efficace le long du génome. Cependant, des régions génomiques dont la forte différenciation interspécifique n’est pas compatible avec des flux de gènes libres dans un modèle neutre ont été détectées, signe d’un l’isolement reproductif (Navarro & Barton, 2003,

Feder et al., 2012). Ces régions du génome fortement différenciées ont notamment été repérées sur les chromosomes sexuels ainsi que sur les chromosomes ayant subi des réarrangements (Faria & Navarro, 2010, Noor & Bennett, 2009, Rieseberg, 2001). Il est dit dans la littérature que les inversions sont nécessaires pour la persistance des espèces face au flux de gènes chez certaines drosophiles (Kulathinal et al., 2009); et que les incompatibilités chromosomiques seraient à l'origine par exemple de la divergence des équidés, bien que la forte variation du nombre de chromosomes n'empêche pas le flux de gènes (Jónsson et al., 2014). Chez les *Jaera*, des réarrangements chromosomiques ont eu lieu chez une ou les deux espèces, faisant varier le nombre de chromosomes au sein et entre espèces, ce qui ne limite pour autant pas l'hybridation, mais semble empêcher l'introgession à ces chromosomes réarrangés en Normandie.

Les loci à forte différenciation interspécifique n'étant pas tous impliqués dans les barrières à la reproduction, il est maintenant nécessaire d'aller plus loin, par le biais de recherche de loci liés à des traits directement impliqués dans l'isolement (Ravinet et al., 2017) afin de mener à une compréhension plus globale du paysage de divergence génomique. Plusieurs traits peuvent contribuer à l'isolement (Coyne & Orr, 2004) et il y peut y avoir un effet de pléiotropie, *i.e.* implication d'un allèle dans plusieurs traits différents, avec une chaîne d'association entre plusieurs traits participant à l'isolement (Figure D.4). Une longue chaîne aura plus de chance d'être rompue par le flux de gènes que des chaînes courtes (Smadja & Butlin, 2011). La recherche de telles chaînes serait une perspective intéressante, après identification des loci liés à aux barrières, dans le cas où cela pourrait éventuellement expliquer le changement d'adaptation à l'habitat en Normandie.

La difficulté est de différencier les loci impliqués dans l'isolement reproductif et partagés entre répliques qui dateraient donc de la divergence primaire de ceux qui ont pu s'installer depuis le contact secondaire (Bierne et al., 2013, Welch & Jiggins, 2014). De plus, les loci impliqués dans l'isolement reproductif peuvent être variables entre espèces et entre sites, en fonction des zones, du choix de partenaires, de l'adaptation à l'habitat, et des forces évolutives sous-jacentes (Nolte et al., 2009, Teeter et al., 2010).

L'isolement comportemental semblant être la barrière principale, une recherche de QTL sur les traits sexuels secondaires (*e.g.*, Veltsos et al., 2015), *i.e.* soies sur les périopodes des mâles, devrait être réalisée afin de détecter leur emplacement le long du génome. Une étude de Mifsud (2011) a suggéré que des QTL seraient préférentiellement localisés sur des chromosomes réarrangés ou liés au sexe. Au vue du maintien des morphes parentaux en Normandie on pourrait s'attendre à ce que de tels caractères soient situés sur des îlots

génomiques mais l'étude de Ruegg et al. (2014) montre que les loci associés au comportement migratoire des grives *Catharus ustulatus* sont situés en dehors des îlots génomiques. Une autre hypothèse serait que ces caractères soient situés sur le chromosome sexuel (Iyengar et al., 2002, Mank et al., 2007).

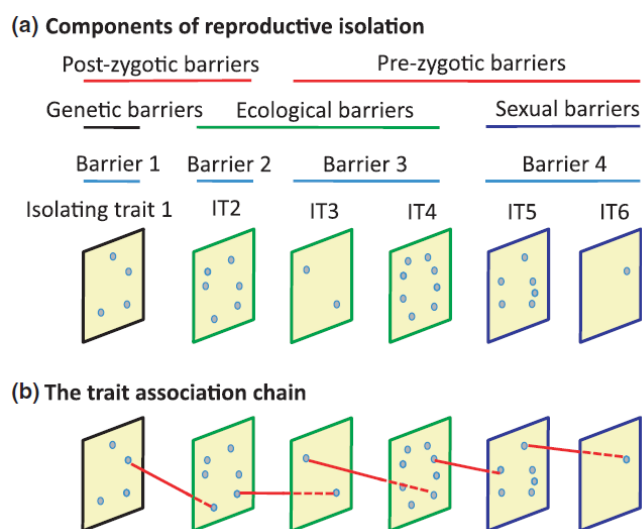


Figure D.4 : Chaines d'association de traits impliqués dans des barrières à la reproduction.
D'après Smadja & Butlin (2011).

Conclusion générale

Les études de la spéciation entre paires de populations divergentes répliquées occupant différents habitats et dont les schémas de différenciation génomique sont hétérogènes ont depuis peu fait leur apparition dans la littérature (Gagnaire et al., 2013, Rougemont et al., 2016a, Jones et al., 2012, Deagle et al., 2012, Westram et al., 2014). Ces études comparatives permettent de pouvoir détecter si les bases génétiques de la divergence sont les mêmes, comprendre les caractéristiques démographiques et comparer les histoires à différents sites ; de pouvoir identifier les régions génomiques résistantes à l'introgression, parfois, dues à des réarrangements chromosomiques ; ou encore de détecter l'influence de la sélection d'arrière-plan ou la force du flux de gènes.

Notre travail sur la paire d'espèces *Jaera albifrons* / *J. praehirsuta* pointe vers un événement de spéciation allopatrique unique suivi de contacts secondaires qui ont entraîné des histoires évolutives contrastées entre zones géographiques. Ces histoires évolutives différentes semblent impliquer essentiellement des variations d'isolement sexuel (très probablement la composante comportementale de cette barrière), car fort et asymétrique dans tous les sites étudiés, mais parfois incomplet (e.g. Normandie), et des variations de disponibilité locale d'habitat et d'adaptation à ces habitats. Le rôle d'une barrière de type post-zygotique n'est pas exclu en Bretagne et Normandie mais semble limité, et son impact reste à mesurer au site des Scilly. *Wolbachia* ne semble pas avoir de rôle particulier en Normandie mais pourrait être source d'incompatibilités en Bretagne (une hypothèse que nous allons bientôt pouvoir confirmer ou réfuter). Sa présence serait également à rechercher aux Iles Scilly. Enfin, la dynamique des plages pourrait avoir un rôle important, affectant la force de la barrière comportementale et permettant ainsi une introgression en Normandie.

Une balance entre isolement sexuel et isolement écologique semble donc être au cœur de l'évolution et de la coexistence des deux espèces, et la force de la barrière comportementale par rapport aux autres barrières pourrait témoigner d'une spéciation conduite par une sélection sexuelle divergente (Seehausen et al., 2014), avec plusieurs populations se trouvant à différents niveaux du continuum de spéciation. Bien qu'il laisse de nombreuses questions ouvertes, ce travail a permis de confirmer l'intérêt de ce complexe d'espèces pour l'étude plus globale de la spéciation, notamment par l'analyse comparative de zones sympatriques, montrant un continuum de spéciation (hybridation introgressive à isolement reproductif total).

Bien connaître son modèle requiert beaucoup d'analyses. Ainsi en 2015, une revue basée sur le modèle des littorines reprenait 269 publications écrites en 18 ans dans le domaine de l'écologie évolutive (Rolán-Alvarez et al., 2015). Durant cette thèse nous nous sommes concentrés à

répondre à quelques grandes questions, qui viennent ici apporter des perspectives intéressantes.

Notre étude s'apparente au modèle des lamproies *Lampetra fluviatilis* et *L. planeri* décrit par Rougemont et al. (2015). En effet l'étude de plusieurs populations en sympatrie montre des niveaux de différenciation variables, de la panmixie à une différenciation légère, avec maintien des morphes parentaux ; et des populations en parapatrie montrent un isolement très fort, notamment dû aux barrières géographiques. Cependant, dans nos populations de *Jaera* bretonnes, au contraire, un léger chevauchement de niche a été montré (Ribardière et al., 2017) signifiant que nos espèces peuvent être trouvées dans des habitats propres à la seconde espèce et qu'une forte différenciation interspécifique et un fort isolement ont donc lieu au-delà de l'isolement écologique et géographique.

C'est l'un des premiers points importants que j'ai soulevé durant ma thèse : *Jaera albifrons* et *Jaera praehirsuta* sont retrouvés en milieu naturel en sympatrie dans plusieurs sites probablement indépendants et montrent selon les sites une différenciation génétique interspécifique différente impliquant des processus divers. De plus, nous avons montré que ces espèces sont issues d'une divergence allopatrique probablement due à la sélection sexuelle que le contact secondaire n'a pas beaucoup érodée même en présence d'hybridation introgressive. Les barrières à la reproduction bien que fortes ne sont pas strictes et le chromosome sexuel ainsi que des régions génomiques ayant subi des réarrangements semblent être les principales limites à l'introgression.

ANNEXES

ANNEXE 1 : Construction d'une banque ddRAD

Pour espérer obtenir de bons marqueurs génétiques la première étape, celle en laboratoire, est primordiale. Choisir le protocole adapté au modèle d'étude, puis réaliser les manipulations avec minutie sont les clés de la première étape vers la réussite. Ces étapes sont décrites dans cette annexe et ont été supervisées par Claire Daguin, ingénieure de recherche.

Le séquençage de marqueurs ADN associés à des sites de restrictions (RAD-seq) par double digestion (ddRAD) est une méthode faisant intervenir 2 enzymes de restriction avec sélection de taille précise permettant d'obtenir des fragments d'ADN (RAD-tags) répartis tout au long du génome, en excluant les sites de digestion trop proches ou trop éloignés (Peterson, 2012). La méthode que nous avons choisi d'utiliser a été adaptée du protocole de Brelsford (2016), lui-même dérivé du protocole de Peterson (2012) et de Parchman (2012). L'avantage majeur de ce protocole est que les différents échantillons y sont traités séparément jusqu'à ce qu'ils soient amplifiés par PCR. Cela permet donc de pouvoir visualiser le produit de PCR de chacun des échantillons avant le pool et de garantir ainsi une représentativité optimale de chacun dans la banque ainsi obtenue. Les autres caractéristiques importantes de ce protocole sont :

- le multiplexage d'un maximum de 288 individus par banque, par la combinaison de 24 barcodes (retrouvés en début de read) et de 12 indexes illumina lus lors du séquençage (voir Figure 1).
- la purification par billes magnétiques des produits de ligation, de façon à éliminer efficacement les fragments très petits et les adaptateurs en excès.
- la sélection de taille à l'aide du pippin prep (Sagescience), qui permet de ne garder dans la banque qu'une certaine gamme de taille de fragments, et ceci de façon précise et répétable.

ddRAD

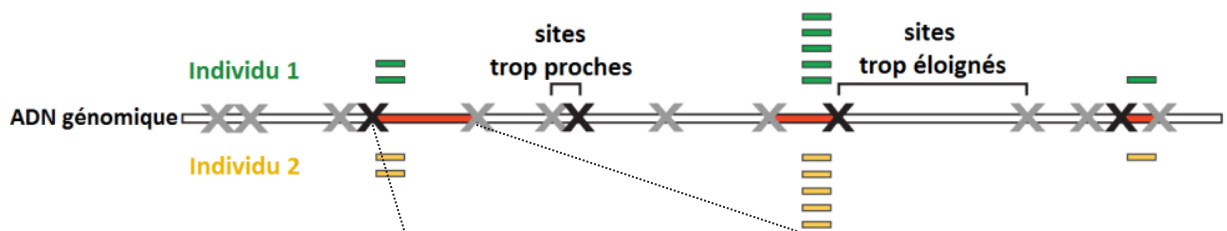
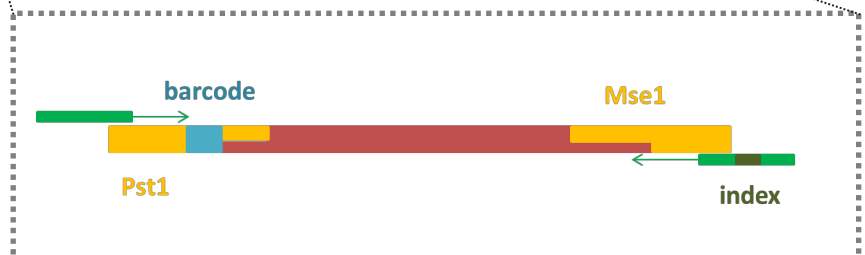


Figure 1 – Descriptif de la double digestion de l'ADN génomique, modifié d'après Peterson (2012), et schéma d'une combinaison index/barcode, les adaptateurs utilisés dans la ligation étant représentés en jaune, les amorces utilisées dans la PCR en vert, et l'ADN génomique de l'échantillon en rouge.



Mise au point du protocole : 1^{ère} expérience en séquençage MiSeq Illumina

Aucun protocole de RAD-seq n'ayant encore été testé sur des espèces du complexe *Jaera albifrons*, nous avons décidé de construire en premier lieu une banque contenant 16 individus, et séquencée sur le MiSeq (Illumina) de la plate-forme génomique de la SBR, de façon à tester et adapter le protocole de construction de banque ddRAD. En l'absence de données génomiques permettant d'estimer *in silico* le meilleur couple d'enzymes adapté à notre capacité de multiplexage et de séquençage, le même protocole que celui utilisé pour les cionos (autre espèces étudiée dans notre laboratoire) a été testé dans un premier temps (voir le diagramme en Figure 2).

Les 16 individus étaient tous des mâles, 6 *J. albifrons*, 7 *J. praehirsuta* et 3 morphotypes intermédiaires, provenant de Bretagne (Inizan, site 15 sur la Figure 1 du Chapitre 2) ou de Normandie (Grandcamp et Sainte-Honorine, sites 30 et 32 sur la Figure 1 du Chapitre 2). Une concentration d'ADN moyenne de 2.04 ng/ μ l [1.44-2.92] a été mesurée sur l'ensemble de ces individus. La banque a été construite à partir de 50 ng d'ADN, et les séquences ont été obtenues par double digestion à l'aide des enzymes Pst1 et Mse1. Huit barcodes ont été ajoutés lors de la ligation ainsi que deux index lors de la PCR normalisante (cf. diagramme Figure 2). Après une sélection de taille des fragments entre 270 et 600 pb, les échantillons ont été passés au bioanalyser révélant une sélection de taille plus large (210-600 pb). Finalement un séquençage Miseq a été réalisé à l'aide d'une cassette 150-V3, après dilution de la banque pour obtenir une concentration de 4 nM. Le Mix final était de 600 μ l dont 5% de PhiX (ADN phage), pour garantir la réussite du séquençage malgré la faible diversité des premières bases le long des fragments séquencés, selon les recommandations d'Illumina.

Ce séquençage a permis d'obtenir 25 millions de reads, dont 9 400 773 de reads après séquençage et 17 millions de reads une fois tronquée à 110 pb, tous de très bonne qualité. Le démultiplexage a permis de retrouver les 16 individus, en proportions équivalentes. Nous avons ainsi validé le protocole testé et utilisé ces premiers résultats pour estimer la composition des banques suivantes et l'effort de séquençage à réaliser.

Ce protocole a donc été choisi pour le séquençage de nos données en HiSeq Illumina, et est décrit plus en détail ci-après.

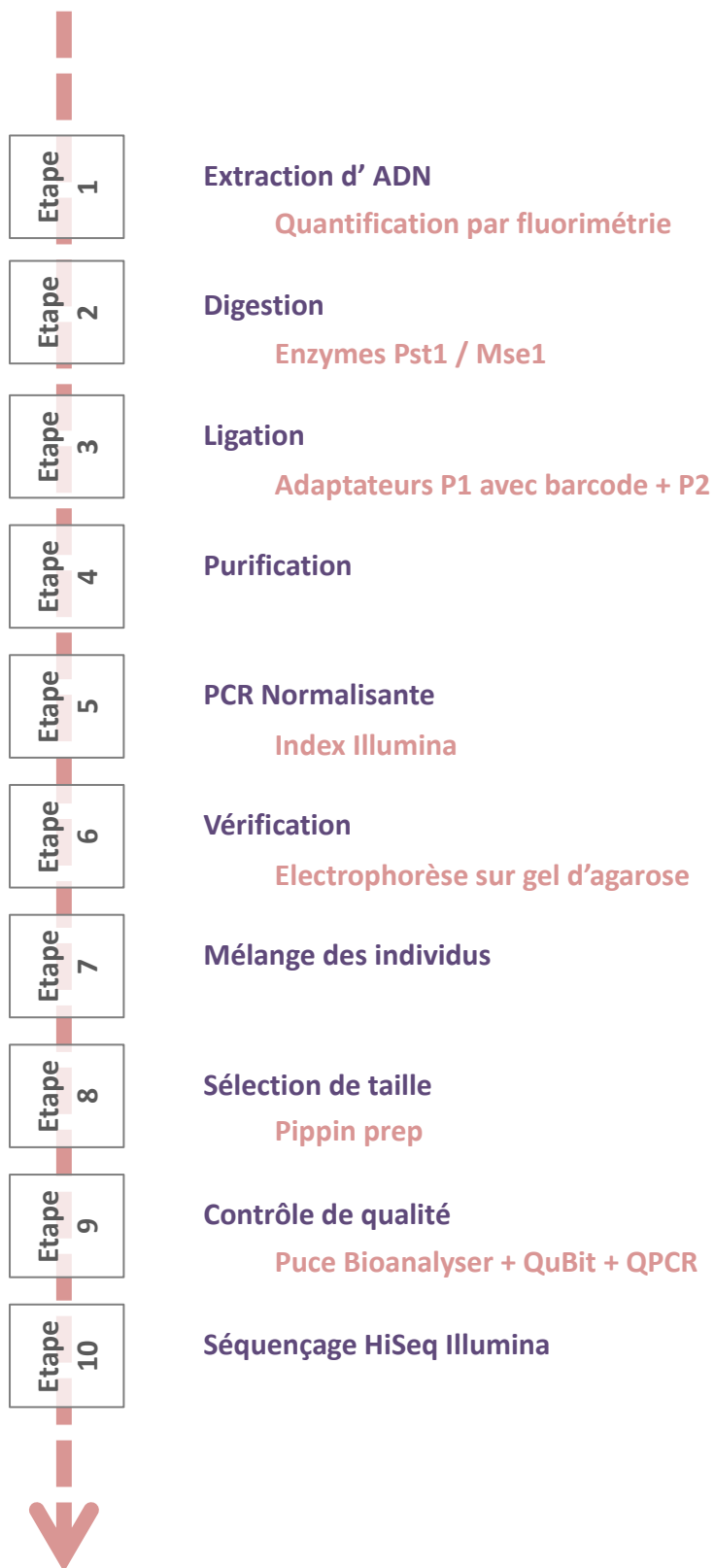


Figure 2. Diagramme du protocole ddRAD suivi

Etape 1.1 - Extraction d'ADN

Nous avons utilisé le kit NucleoSpin® Tissue (Macherey-Nagel) principalement en plaque 96 (protocole décrit ici) et parfois en colonnes individuelles.

- 1 → placer l'individu entier dans un tube de 2 mL
- 2 → ajouter 180 µL de tampon de lyse T1 et 25 µL de Proteinase K, puis vortexer et centrifuger 15 sec à 1500 g
- 3 → Incuber à 56°C pendant 3 heures sous agitation lente (ou avec agitation au vortex au bout de 1h30)
- 4 → Vortexer et centrifuger 15 sec à 1500 g, puis ajouter 200 µL de tampon BQ1 et 200 µL d'éthanol à 100% et enfin vortexer puis centrifuger 10 sec à 1500 g
- 5 → transférer les produits dans une plaque avec colonnes puis centrifuger à 6 000 g pendant 10 min
- 6 → ajouter 500 µL de buffer BW et centrifuger à 6 000 g pendant 2 min
- 7 → ajouter 700 µL de buffer B5 et centrifuger à nouveau à 6 000 g pendant 4 min
- 8 → placer dans un incubateur pendant 10 min à 70°C pour évaporer l'alcool restant
- 9 → éluer l'ADN extrait avec 50 µL de tampon d'élution BE (préchauffer à 70°C), puis incuber pendant 1 minute à température ambiante avant centrifugation pendant 2 minutes à 6 000 g. Cette étape est à répéter deux fois pour obtenir un volume d'élution de 100 ou 150µL.

Etape 1.2 - Quantification de l'ADN extrait par fluorimétrie

Le kit Quant-IT dsDNA picogreen (Life technologies) a été utilisé afin de mesurer la concentration d'ADN double brin après extraction, par technique de fluorescence. Les extraits d'ADN sont dilués au 1/50^{ème} avec du buffer TE1X. Une gamme étalon (de 0, 5, 10, 50, 100, 250, 500 à 1000 pg/µl) est préparée à partir d'un ADN de référence de concentration connue. En moyenne sur un échantillon de 420 individus mâles et femelles de l'espèce *J. albifrons* ou *J. praehirsuta* nous avons mesuré une concentration moyenne de 2.97 ng/µl par individu (min : 0.37 ng/µl, max : 25.56 ng/µl).

Cependant, ces mesures sont très variables entre individus, et notamment plus élevées chez les femelles qui sont beaucoup plus grosses que les mâles (2 à 3 mm de différence, en moyenne respectivement 4.49 ng/µl et 1.78 ng/µl).

Une seconde quantification par QuBit a été réalisée sur un échantillon de 50 individus mâles afin de confirmer les résultats obtenus par picogreen, et a permis de mesurer une concentration moyenne de 2.2 ng/µl par individu (min : 0.3 ng/µl, max : 5.8 ng/µl).

Les banques RADseq ont été construites en utilisant environ 50 ng d'ADN (basé sur une quantité de 2 ng/µl d'ADN par individu) au départ pour chacun des échantillons.

Etape 2 - Double digestion de l'ADN génomique

Deux enzymes de restriction sont utilisées dans ce protocole, qui reconnaissent et coupent des séquences de nucléotides caractéristiques (cf. Figure 3), *i.e.* appelés sites de restriction, présents tout au long du génome. L'une est PstI, et coupe rarement car reconnaît des séquences plus longues et donc plus rares, tandis que la seconde, MseI, coupe beaucoup plus fréquemment.

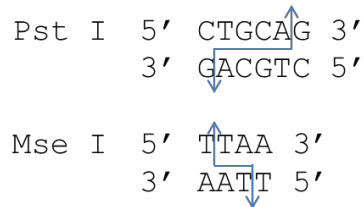


Figure 3. Séquence des sites de restriction ciblés par les enzymes utilisées pour ce protocole

Pour éviter les contaminations des banques par des produits de PCR, la digestion a été réalisée dans un thermocycler placé dans le laboratoire où les manipulations AVANT PCR ont lieu, dans un mélange réactionnel comme décrit dans le Tableau 1.

Tableau 1 – Mix de préparation pour la digestion

		concentrations stock	volumes à ajouter (µl)
Mix (5 µl)	Cutsmart buffer NEB	10X	3
	Enzyme 1 - Pst1	20u/µL	0.4
	Enzyme 2 - Mse1	10u/µl	0.2
	H2O		1.4
	ADN	2 ng/µl	25
TOTAL (mix + ADN)			30
37°C pendant 10h puis 20mn à 65°C puis 12°C			

Etape 3 - Ligation

Le multiplexage des individus dans une seule et même banque est possible par l'utilisation combinée de barcodes, insérés lors de la ligation, et d'index illumina, insérés lors de la PCR (voir étape 5). Lors de la ligation, des adaptateurs vont être placés à chaque extrémité des fragments préalablement découpés lors de la digestion. L'adaptateur P1, correspondant à PstI, possède un barcode qui est spécifique et permet de reconnaître les individus entre eux. 24 barcodes ont été utilisés (2-26, voir Tableaux 2 et 3). La ligation est réalisée par ajout du mélange de ligation (Tableau 4) dans la plaque où s'est déroulée la digestion, et effectuée dans le même thermocycler à 16°C pendant 6h.

Tableau 2 – Barcodes utilisés lors de la ligation

Barcode_ID	Barcode_séquence	Barcode_ID	Barcode_séquence	Barcode_ID	Barcode_séquence
2	CTGGTT	10	CATCAA	19	CAACCG
3	AAGATA	11	GCCTGG	22	CCAGCT
4	ACTTCC	13	TCGCAT	20	GTATGA
5	TTACGG	14	GGTAGA	21	TGGATT
6	AACGAA	15	GGAGCG	23	AACTCG
7	ATTCAT	16	TTGAAC	24	ACCAGA
8	CCGACC	17	GATTAC	25	TCAGGA
9	ATCGTC	18	CGAGGC	26	CGCTAC

Tableau 3 – Plan de plaques des numéros de barcodes utilisés

	1	2	3	4	5	6	7	8	9	10	11	12
A	25	9	17	25	9	17	25	9	17	25	9	17
B	2	10	18	2	10	18	2	10	18	2	10	18
C	3	11	19	3	11	19	3	11	19	3	11	19
D	4	26	22	4	26	22	4	26	22	4	26	22
E	5	13	20	5	13	20	5	13	20	5	13	20
F	6	14	21	6	14	21	6	14	21	6	14	21
G	7	15	23	7	15	23	7	15	23	7	15	23
H	8	16	24	8	16	24	8	16	24	8	16	24

Tableau 4 – Mix de préparation pour la ligation

		concentrations stock	volumes à ajouter (µl)
Mix (9.2 µl)	Adaptateur P2	10µM	2.2
	Cutsmart buffer	10X	1.2
	ATP	10mM	4.2
	ADN ligase T4	400u/µl	0.4
	H2O		1.2
	Adaptateur P1 (barcode)	1 µM	2.8
	ADN digéré		30
	TOTAL		42
16°C pendant 6h puis 12°C			

Etape 4 - Purification

Après la ligation, une étape de purification est réalisée à l'aide de billes magnétiques AMPure XP (Beckman), afin d'éliminer les fragments d'ADN trop petits, et les adaptateurs en excès non incorporés durant la ligation. Le ratio solution de billes / solution d'ADN utilisé est de 1.5X. Cette étape est réalisée en format 96 à l'aide d'une pipette 8 canaux, en procédant par lots de 3 à 4 colonnes maximum, à l'aide d'un aimant spécialisé pour plaque 96 (Figure 4).



Figure 4 – Aimant utilisé lors de la purification avec les billes AMPure XP.

- 1 → ajouter 60 μ l de billes stabilisées à température ambiante à 40 μ l de la ligation précédemment effectuée : mélanger une dizaine de fois par resuspension, puis laisser 5 mn à température ambiante.
- 2 → placer la plaque sur la plaque aimantée (Figure 4) et attendre 5 minutes à température ambiante. Pipeter le surnageant et le jeter. Laver avec 200 μ l d'EtOH 70 % fraîchement préparé. Attendre 30 secondes et retirer l'éthanol. Faire un second lavage de la même manière (200 μ l d'EtOH 70 %). Laisser sécher l'éthanol à l'air libre 5 mn environ.
- 3 → retirer la plaque de l'aimant après séchage et avant que les billes ne craquellent, puis ajouter immédiatement 40 μ l de Tris 10 mM, et mélanger une dizaine de fois par resuspension les billes, puis laisser 5 mn à température ambiante.
- 4 → remettre la plaque sur l'aimant et attendre 5 mn à température ambiante. Récupérer le surnageant dans une nouvelle plaque sans prélever de billes, puis conserver cette plaque au congélateur.

Etape 5 - PCR Normalisante

Afin d'augmenter notre capacité de multiplexage au sein d'une même banque, l'ajout d'un index (A, B, C, ..., L ; voir Tableau 5) lors de la PCR est réalisé, afin d'obtenir une combinaison index/barcode unique pour chaque individu composant la banque (Tableau 6). La PCR est dite normalisante car dans une première partie, les amorces y sont en quantité limitante. La quantité de fragments amplifiés produite est donc théoriquement proportionnelle à la quantité d'amorces et non pas à la quantité de matrice d'ADN (les fragments issus de la digestion-ligation) présente au départ. Elle est réalisée dans un mélange réactionnel (voir Tableau 7) pendant 15 cycles de façon à éviter d'introduire des erreurs de PCR suivi d'un cycle final où les amorces sont en très large excès (Tableau 8), de façon à parfaire la construction de tous les fragments.

L'enzyme utilisée pour cette PCR est une enzyme dite haute-fidélité, la Q5 Hot Start polymerase (New England Biolabs) qui par sa propriété « proof reading » minimise les erreurs lors de la PCR.

Tableau 5 – Index Truseq illumina IIIPCR2 pouvant être utilisés lors de la PCR

Index_IIIPCR2_ID	Index_séquence
A	ATCACG
B	CGATGT
C	TTAGGC
D	TGACCA
E	ACAGTG
F	GCCAAT
G	CAGATC
H	ACTTGA
I	GATCAG
J	TAGCTT
K	GGCTAC
L	CTTGTA

Tableau 6 – Exemple de plan des combinaisons index/barcodes utilisés lors de la PCR

25	9	17	25	9	17	25	9	17	25	9	17
2	10	18	2	10	18	2	10	18	2	10	18
3	11	19	3	11	19	3	11	19	3	11	19
4	26	22	4	26	22	4	26	22	4	26	22
5	13	20	5	13	20	5	13	20	5	13	20
6	14	21	6	14	21	6	14	21	6	14	21
7	15	23	7	15	23	7	15	23	7	15	23
8	16	24	8	16	24	8	16	24	8	16	24
IIIPCR2 - A			IIIPCR2 - B			IIIPCR2 - C			IIIPCR2 - D		

Tableau 7 – Mix de préparation pour la PCR

		concentrations stock	volumes à ajouter (µl)
Mix (30 µl)	Q5 buffer	5X	8
	dNTP	25mM chaque	0.32
	PCR primer mix (iIIPCR1 et iIIPCR2index)	5µM chaque	1.36
	Q5 Hot start DNA polymerase NEB	2u/µl	0.4
	High GC enhancer NEB	5X	8
	H2O		11.92
	ADN digéré-ligué-purifié		10
	TOTAL		30

Programme PCR :

98°C 30s

98°C 20s

60°C 30s

72°C 40s

72°C 10min

15 cycles

Tableau 8 – Mix de préparation pour le cycle final de la PCR

final cycle	concentrations stock	volumes à ajouter (µl)
Q5 buffer	5X	0.2
dNTP	25mM	0.08
Mix d'amorces (index iIIPCR1 et iIIPCR2)	5µM chaque	0.4
H2O		0.32
TOTAL à rajouter dans la plaque		1

Programme PCR :

98°C 3mn

60°C 2min

72° 12min

cycle final

De façon à s'affranchir des biais stochastiques de la PCR, trois PCR sont réalisées pour chaque échantillon, en parallèle et mélangées à l'issue de la PCR.

Etape 6 - Vérification par électrophorèse sur gel d'agarose

Le succès de la PCR est vérifié par électrophorèse de 5 μ l de produit de PCR sur gel d'agarose à 1.5 %, où les produits de PCR sont colorés avec du bromure d'éthidium et visualisés sous lumière UV (Figure 5).

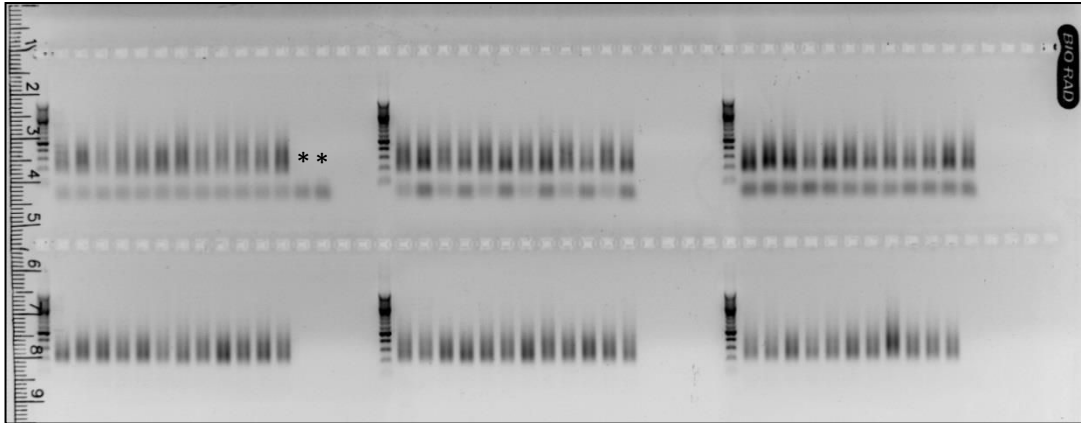


Figure 5 – Exemple de produits de PCR obtenus pour 74 échantillons dont 2 témoins négatifs(*), après électrophorèse sur gel d'agarose

Etape 7 - Mélange des individus

L'intensité relative des smears d'ADN obtenus dans l'étape précédente permet d'ajuster le volume de produit de PCR à ajouter pour chaque individu, afin de les réunir en une seule banque à séquencer. Un volume de 5 μ l a été prélevé par individu, diminué à 3 μ l si cet individu montrait une bande forte sur gel, et augmentée à 10 μ l si la concentration d'ADN pour un individu semblait faible.

Etape 8 - Sélection de taille

L'objectif de cette étape est de ne garder dans la banque que les fragments d'ADN inclus dans une gamme de taille précise (300-600) pour éliminer les amorces en excès après la PCR, les dimères d'amorces éventuellement produits par la PCR, les fragments d'ADN trop petits au regard de la longueur prévue pour le séquençage Illumina, ou trop grands pour permettre un séquençage optimal, et ceci de façon automatique, précise et répétable. 30 μ l du pool d'individus sont chargés dans une cassette de 1.5% Agarose, Dye-Free et soumis à une électrophorèse dans une blue pippin (SAGE SCIENCE). A la fin de cette étape, les produits correspondant à la gamme de taille souhaitée sont récupérés dans un volume d'environ 40 μ l.

Etape 9 - Contrôle qualité des banques

Un profil bioanalyser (Figure 6) est réalisé pour examiner la gamme de taille, estimer la taille moyenne (pb) des fragments et la concentration [en pg/ μ l] de chaque banque. Cette étape se fait à l'aide d'une puce microfluidique HS (High Sensitivity) sur un appareil Bioanalyser 2100 (Agilent).

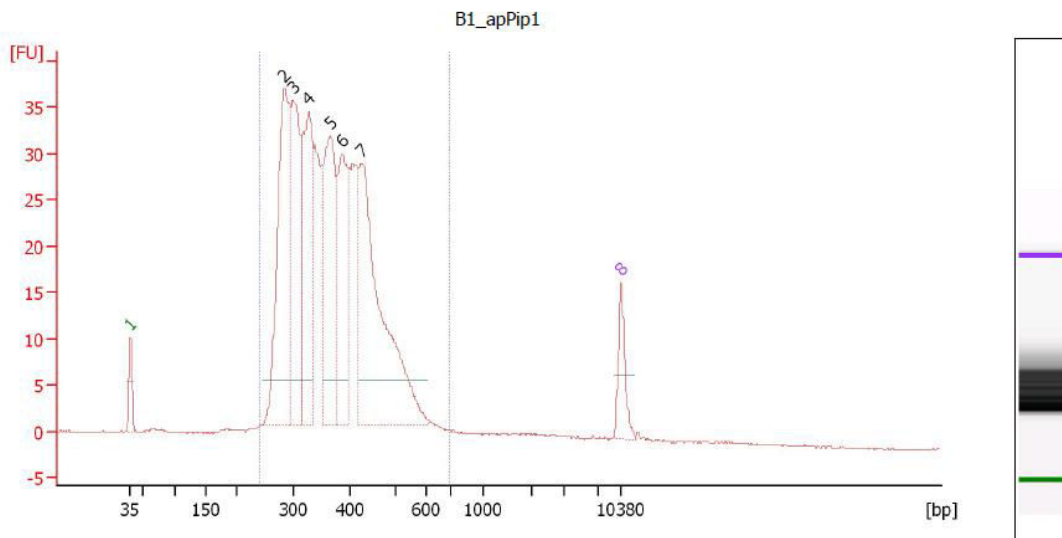


Figure 6 – Profil de migration de la Banque 1 obtenue à l'aide du Bioanalyser. Le premier et le dernier pic (1 et 8) correspondent aux marqueurs internes du kit DNA HS.

Chaque banque est également quantifiée par fluorimétrie, à l'aide d'un kit Qubit Broad Range, avec une dilution de l'ADN au $1/200^{\text{ème}}$, pour déterminer la concentration d'ADN double brin dans un échantillon.

Enfin, une PCR quantitative est réalisée car c'est la méthode la plus fiable pour d'une part vérifier que la PCR a inséré les séquences spécifiques des banques Illumina à chaque extrémité des fragments, et d'autre part en déterminer très précisément la concentration, seuls les fragments bien construits amplifiant avec les amorces fournies dans le kit Kapa Library Quantification kit (K-biosciences). Cette qPCR permet de mesurer la quantité initiale d'ADN, grâce au point seuil qui correspond au nombre de cycles à partir desquels on obtient un maximum d'intensité de bandes au cours d'une électrophorèse, le CT ou CP. Une courbe étalon est analysée en parallèle pour connaître l'efficacité de la PCR et calculer la concentration d'ADN de chaque échantillon. Pour cela, le kit kapa (K-biosciences) est utilisé et la PCR quantitative réalisée dans un thermocycler Lightcycler Roche 480, avec 4 dilutions de l'échantillon ($1/1000^{\text{ème}}$, $1/2000^{\text{ème}}$, $1/4000^{\text{ème}}$, $1/8000^{\text{ème}}$), et trois réplicas d'un volume de 10 μ L (4 μ L d'ADN et 6 % de mélange réactionnel).

Etape 10 – Séquençage sur HiSeq Illumina

La banque est enfin envoyée à séquençer à la concentration requise (ici un volume final de 20 μ l, à 10 nM) par le prestataire de séquençage Illumina sur HiSeq 2500.

Bilan des séquençages

Durant ma thèse, six banques ont été réalisées (en plus de la banque Miseq test), résumées dans le Tableau 9. Le séquençage réalisé était toujours du Single Read 125 avec une lecture d'index. Deux enzymes de restriction ont été utilisées, Pst1 et Mse1 et ces banques ont été réalisées en deux fois, les trois premières banques ont été séquencées en Juillet 2015 (séquençage par Eurofins), et les trois suivantes en Janvier 2017 (séquençage par Fasteris) (Figure 7). Tous les individus ont été mis en proportions équimolaires et les banques (coupées entre 300 et 600 pb lors du Pippin-prep) montraient des tailles moyennes de fragments et des concentrations d'ADN similaires entre elles (Tableau 9).

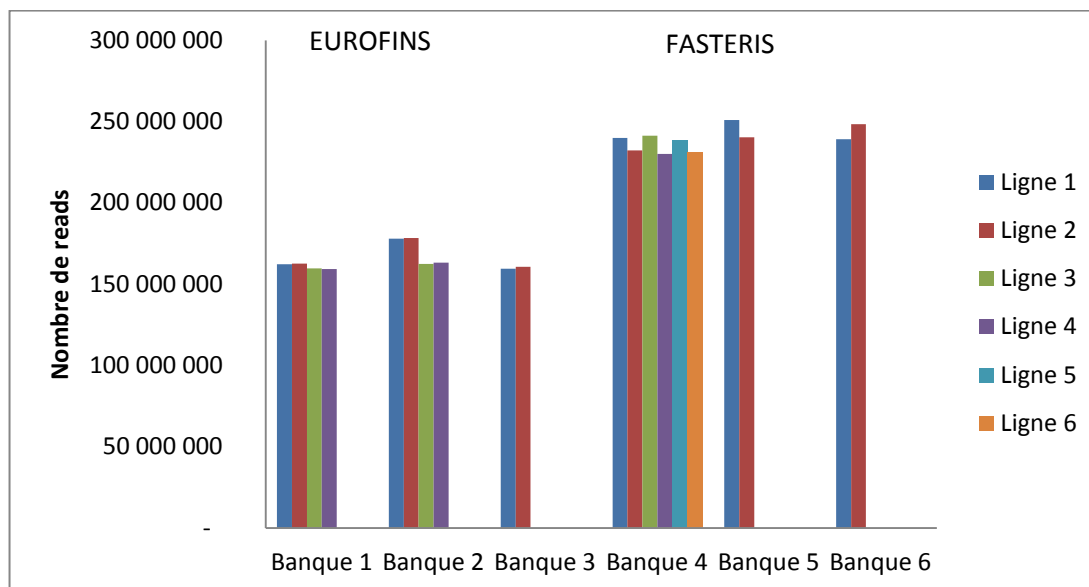


Figure 7. Résumé des nombres de reads obtenus après séquençage pour chacune des lignes des 6 banques

Au total, 963 échantillons ont été séquencés en 6 banques, comprenant 263 individus pour la cartographie, 185 individus pour la recherche de QTL (non analysés dans le cadre de cette thèse), et 469 individus (donc 379 mâles et 90 femelles) utilisés pour l'analyse de populations naturelles. Les populations naturelles comprenaient 244 *J. albifrons*, 126 *J. prae-hirsuta*, 29 morphotypes intermédiaires, 10 *J. forsmanni* (dont 1 morphotype intermédiaire avec *J. prae-hirsuta*), 10 *J. ischiosetosa* et 50 femelles du complexe *Jaera albifrons* non identifiés. En plus de tous ces échantillons cités, 3 échantillons de mouette ivoire (*Pagophila eburnea*, individus ajoutés pour une collaboration externe à cette étude), 19 témoins négatifs (10 témoins d'extractions, 6 de construction de banques et 3 de séquençage) ainsi que 24 réplicas (dont 16 réplicas inter-banques et 8 intra-banques) ont été séquencés.

En se basant sur les premières analyses effectuées, les témoins négatifs présentaient en moyenne 7 564 reads pour le premier séquençage (banques 1-3) et 22 986 reads moyen pour le second séquençage (banques 4-6), tandis que les individus présentaient en moyenne 3 241 927 lors du premier séquençage et 2 896 147 reads lors du second. De plus la bonne qualité du séquençage peut être visualisée par la ressemblance entre réplicas (plus de 96.5%) et le peu d'individus présentant des données manquantes (seulement 1.7 % d'individus ayant plus de 70% de données manquantes).

Tableau 9. Bilan des six banques séquencées

Nom de la banque	Expérience	Nombre d'individus multiplexés	qubit concentration (ng/ μ l)	qPCR concentration (ng/ μ l)	qPCR concentration (nM)	taille moyenne (pb)	Nombre de lignes HiSeq 2500	Prestataire
Banque 1	Analyse des populations naturelles	164	4.02	5.1	23.7	348	4	Eurofins Genomics
Banque 2	Analyse des populations naturelles	153	4.32	6.15	28.5	328	4	Eurofins Genomics
Banque 3	Analyse des populations naturelles	87	4.79	6.7	31.1	363	2	Eurofins Genomics
Banque 4	Cartographie	269	4.5	5.7	27.2	340	6	Fasteris
Banque 5	Recherche de QTL	191	4.2	6.9	32.8	339	2	Fasteris
Banque 6	Analyse des populations naturelles	99	3.24	4.3	21.4	323	2	Fasteris

ANNEXE 2 : Analyses bio-informatiques des séquences d'ADN

La route est longue et parsemée d'embuches entre la réception des séquences ADN et l'obtention de résultats propres. Cette annexe a pour objectif de rassembler et décrire le protocole bio-informatique utilisé, ainsi que des exemples de différents filtres et paramètres appliqués sur nos données, jusqu'à l'obtention d'un fichier final type (voir résumé Figure 1). Les scripts cités ci-dessous sont rassemblés à la fin de l'annexe. Cette partie du projet a été menée en collaboration avec Gildas Le Corguillé et la Plateforme Abims.

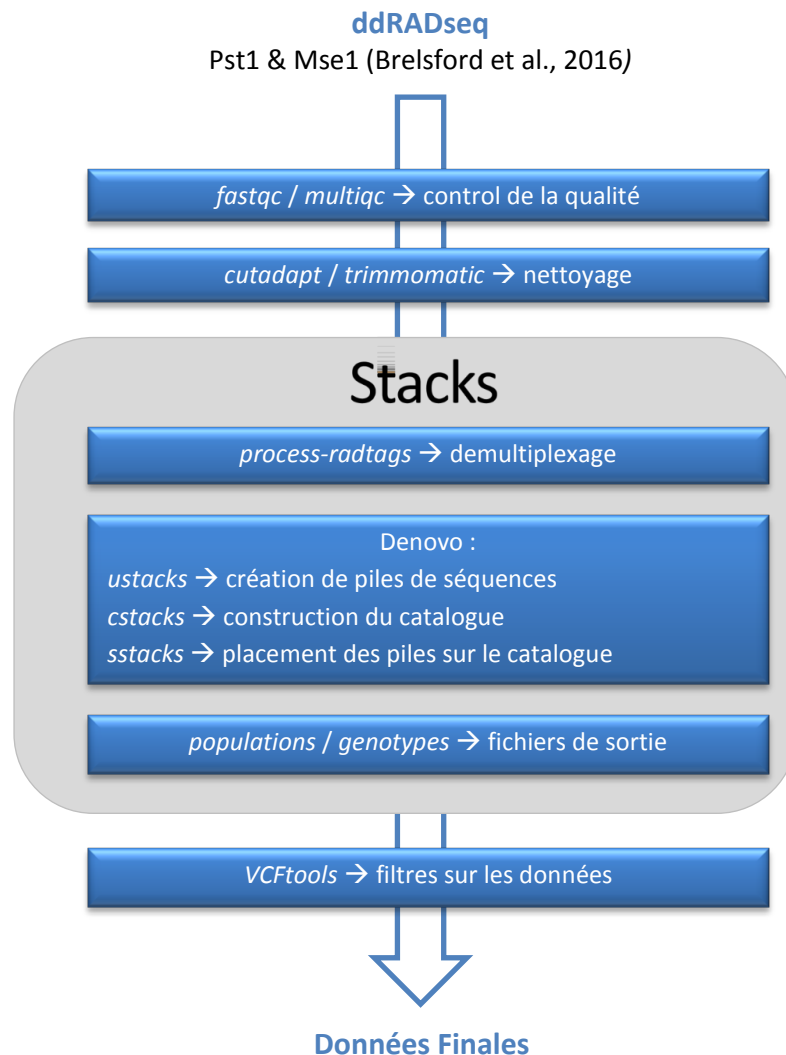


Figure 1. Schéma résumant les différentes étapes bio-informatique de l'acquisition des séquences au jeu de données finales

1. Vérifier la qualité des séquences à leur réception

script a

A la réception des séquences ADN, celles-ci sont groupées dans un fichier fastq.gz, pour chacun des index et des lignes de séquençage réalisé (2 à 6 lignes par banque) et mesurent 125 bases. En effet, les prestataires de séquençage réalisent eux-mêmes le démultiplexage des données selon l'index déterminé au cours du séquençage.

Une vérification de la qualité est tout d'abord effectuée, par analyse du score de qualité à l'aide du programme *fastqc* (Andrews). Ce score Q indique la probabilité qu'une base donnée ait mal été séquencée et est défini comme une propriété logarithmiquement liée à la probabilité d'erreur de séquençage (P)² : $Q = -10 \log_{10} P$. Un score de 30 (Q30) représente donc la probabilité d'une erreur sur 1000 bases séquencées, soit 99.9% de bases correctement séquencées. Le programme *multiqc* (Ewels et al., 2016) permet de compiler les sorties graphiques et les tableaux du programme *fastqc* pour chacun des fichiers fastq bruts. La distribution des scores de qualité le long des reads de 125 bases, pour l'ensemble des fichiers fastq de données brutes obtenus en 2017 de Fasteris, montrent que la qualité est très bonne, même aux extrémités des fragments (Figure 2). Une analyse similaire sur les données obtenues d'Eurofins en 2015 montre une distribution similaire (non montré).

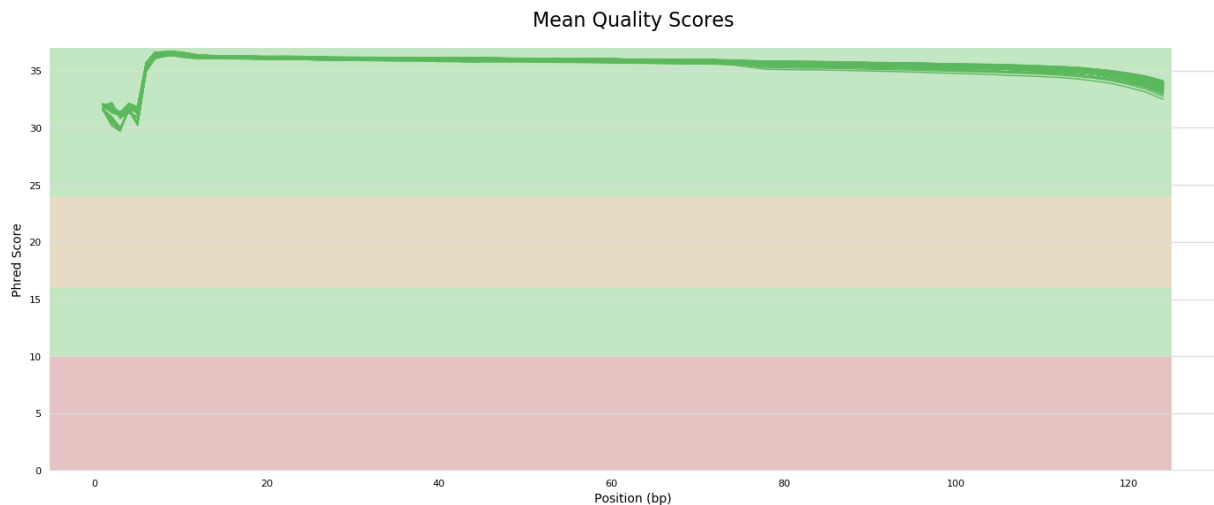


Figure 2. Distribution des scores de qualité par base le long des reads, compilé pour l'ensemble des 102 fichiers .fastq de données brutes obtenus de Fasteris en janvier 2017.

2. Couper les adaptateurs et raccourcir les séquences à 110 bases

script b

Le programme *fastqc* a permis également de détecter la présence des adaptateurs P2 illumina dans les données. Cela s'explique car pour certains des fragments courts composant la banque, le séquençage va au-delà du fragment d'ADN génomique inséré entre les adaptateurs. Le programme *cutadapt* (Martin, 2011) a été utilisé pour enlever les adaptateurs P2 des données brutes et les reads correspondant ainsi tronqués, produisant des reads de longueurs variables. Par la suite, et afin de maximiser le nombre de séquences sans trop diminuer leur taille, les séquences ont été coupées à 110 bases à l'aide du programme *trimmomatic* (Bolger et al., 2014). Le format fastq.gz des fichiers est conservé suite à ces deux étapes de nettoyage, de façon à minimiser le volume nécessaire au stockage des données sur le serveur ABIMS.

3. Stacks

Stacks (Catchen et al., 2013) est un logiciel permettant de construire des loci à partir de séquences courtes provenant de plusieurs individus.

a. Dé-multiplexer les données et réunir les séquences appartenant à un même individu

script c

Jusqu'ici, les séquences de plusieurs individus sont toujours réunies dans un même fichier fastq.gz. Il s'agit des individus portant le même index, et provenant d'une même ligne de séquençage, mais ayant des barcodes différents. Un démultiplexage des différents individus va donc être effectué à l'aide du programme *process-radtags inclus* dans *Stacks* (Catchen et al., 2013) et d'un fichier contenant le nom des individus et leur barcode pour chacun des index correspondant. Une attention particulière a été portée pour donner des noms d'individus à la fois informatifs et facilement manipulables, notamment pour traiter les répliques. Ensuite, les séquences d'un même individu mais de plusieurs lignes ont été compilées dans un même fichier, si bien qu'à l'issue de cette étape, on obtient pour chacun des individus inclus dans la banque, un seul et unique fichier fastq.gz portant le nom de l'individu.

b. Choix des paramètres

script d

Une fois les données dé-multiplexées, une des premières étapes consiste à aligner les séquences entre elles. Pour cette étape il est nécessaire d'appliquer certains paramètres tels que :

m - Profondeur minimale requise pour créer une pile de séquences (stack)

M - Distance maximale en termes de nucléotides autorisée entre les stacks pour les assembler (différences entre nucléotides pouvant être dues à du polymorphisme entre allèles, ou à des erreurs de séquençage).

Un moyen de choisir ces paramètres est de faire une pré-analyse de type m/M Explorer, script qui va permettre de quantifier pour chaque valeur de m et de M données (de 1 à 5 par exemple) le nombre de loci

(Tableau 1) et SNPs (Tableau 2) hétérozygotes qui seront obtenus en fonction des paramètres appliqués, afin de choisir le meilleur duo de valeurs de paramètres m/M.

Tableau 1. Nombre de locus hétérozygotes

m/M	M1	M2	M3	M4	M5
m2	6533	9538	10536	10968	11200
m3	6135	9090	10060	10479	10693
m4	5902	8808	9718	10169	10349
m5	5711	8568	9437	9846	10039
m6	5526	8285	9179	9577	9781
m7	5378	8029	8895	9293	9496
m8	5210	7783	8620	9011	9198
m9	5033	7544	8344	8726	8936
m10	4889	7327	8116	8480	8650

Tableau 2. Nombre de SNPs hétérozygotes

m/M	M1	M2	M3	M4	M5
m2	7491	16025	21990	26116	29217
m3	7029	15606	21158	25372	28265
m4	6745	15212	20529	24704	27420
m5	6535	14985	20047	24119	26907
m6	6325	14432	19568	23417	26479
m7	6160	13916	18855	22884	25777
m8	5976	13608	18698	22156	25013
m9	5774	13184	17810	21566	24369
m10	5608	12861	17605	21061	23498

Plus on augmente la valeur de m (profondeur de couverture) moins nous avons de données mais plus elles sont fiables (moins de risque que ce soit dû à des erreurs de séquençage). Plus la valeur de M (distance entre stacks) est augmentée, plus nous avons de données, mais il est nécessaire ici de faire attention afin de ne pas assembler plusieurs stacks différents. Pour cet exemple, un duo m5/M5 aurait probablement été gardé.

c. *Denovo-map*

script e

Trois étapes sont nécessaires pour obtenir des loci à partir des séquences de départ, dans le cas où aucun génome de référence n'est connu. Le programme *ustacks* permet d'assembler les séquences similaires en piles (stacks), *cstacks* permet de construire un catalogue en créant des loci consensus grâce à l'assemblage des allèles préalablement déterminés. Enfin *sstacks* permet de placer les loci obtenus avec *ustacks* en fonction du catalogue obtenu par *cstacks*. Il est possible de faire tourner ces programmes à la suite les uns des autres à l'aide du programme *denovo-map*. Les paramètres utilisés dans les différentes analyses sont :

Ustacks :

- m : Profondeur minimale requise pour créer une pile de séquences (stack)
- M : Distance maximale en termes de nucléotides autorisée entre les stacks
- N : Distance maximale autorisée pour ajouter les séquences secondaires (pas assez nombreuses pour créer une pile lors de l'étape précédente) aux stacks déjà créés (par défaut M+2)

Attention, lors de *ustacks* il est nécessaire de donner un nom unique à chaque échantillon (-i) afin de pouvoir retrouver ces individus sur une database SQL, en ligne. Cela est notamment nécessaire lors de l'analyse de familles.

Cstacks :

- n : nombre de nucléotides différents autorisés entre les loci lors de la construction du catalogue.
- Lors de cette étape il est important de choisir n en fonction de la divergence des espèces analysées. Dans le cas de familles génétiques, le catalogue sera créé uniquement à partir des parents.

d. *Rxstacks*

script f

Le programme *rxstacks* permet de faire une correction sur les génotypes et haplotypes détectés individuellement grâce à une analyse à l'échelle de la population. Trois types de corrections ont été utilisés par ce programme.

- Correction du modèle : *rxstacks* réévalue les variants qui n'étaient pas significatifs pour chaque individu, en utilisant le catalogue pour rechercher une position de nucléotide particulière dans la population et filtrer les erreurs de séquençage putatives de cet individu (avec une limite de taux d'erreur dans mes analyses de 5%).

- Filtre sur le log-likelihood : lors de *ustacks*, une probabilité lors de l'assemblage de la pile est donnée. Ce programme permet de filtrer les loci ayant un log-likelihood en dessous d'un seuil défini car ils auront tendance à avoir une faible couverture et/ou un taux d'erreur de séquençage élevé. Plus le seuil est proche de zéro, plus la probabilité que les loci sont incorrect sera faible. Ainsi pour l'exemple donné par la Figure 3, le seuil retenu serait de -8.

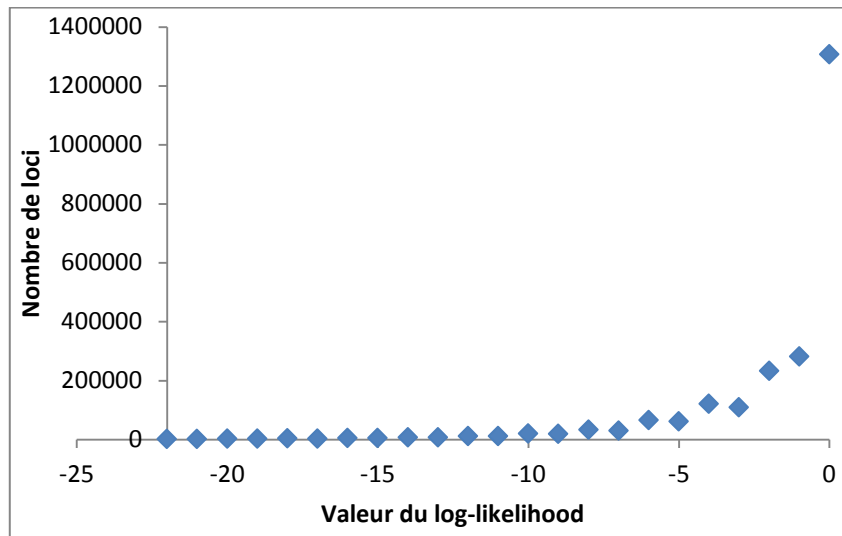


Figure 3. Exemple de graphique montrant le nombre de loci en fonction la valeur du log-likelihood et permettant le choix du seuil à appliquer sous *rxstacks*.

- Filtre sur les loci mal attribués: Il est attendu que dans *sstacks*, chaque individu ait un seul locus correspondant à celui du catalogue. Il arrive parfois cependant que plusieurs loci d'un individu correspondent à un seul locus de catalogue, et il faut donc éliminer le locus du catalogue si celui présente plus d'un certain pourcentage d'individus ayant plusieurs loci pouvant correspondre à ce locus (seuil fixé à 75% des individus dans mes analyses).

e. cstacks + sstacks

script g

Après les corrections sous *rxstacks* il est important de refaire tourner *cstacks* puis *sstacks* afin de mettre à jour les données finales avec ces corrections. Les mêmes paramètres qu'utilisés dans le *denovo-map* ont été appliqués.

f. Populations ou Genotypes

script h

- *populations* : ce programme est utilisé pour analyser des populations d'individus. Un fichier texte avec le nom de la population à laquelle appartient chaque individu est nécessaire. *Populations* permet d'obtenir des fichiers de sorties utiles pour certains programme de génétiques des populations, tels que *Genepop*, *Structure*, *Treemix*, mais il peut aussi être utilisé pour générer des statistiques sur ces populations. Le fichier de sortie habituel après *populations*, et utilisable dans beaucoup d'autres programmes d'analyses

génomiques, est un vcf. Enfin, il est possible de demander à ne garder aléatoirement qu'un SNP par tag, utile pour éviter les déséquilibres de liaison forts dans nos analyses.

- *genotypes* : ce programme, utile pour l'étude de familles génétiques, exporte les données comme un ensemble d'haplotypes encodés dans les génotypes. L'option -r permet de ne garder les loci existants que chez un certain nombre de juvéniles, et l'option -m peut être utilisée pour retenir les loci ayant une certaine profondeur de couverture minimale.

4. *Vcftools*

script i

VCFtools (Danecek et al., 2011) est un programme qui permet de comparer, filtrer ou encore convertir des données de type vcf, tel qu'obtenu par *populations*. Dans nos analyses il est utilisé après chaque étape *populations* pour soit :

- filtrer les loci sur un nombre d'individu manquants dans chaque population avec l'option --max-missing
 - filtrer les loci n'étant pas à l'équilibre d'Hardy-Weinberg dans chaque population avec l'option --hwe
 - filtrer le nombre d'allèle minimum autorisé sur l'ensemble des populations réunies avec l'option --mac
- Certain de ces filtres doivent être appliqués population par population, il faut donc veiller à séparer les populations avant filtres puis les réunir après (merge) afin de continuer les analyses sur l'ensemble des populations.

Script a

```
fastqc -o . --extract --threads 16 *.fastq.gz
```

Script b

Pour couper les adaptateurs :

```
for i in $(ls *.fastq.gz); do qsub cutadapt.qsub $i; done
```

```
cutadapt.qsub : cutadapt --adapter=TTAAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -e 0.1 -O 8 -o  
$1.cutadapt.fastq.gz $1
```

avec :

-e : taux d'erreur admis (10%)

-O nombre de bases minimum pour détecter l'adaptateur

-o : nom du fichier de sortie

Pour couper les séquences à 110 pb :

```
for i in $(ls *.fastq.gz); do qsub trimmomatic110.qsub $i; done
```

```
trimmomatic110.qsub : trimmomatic SE -threads 10 -phred33 $1 $1_trim.fq.gz MINLEN:110 CROP:110
```

avec :

MINLEN:110 : ne garder que les séquences faisant au moins 110 pb

CROP:110 : couper les séquences faisant plus de 110 pb à cette taille.

Script c

Pour démultiplexer (exemple de la Banque 1 "Lib1") :

```
for i in $(ls Lib1*); do qsub process_radtags_lib1.qsub $i; done
```

```
process_radtags_lib1.qsub : process_radtags -i gzfastq -f $1 -o $1"_pr" -y gzfastq -b barcodes_lib3.txt --  
inline_index -q -D --renz_1 pstl --renz_2 msel --adapter_1  
TAAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
```

Pour pooler les individus ensemble :

```
for i in $(ls -d *_pr | sed -r "s/_L0.*$//" | sort | uniq); do for j in $(ls $i*/*.gz | sed -r L00[12345678]/$/" |  
sort | uniq); do zcat $(echo $j | sed -r 's/\/$/*/') | gzip -c > $(basename $j);done ;done
```

Script d

```
for m in `seq 1 5`; do for M in `seq 1 5`; do qsub -v m=$m -v M=$M -t 1-`ls *.fq.gz | wc -l`
/projet/sbr/isobar/15-09-popnat/script/stacks_ustacks_array.qsub; done; done
```

stacks_ustacks_array.qsub:

```
INPUT=`ls *.fq.gz | awk "NR==$SGE_TASK_ID"
```

```
OUTPUTDIR="1-ustacks-m"$m"-M"$M
```

```
mkdir $OUTPUTDIR 2> /dev/null
```

```
mkdir $OUTPUTDIR/log 2> /dev/null
```

```
echo -e `date '+%y%m%d-%H:%M'`\t"$SGE_TASK_ID"\t"$INPUT >> $OUTPUTDIR/qsub_array_files.tab
```

```
ustacks -t gzfastq -f $INPUT -i $SGE_TASK_ID -o $OUTPUTDIR -m $m -M $M -N 0 -p 1 -d -r --
max_locus_stacks 2 > $OUTPUTDIR/log/$INPUT.ustacks.stdout 2> $OUTPUTDIR/log/$INPUT.ustacks.stderr
&&
```

```
mv $OUTPUTDIR/log/$INPUT.ustacks.stderr $OUTPUTDIR/$INPUT.ustacks.log
```

créer les tableaux :

```
stacks_ustacks_stats_mM.sh 1-*
```

stacks_ustacks_stats_mM.sh :

```
inds=$(for file_i in $(ls $1/*.alleles.tsv); do echo -ne " $(basename $file_i .alleles.tsv) ; done)
```

```
stats_nb_locus_hetero=stats_ustacks_nb_locus_hetero.tab
```

```
echo -e "dir\tm\tM\t$(echo $inds | sed "s/ /\t/g")> $stats_nb_locus_hetero
```

```
for ustacks_dir in $*; do
```

```
    m=`echo $ustacks_dir | sed -r "s/^.*-m([1-5]+).*/\1/"`
```

```
    M=`echo $ustacks_dir | sed -r "s/^.*-M([1-5]+).*/\1/"`
```

```
    echo -en $ustacks_dir"\t"$m"\t"$M >> $stats_nb_locus_hetero
```

```
    for alleles_file in `ls $ustacks_dir/*.alleles.tsv`; do
```

```
        echo -en "\t"\t"cut -f 3 $alleles_file | uniq | wc -l` >> $stats_nb_locus_hetero
```

```
    done
```

```
    echo "" >> $stats_nb_locus_hetero
```

```
done
```

```
$stats_nb_locus_hetero
```

```
echo -e $stats_nb_locus_hetero"\tdone"
```

```
stats_nb_locus_hetero_matrix=stats_ustacks_nb_locus_hetero_matrix.tab
```

```
m_min=`sed "1d" $stats_nb_locus_hetero | cut -f 2 | sort -n | head -n1`
```

```
m_max=`sed "1d" $stats_nb_locus_hetero | cut -f 2 | sort -nr | head -n1`
```

```
M_min=`sed "1d" $stats_nb_locus_hetero | cut -f 3 | sort -n | head -n1`
```

```
M_max=`sed "1d" $stats_nb_locus_hetero | cut -f 3 | sort -nr | head -n1`
```

Annexe 2

```
(echo -en "m/M\t";seq $M_min $M_max | sed "s/^/M/" | tr "\n" "\t";echo "") >
$stats_nb_locus_hetero_matrix
for m in `seq $m_min $m_max`; do
    echo -en "m"$m >> $stats_nb_locus_hetero_matrix
    for M in `seq $M_min $M_max`; do
        echo -ne "\t" >> $stats_nb_locus_hetero_matrix
        awk -v m=$m -v M=$M '$2 == m && $3 == M { printf $4}' $stats_nb_locus_hetero >>
$stats_nb_locus_hetero_matrix
    done
    echo "" >> $stats_nb_locus_hetero_matrix
done
echo -e $stats_nb_locus_hetero_matrix"\tdone"
stats_nb_snps=stats_ustacks_nb_snps.tab
echo -e "dir\tm\tM\t$(echo $inds|sed "s/ /\t/g")> $stats_nb_snps

for ustacks_dir in $*; do
    m=`echo $ustacks_dir | sed -r "s/^.*-m([0-9]+).*/\1/"`
    M=`echo $ustacks_dir | sed -r "s/^.*-M([0-9]+).*/\1/"`
    echo -en $ustacks_dir"\t"$m"\t"$M >> $stats_nb_snps
    for snps_file in `ls $ustacks_dir/*.snps.tsv`; do
        echo -en "\t"`awk '$5 == "E" { count ++ } END { print count }' $snps_file` >>
$stats_nb_snps
    done
    echo "" >> $stats_nb_snps
done
echo -e $stats_nb_snps"\tdone"

stats_nb_snps_matrix=stats_ustacks_nb_snps_matrix.tab
m_min=`sed "1d" $stats_nb_snps | cut -f 2 | sort -n | head -n1`
m_max=`sed "1d" $stats_nb_snps | cut -f 2 | sort -nr | head -n1`
M_min=`sed "1d" $stats_nb_snps | cut -f 3 | sort -n | head -n1`
M_max=`sed "1d" $stats_nb_snps | cut -f 3 | sort -nr | head -n1`
(echo -en "m/M\t";seq $M_min $M_max | sed "s/^/M/" | tr "\n" "\t";echo "") > $stats_nb_snps_matrix
for m in `seq $m_min $m_max`; do
    echo -en "m"$m >> $stats_nb_snps_matrix
    for M in `seq $M_min $M_max`; do
        echo -ne "\t" >> $stats_nb_snps_matrix
```

```
awk -v m=$m -v M=$M '$2 == m && $3 == M { printf $4}' $stats_nb_snps >>
$stats_nb_snps_matrix
done
echo "" >> $stats_nb_snps_matrix
done
echo -e $stats_nb_snps_matrix"\tdone"
```

Script e

```
denovo_map="denovo_map.pl"
```

```
m=5
```

```
M=5
```

```
N=7
```

```
n=5
```

```
bound_high=0.05
```

```
popmap=pop.txt
```

```
Xs="-X populations:--vcf "
```

```
fqPATH="."
```

```
OUTPUTDIR="."
```

```
mkdir $OUTPUTDIR 2> /dev/null
```

```
$denovo_map $(ls $fqPATH/*.fq.gz | sed "s/^/ -s /" | tr '\n' ' ') -o $OUTPUTDIR -O $fqPATH/$popmap -T
$thread -m $m -M $M -N $N -n $n --bound_high $bound_high -X "ustacks:--max_locus_stacks 2" $Xs -b 1 -S
> $OUTPUTDIR/denovo_map.out 2> $OUTPUTDIR/denovo_map.err
```


Script f

Obtenir le fichier pour estimer le filtre de log-likelihood à appliquer :

```
rxstacks -b 1 -P rxstacks -o Inls --model_type bounded --bound_high 0.05 --lnl_dist
```

Faire tourner rxstacks avec les différents filtres :

```
rxstacks -b 1 -P rxstacks --model_type bounded --bound_high 0.05 --lnl_filter --lnl_lim -15.0 --lnl_dist --  
conf_filter --conf_lim 0.75
```

Script g

Exemples pris de l'analyse de cartographies

cstacks :

```
cstacks -s F1_father_1499 -s F1_mother_1498 -s F2_father_1501 -s F2_mother_1500 -s F3_father_1503 -s  
F3_mother_1502 -s F4_father_E1334 -s F4_mother_E1316 -o . -b 5 -n 3 -p 15
```

sstacks :

```
sstacks -b 5 -c batch_5 -o . -p 10 -s juvénile1 -s juvénile2 -s juvenile3 ...
```

Script h

populations :

```
populations -b 1 -P . -M ../popmap_males.tab --vcf --treemix --genepop -t 16
```

genotypes :

```
genotypes -b 5 -P . -r 30 -c -t 'GEN' -o 'joinmap' -s
```

Script i

VCFtools : exemple à partir d'une population *J. albifrons* de Bretagne (merge effectué sur 5 populations différentes)

```
vcftools --vcf batch_1.vcf --keep Pop1.txt --out BrAl --recode
```

```
vcftools --vcf BrAl.recode.vcf --max-missing 0.8 --hwe 0.01 --out BrAl.filtered --recode
```

```
bgzip BrAl.filtered.recode.vcf
```

```
tabix -p vcf BrAl.filtered.recode.vcf.gz
```

```
vcf-merge BrAl.filtered.recode.vcf.gz BrPr.filtered.recode.vcf.gz NoAl.filtered.recode.vcf.gz  
NoPr.filtered.recode.vcf.gz NoHy.filtered.recode.vcf.gz >merge5pop.vcf  
grep -v "^#" merge5pop.vcf | grep -e "SF=0,1,2,3,4" > SNP5pop.vcf  
→ Attention les en-têtes ont disparus, les rajouter grâce au fichier précédent  
(grep "#" merge5pop.vcf; cat SNP5pop.vcf)> tot5pop_head.vcf
```

Un fichier final filtré par populations a été obtenu, il est maintenant possible de faire d'autres filtres tel qu'un « minimum allele count » ou des statistiques descriptives sur ce jeu de données :

```
vcftools --vcf tot5pop_head.vcf --mac 2 --out Fin_5pop_mac2 --recode  
vcftools --vcf tot5pop_head.vcf --out 5pop_depth --depth  
vcftools --vcf tot5pop_head.vcf --out 5pop_smdepth --site-mean-depth
```

ANNEXE 3 : Scripts relatifs à DADI

Dans cette annexe sont réunis les scripts utilisés lors du chapitre 5, afin d'obtenir *via* le programme $\delta\delta i$ des spectres joints de fréquences alléliques. Ceux-ci ont été modifiés d'après Pierre-Alexandre Gagnaire et Clément Rougeux.

1/ Convertir le vcf en fichier lisible par Dadi

```
perl convert_vcf_to_dadi_input.pl <genome file> <vcf file> <list file>
```

→ récupération d'un fichier de type « file.data »

avec

- **convert_vcf_to_dadi_input.pl** : *script a*

- **<genome file>**

Ce fichier fonctionne quand on a un génome de référence. Comme la ce n'est pas le cas dans notre cas, il faut créer une séquence alignant tous les reads tel un chromosome qu'on appellera ici seqcons.fa. Pour cela, il faut utiliser le fichier « batch_1.catalog.tags.tsv » et faire `grep -v "^#" batch_1.catalog.tags.tsv | cut -f 10 > seqcons.fa`

Enfin, il faut ajouter en en-tete « >un » à ce fichier.

- **<vcf file>**: Le fichier vcf obtenu après vcftools

- **<list file>**: liste d'individus et de populations

2/ Editer le fichier « file.data » afin de bien gérer l'information provenant de l'outgroup.

Pour cela, utiliser le script R **Code_dadi_dat.r** : *script b*

Attention, le fichier de sortie .data doit être sous la forme Allele1 / Pop1 / Pop 2 / outgroup / Allele2 / Pop1 / Pop2/ outgroup. A la fin de R enlever la 1ere colonne et les guillemets du document.

→ un fichier de type « file.edited.data » est obtenu.

3/Activer Dadi

```
./usr/local/genome2/dadi_venv_1.6.3_gensealab/bin/activate
```

4/ Construire le spectre de fréquences alléliques observées

```
python fs_from_data.py
```

avec

- **fs_from_data.py** : *script c*

Indiquer dans ce script le nom des populations, et le nombre n d'observations par populations.

- file.edited.data : le fichier obtenu avec R, sur laquelle le script fs_from_data.py va travailler

→ on obtient alors un fichier et une figure sfs_file_data.fs

5/ Faire tourner les simulations des modèles d'inférences démographiques

bash *modèle.sh*

avec

- **bash *modèle.sh* : script *d***

Dans ce script il faut appeler le script `inference_anneal_models.py` et indiquer le modèle démographique que l'on souhaite faire tourner avec le nombre de runs, les populations étudiées, le nom du spectre des fréquences alléliques observées, et la taille des cases du spectre alléliques théorique que l'on veut `[n,n+10,n+20]`

- **inference_anneal_models.py : script *e***

Script appelé par la commande précédente, permettant de tester les différents modèles démographiques à partir de nos données. Va lui-même appeler le script `modeledemo.py`. Il faut préciser les noms des populations, des modèles que l'on souhaite faire tourner, le nom du spectre des fréquences alléliques observées, et la taille des cases du spectre alléliques théorique que l'on veut `[n,n+10,n+20]`.

- **modeledemo.py : script *f***

Module `modeledemo` contenant les différents modèles démographiques de divergence.

6/ Trier les fichiers et choisir le meilleur modèle

#créer un dossier par modèle

```
mkdir modèle_all
```

#copier tous les fichiers des différents runs dans ce dossier

```
In ./modèle_masked*/.*_AM_all/
```

#aller chercher tous les fichiers .txt d'un dossier en extraire les lignes d'intérêts puis formater le fichier manuellement en retirant les retours de lignes et remplaçant les séparateurs par des tabulations permettant d'obtenir uniquement les valeurs en lignes

```
for i in $(find . | grep -E '_masked_.*\.txt$'); do grep -A 1000 BFGS "$i" | grep -A 1000 parameters | perl -pe 's/,\\n/,/' | perl -pe 's/^\\D+//; s/^\\D+$/\\n/; s/,?[\\t]\\t/g' | perl -pe 's/^\\t+\\t/g' > "$i".extracted.txt; done
```

obtenir les paramètres démographiques

```
for i in $(find . 'extracted.txt$'); do head -n 1 "$i" > "$i"_firstline.txt; done
```

```
cat ./modèle_all/*extracted.txt_firstline* > ./modèle_all/Param_modèle.txt
```

```
rm -f ./*/*_firstline.txt
```

#obtention du MLE

```
for i in $(find . 'extracted.txt$'); do head -n 2 "$i" | tail -1 > "$i"_secondline.txt; done
```

```
cat ./modèle_all/*extracted.txt_secondline* > ./modèle_all/MLE_modèle.txt
```

```
rm -f ./*/*_secondline.txt
```

#obtention du Theta

```
for i in $(find . 'extracted.txt$'); do head -n 3 "$i" | tail -1 > "$i"_thirdline.txt; done
```

```
cat ./modèle_all/*extracted.txt_thirdline* > ./modèle_all/THETA_modèle.txt
```

```
rm -f ./*/*_thirdline.txt
```

#obtention de l'AIC

```
for i in $(find . 'extracted.txt$'); do head -n 4 "$i" | tail -1 > "$i"_fourthline.txt; done
cat ./modèle_all/*extracted.txt_fourthline* > ./modèle_all/AIC_modèle.txt
rm -f ./*/*_fourthline.txt
#obtention d'un fichier global, regroupant paramètres démographiques, MLE, Theta,AIC
paste Param_modèle.txt MLE_modèle.txt THETA_modèle.txt AIC_modèle.txt > texte.txt
cat texte.txt | sort -k7n,7n > modèle_final.txt
```

7/Vérifier l'absence d'erreurs lors de la construction des modèles

→ le spectre des données observées doit être le même que celui obtenu avant inférences des modèles, et vérifier que les résultats des différents runs d'un même modèle sont constants.

Enfin, récupérer le run de chaque modèle ayant l'AIC le plus faible, *i.e.* le meilleur run, dans un dossier final.

8/ Repérer le modèle parmi les 13 ayant l'AIC le plus faible, ainsi que ceux ayant un $\Delta AIC < 10$.

Pour ces modèles, générer 1000 bootstraps afin d'obtenir des intervalles de confiance aux données théoriques obtenues pour chaque modèle.

perl dadi_boot.pl in=pop.edited.list.data boot=1000

avec

- **dadi_boot.pl** : *script g* permettant de générer les bootstraps, avec comme arguments :

in=[filename]

boot=[integer]: nombre de bootstraps (100 par défaut)

python Fs_from_boot.py

avec

- **Fs_from_boot.py** : *script h* permettant de transformer les données des bootstraps en spectres de fréquences alléliques

Indiquer les populations et la taille des cases attendue du spectre allélique.

Python script_uncertainty_analysis_modèle.py

avec

- **script_uncertainty_analysis_modèle.py** : *script i*, permettant d'appliquer les bootstraps aux valeurs des meilleurs modèles obtenus

Changer :

- le nom du modèle

- grid_pts

- les résultats démographiques obtenus pour le meilleur modèle p0[paramètres]

- le chemin d'accès à nos données observées

→ on obtient un fichier 'modèle_uncert_masked.txt' avec les intervalles de confiance pour chaque valeur estimée du modèle théorique.

ANNEXE 4 : Comparaison de spectres joints dépliés vs repliés

Cette annexe regroupe les premiers résultats d'analyses démographiques réalisées afin de tester la méthode sur nos données (spectres joints sans *outgroup*). Les différentes étapes ont été supervisées par Pierre-Alexandre Gagnaire.

Avant d'obtenir des spectres joints dépliés grâce à l'ajout d'un *outgroup* *J. ischiosetosa*, nous avons obtenu des spectres joints repliés, *i.e.* sans distinction des allèles ancestraux des allèles dérivés par ajout d'un *outgroup*, en Bretagne et en Normandie, que nous avons comparés avec les 7 modèles démographiques décrits dans Tine et al., 2014 (Figure 1).

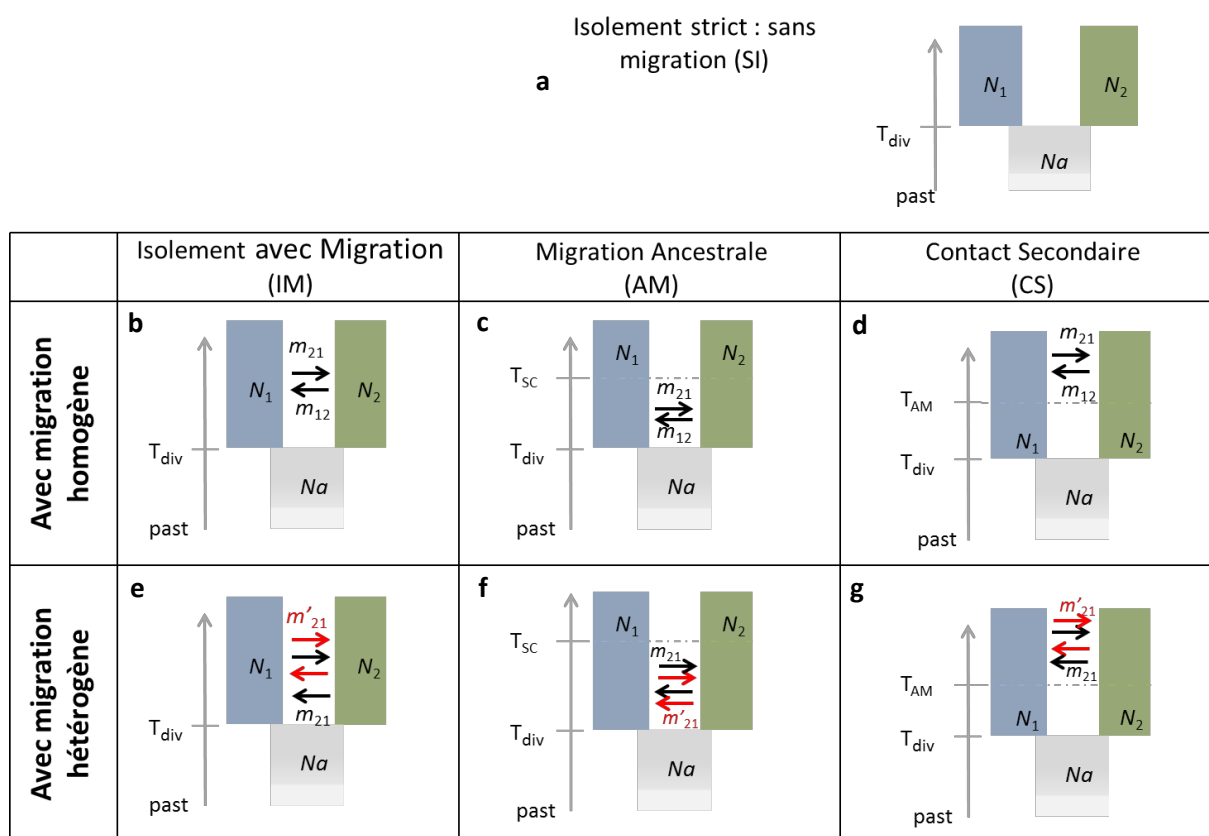


Figure 1. Modèles simplifiés pour inférer l'histoire démographiques entre *J. albifrons* et *J. praeirsuta*. Le modèle basique est l'isolement strict (a) où 2 espèces (N_1 et N_2) émergent à partir d'une espèce mère (N_a). Les autres modèles décrits sont l'isolement avec migration (b,e), la migration ancestrale (c,f) et le contact secondaire (d,g). Certains de ces modèles sont implémentés avec une migration homogène des loci le long du génome (b,c,d) ou avec une migration hétérogène le long du génome (e,f,g). Modifié d'après Rougemont et al. (2016a).

En Bretagne, 4 sites avaient été retenus : Bouillenou, StJean, Lingoz et Tanguy (sites x-y sur fig. x, chapitre z), réunissant 53 *J. albifrons* et 46 *J. praeheirsuta*.

En Normandie, 60 *J. albifrons* et 44 *J. praeheirsuta* provenant des sites Grandcamp, SteHonorine, Longues et Yport (sites 30, 32, 33 et 37 sur la Figure 1 du Chapitre 2) avaient été gardés pour les analyses.

Après nettoyage des données, un *denovo-map* avait été réalisé avec comme paramètres $m=3$, $M=3$, $N=5$ et $n=3$. *VCFTools* avait ensuite été directement utilisé pour filtrer les données avec comme paramètres : $\text{max-missing} = 0.8$ et $\text{HW} = 0.01$ pour chacune des deux populations de Bretagne ; et seulement $\text{max-missing} = 0.8$ pour chacune des deux populations de Normandie. Après avoir fusionné les deux populations au sein de chaque région, un SNP par RAD avait été retenu. En Bretagne 12 183 SNPs avaient donc été utilisés pour l'obtention des spectres joints de fréquences alléliques contre 11 834 SNPs en Normandie.

En Bretagne et en Normandie, un spectre de fréquences alléliques replié a été réalisé avec une résolution basée sur $2N=50$. Ce spectre a été comparé sur 20 runs pour chacun des 7 modèles démographiques théoriques décrits plus haut. Que ce soit en Bretagne (Figure 2a) ou en Normandie (Figure 2b) le meilleur modèle démographique expliquant nos données était le contact secondaire avec migration hétérogène (SC2m), sans doute possible parmi les 7 modèles, les 20 runs des 6 autres modèles ayant tous un $\Delta\text{AIC} > 10$ de celui-ci. Ce qui apparaît lorsque l'on compare les données obtenues (Fig. 4), c'est la grande ressemblance entre données obtenues et données théoriques tant d'un point de vue visuel, que par la distribution des résidus.

L'analyse des paramètres obtenus pour ce modèle dans chacune des deux régions (Tableau 1) permet de faire plusieurs prédictions sur les histoires démographiques passées entre *Jaera albifrons* et *Jaera praeheirsuta*, basées sur les rapports entre les tailles de populations et les taux de migration.

Deux méthodes d'inférence de l'histoire de la divergence par $\delta a \delta i$ ont été utilisées rendant les analyses des résultats complexes car les modèles utilisés dans les deux méthodes ne sont pas les mêmes, et l'une a été réalisée sans *outgroup* (spectres repliés, présentés ici) et la seconde avec l'ajout de l'*outgroup J. ischiosetosa* (spectres dépliés, chapitre 5). Cependant, et dans tous les cas, *i.e.* dans chaque modèle de chaque région, le meilleur modèle obtenu étant le contact secondaire, nous pouvons tout de même faire une étude comparative des résultats obtenus par les différentes méthodes au sein de chaque région, afin de vérifier si l'information globale est sensible ou non à la méthode utilisée.

La valeur de θ étant dans tous les cas plus forte en Normandie qu'en Bretagne (spectres dépliés : $\theta_N = \theta_B \times 6$, spectres repliés : $\theta_N = \theta_{NB} \times 2$) nous ne comparerons que les paramètres qui ne dépendent pas directement de θ . En Bretagne comme en Normandie, le temps passé en allopatrie (T_{div}) représente 80% à 90 % du temps depuis la divergence, et la durée relative du contact secondaire ($\frac{T_{\text{sc}}}{T_{\text{div}}+T_{\text{sc}}}$) semble relativement équivalente en Bretagne et Normandie, respectivement 17% vs 11% concernant les spectres dépliés et 7.9% vs 6.4% concernant les spectres repliés. La proportion de loci soumis à un taux de migration efficace neutre est équivalente avec les deux méthodes, soit 37-49% du génome en Bretagne et 67% du génome en Normandie. Pour tous les modèles (analyses dépliées et analyses repliées), il y a une asymétrie marquée dans le taux de

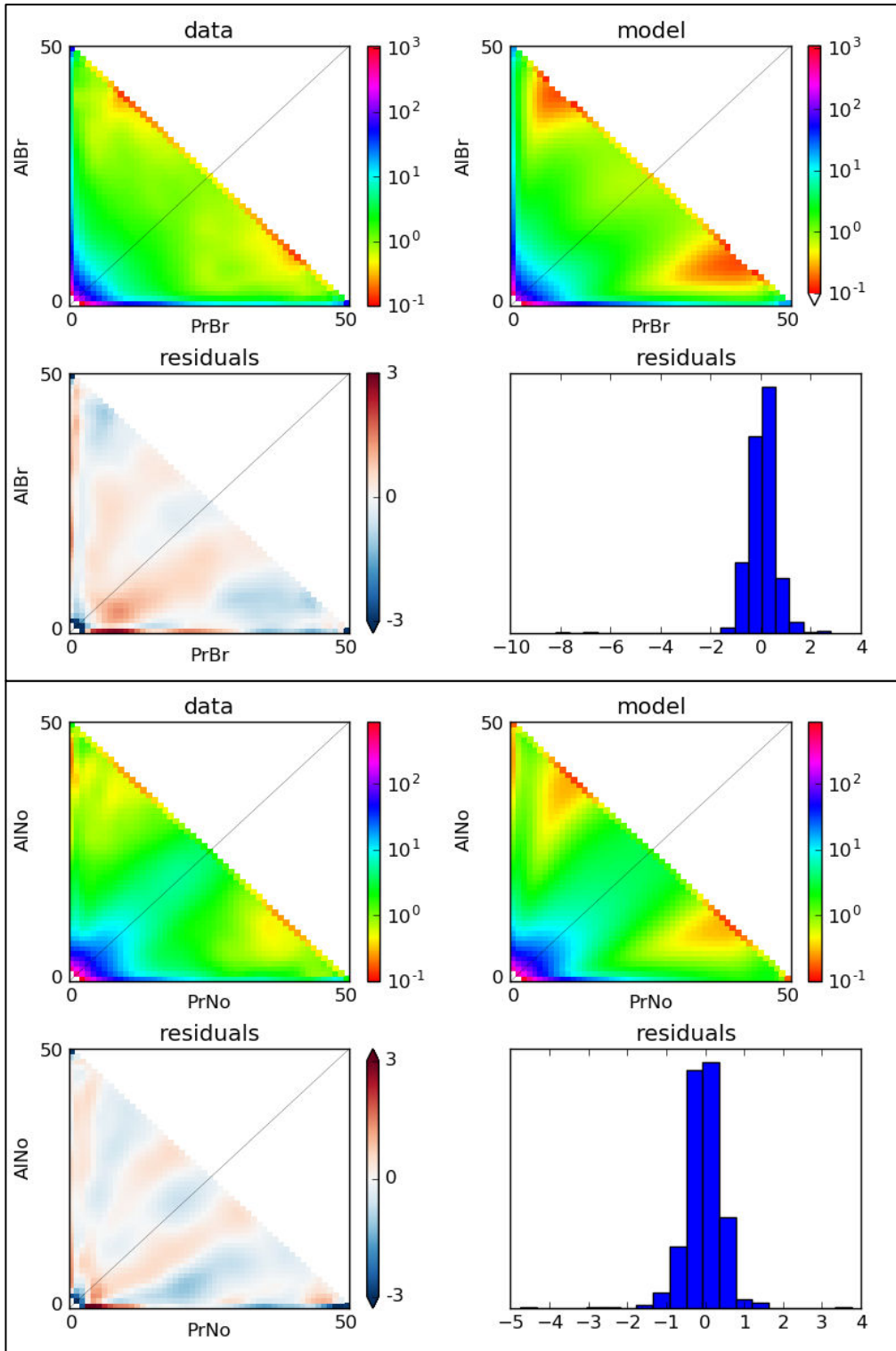


Figure 2. Comparaison des données observées (data) et du meilleur modèle obtenu (model), *i.e.* SC2m en a) Bretagne et en b) Normandie. Les résidus (residuals) correspondent à la différence entre les valeurs observées et les valeurs modélisées.

migration efficace neutre de *J. praeirsuta* dans *J. albifrons* ($\frac{m_{12}}{m_{21}} = \frac{M_{12}}{2N_{ref}} \times \frac{2N_{ref}}{M_{21}} = \frac{M_{12}}{M_{21}} = 1.9-2.8$ en Bretagne et 1.7-1.9 en Normandie). Le taux de migration efficace réduit s'applique à une proportion $1-P$ de loci dont la migration interspécifique est réduite par rapport au reste du génome, et semble légèrement plus important de *J. albifrons* vers *J. praeirsuta* en Normandie ($\frac{m'_{21}}{m'_{12}} = \frac{M'_{21}}{2N_{ref}} \times \frac{2N_{ref}}{M'_{12}} = \frac{M'_{21}}{M'_{12}} = 1.4$). En Bretagne en revanche, ce taux de migration efficace réduit est plus variable, avec un facteur de *J. albifrons* vers *J. praeirsuta* de 0.8 obtenu par le spectre plié et de 1.2 obtenu par le spectre déplié.

L'introgession différentielle à ces loci est ralentie d'un facteur moyen $\frac{1}{2} \left(\frac{m_{12}}{m'_{12}} + \frac{m_{21}}{m'_{21}} \right) = \frac{1}{2} \left(\frac{M_{12}}{M'_{12}} + \frac{M_{21}}{M'_{21}} \right) = 22-24$ en Bretagne et 11-18 en Normandie. Cependant, ce "ralentissement" semble plus fort dans le sens *J. praeirsuta* vers *J. albifrons* ($\frac{m_{12}}{m'_{12}} = \frac{M_{12}}{2N_{ref}} \cdot \frac{2N_{ref}}{M'_{12}} = \frac{M_{12}}{M'_{12}} = 32-33$ en Bretagne et 17-26 en Normandie) que dans le sens opposé ($\frac{m_{21}}{m'_{21}} = 13-15$ en Bretagne et 7-15 en Normandie) (Figure 3).

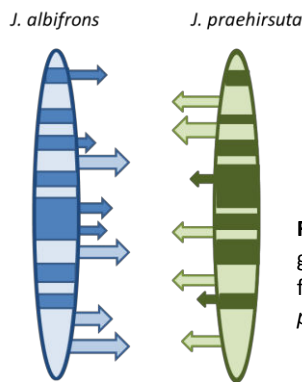


Figure 3. Représentation schématique de l'asymétrie de migration le long du génome : exemple de la Bretagne, où 62% du génome, représenté par les zones foncées, va migrer 15 fois moins que le reste du génome de *J. albifrons* vers *J. praeirsuta*, et 32 fois moins de *J. praeirsuta* vers *J. albifrons*.

Le flux de gènes efficace moyen (produit Nm) de *J. praeirsuta* vers *J. albifrons* est estimé en Bretagne à 1.63 migrants/génération avec *outgroup* et 1.88 migrants/génération sans *outgroup* ; et en Normandie à 1.24 migrants/génération avec *outgroup*, et 1.28 migrants/génération sans *outgroup*. Le flux de gènes efficace moyen de *J. albifrons* vers *J. praeirsuta* est quant à lui estimé en Bretagne à 1.03 migrants/génération avec *outgroup* et 4.27 migrants/génération sans *outgroup* et en Normandie il est estimé à 3.17 migrants/génération avec *outgroup* et 9.85 migrants/génération sans *outgroup* en Normandie.

Concernant les tailles de populations, on retrouve dans tous les cas que *J. praeirsuta* (n_2) montre la plus grande taille efficace de population, et que cette taille efficace est 1,5 à 2 fois plus grande que celle de *J. albifrons* (n_1) en Bretagne alors qu'elle est plus de 4x plus grande en Normandie.

Une forte similarité dans les résultats des modèles pliés et dépliés est observable. En effet, un contact secondaire a dans tous les cas été détecté, avec des taux de migration hétérogènes et asymétriques similaires. La méthode sans *outgroup* estime le même flux génique que la méthode avec *outgroup* dans le sens *J. praeirsuta* vers *J. albifrons*, en revanche elle a tendance à surestimer le nombre de migrants par génération de

J. albifrons vers *J. prae-hirsuta* en comparaison avec la seconde méthode. Enfin une même estimation de la proportion du génome concerné par ces taux de migration neutre et réduit est donnée par les deux méthodes. Cette similarité dans les résultats à travers les méthodes utilisées soutient d'autant plus le modèle de contact secondaire comme étant le modèle parmi ceux testés expliquant le mieux l'histoire démographique de *J. albifrons* et *J. prae-hirsuta* en Bretagne et en Normandie.

Tableau 1. Description des valeurs de paramètres obtenues pour le meilleur modèle déplié ou plié dans chaque région

	AIC	θ	n1	n2	hrf	M12	M21	M'12	M'21	Tdiv	Tpostdiv	P	Q	O
Spectres dépliés														
Bretagne														
SC2N2m	569.182	16.229	5.509	6.248	0.133	2.653	1.392	0.079	0.094	4.282	0.870	0.376	0.499	0.972
Normandie														
SC2m	684.505	103.642	0.363	1.551	-	9.930	5.811	0.387	0.536	0.784	0.094	0.674	-	0.963
Spectres repliés														
Bretagne														
SC2m	3879.504	597.174	1.915	3.537	-	3.901	1.370	0.122	0.103	2.720	0.217	0.494	-	-
Normandie														
SC2m	4041.974	1526.717	0.411	1.750	-	29.717	15.442	1.732	2.374	0.522	0.036	0.678	-	-

*Le tableau contient dans l'ordre, la valeur de l'AIC associé au modèle, l'estimation de la diversité génétique Theta de la population ancestrale, la taille efficace pondérée de la population *J. albifrons* (N1) et *J. praehirsuta* (N2), la valeur du facteur Hill-Robertsonien (hrf) correspondant à l'estimation de la réduction de la taille efficace dans des zones du génome étant soumises aux effets du balayage sélectif, les taux de migration neutre pondérés de *J. praehirsuta* dans *J. albifrons* (M12) et inversement (M21) ainsi que les taux de migration réduit pondérés correspondant aux régions génomiques sous sélection de *J. praehirsuta* dans *J. albifrons* (M'12) et inversement (M'21), l'estimation de la durée depuis la divergence des deux espèces (Tdiv) et le temps depuis que la migration est arrêtée dans le cas de la migration ancestrale ou a recommencé après un contact secondaire (Tpostdiv). Enfin, la proportion du génome avec une taille effective neutre (P), la proportion du génome avec une taille effective réduite due à une sélection aux sites liés (Q) et la bonne orientation des SNPS (ancestraux ou dérivés; O) sont également donnés.

REFERENCES

A

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Most, M., Mullen, S., Nichols, R., Nolte, A. W., Parisod, C., Pfennig, K., Rice, A. M., Ritchie, M. G., Seifert, B., Smadja, C. M., Stelkens, R., Szymura, J. M., Vainola, R., Wolf, J. B. & Zinner, D. 2013. Hybridization and speciation. *J Evol Biol* **26**: 229-46.
- Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. available at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- April, J., Hanner, R. H., Dion-Cote, A. M. & Bernatchez, L. 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Mol Ecol* **22**: 409-22.
- Archer, F. I., Adams, P. E. & Schneiders, B. B. 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. *Molecular ecology resources* **17**: 5-11.
- Avtzis, D. N., Doudoumis, V. & Bourtzis, K. 2014. Wolbachia Infections and Mitochondrial Diversity of Two Chestnut Feeding Cydia Species.
- Baldo, L. & Werren, J. H. 2007. Revisiting Wolbachia supergroup typing based on WSP: spurious lineages and discordance with MLST. *Current microbiology* **55**: 81-87.
- Bandi, C., Anderson, T. J., Genchi, C. & Blaxter, M. L. 1998. Phylogeny of Wolbachia in filarial nematodes. *Proceedings of the Royal Society of London B: Biological Sciences* **265**: 2407-2413.
- Bank, C., Bürger, R. & Hermisson, J. 2012. The limits to parapatric speciation: Dobzhansky–Muller incompatibilities in a continent–island model. *Genetics* **191**: 845-863.
- Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M. & Meyer, A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* **439**: 719.
- Barton, N. & Bengtsson, B. O. 1986. The barrier to genetic exchange between hybridising populations. *Heredity* **57**: 357-376.
- Barton, N. H. & de Cara, M. A. 2009. The evolution of strong reproductive isolation. *Evolution* **63**: 1171-90.
- Barton, N. H. & Hewitt, G. M. 1985. Analysis of hybrid zones. *Annual Review of Ecology, Evolution, and Systematics* **16**: 113-148.
- Beaumont, M. A. & Nichols, R. A. 1996. Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society B: Biological Sciences* **263**: 1619-1626.
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R. & Bernatchez, L. 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Mol Ecol* **24**: 3299-315.
- Bernatchez, L. & Wilson, C. C. 1998. Comparative phylogeography of

B

- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. & Johnson, E. A. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One* **3**.

- Nearctic and Palearctic fishes. *Molecular Ecology* **7**: 431-452.
- Beukeboom, L. W. & Perrin, N. 2014. *The evolution of sex determination*. Oxford University Press, USA.
- Bierne, N., Bonhomme, F., Boudry, P., Szulkin, M. & David, P. 2006. Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proceedings of the Royal Society of London B* **273**: 1253-60.
- Bierne, N., Bonhomme, F. & David, P. 2003a. Habitat preference and the marine-speciation paradox. *Proceedings of the Royal Society of London B* **270**: 1399-406.
- Bierne, N., Borsa, P., Daguin, C., Jollivet, D., Viard, F., Bonhomme, F. & David, P. 2003b. Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology* **12**: 447-461.
- Bierne, N., David, P., Boudry, P. & Bonhomme, F. 2002. Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Evolution* **56**: 292-298.
- Bierne, N., Gagnaire, P. A. & David, P. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology* **59**: 72-86.
- Bierne, N., Welch, J., Loire, E., Bonhomme, F. & David, P. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol Ecol* **20**: 2044-72.
- Bocquet, C. (1953) Recherches sur le polymorphisme naturel des *Jaera Marina* (Fabr.) (Isopodes Asellotes).: Essai de systématique évolutive. pp. Centre national de la recherche scientifique.
- Bocquet, C. 1954. Evolution of a Superspecies of Marine Isopods. *Systematic Zoology* **3**: 148-162.
- Bocquet, C. 1972. La spéciation des *Jaera albifrons*. Etat présent des connaissances et des problèmes. *Fifth European Marine Biology Symposium*: 131-137.
- Bocquet, C. & Solignac, M. 1969. Étude morphologique des hybrides expérimentaux entre *Jaera (albifrons) albifrons* et *Jaera (albifrons) praehirsuta*. *Arch Zool Exp Gén* **110**: 435-452.
- Bolger, A. M., Lohse, M. & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.
- Bonnet, T., Leblois, R., Rousset, F. & Crochet, P. A. 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*.
- Bordenstein, S. R., O'Hara, F. P. & Werren, J. H. 2001. Wolbachia-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature* **409**: 707-710.
- Bordenstein, S. R. & Werren, J. H. 1998. Effects of A and B Wolbachia and host genotype on interspecies cytoplasmic incompatibility in *Nasonia*. *Genetics* **148**: 1833-1844.
- Bordenstein, S. R. & Werren, J. H. 2007. Bidirectional incompatibility among divergent Wolbachia and incompatibility level differences among closely related Wolbachia in *Nasonia*. *Heredity (Edinb)* **99**: 278-87.
- Bouchon, D., Cordaux, R. & Grève, P. 2008. Feminizing Wolbachia and the evolution of sex determination in isopods. *Insect symbiosis* **3**: 273-294.
- Bouchon, D., Rigaud, T. & Juchault, P. 1998. Evidence for widespread Wolbachia infection in isopod crustaceans: molecular identification and host feminization. *Proc Biol Sci* **265**: 1081-90.
- Braig, H. R., Zhou, W., Dobson, S. L. & O'Neill, S. L. 1998. Cloning and Characterization of a Gene Encoding

- the Major Surface Protein of the Bacterial Endosymbiont *Wolbachia pipientis*. *Journal of bacteriology* **180**: 2373-2378.
- Breeuwer, J. & Jacobs, G. 1996. *Wolbachia*: intracellular manipulators of mite reproduction. *Experimental and Applied Acarology* **20**: 421-434.
- Brelsford, A., Dufresnes, C. & Perrin, N. 2016. High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity (Edinb)* **116**: 177-81.
- Briggs, J. C. 1999. Modes of speciation: marine Indo-west Pacific. *Bulletin of Marine Science* **65**: 645-656.
- Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889-890.
- Buckler, E. S., Phelps-Durr, T. L., Buckler, C. S. K., Dawe, R. K., Doebley, J. F. & Holtsford, T. P. 1999. Meiotic drive of chromosomal knobs reshaped the maize genome. *Genetics* **153**: 415-426.
- Buggs, R. J. 2007. Empirical study of hybrid zone movement. *Heredity (Edinb)* **99**: 301-12.
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S. & Garamszegi, L. Z. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome research* **25**: 1656-1665.
- Butlin, R. K., Galindo, J. & Grahame, J. W. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **363**: 2997-3007.
- Butlin, R. K., Saura, M., Charrier, G., Jackson, B., Andre, C., Caballero, A., Coyne, J. A., Galindo, J., Grahame, J. W., Hollander, J., Kemppainen, P., Martinez-Fernandez, M., Panova, M., Quesada, H., Johannesson, K. & Rolan-Alvarez, E. 2014. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* **68**: 935-49.
- C**
- Cariou, M.-L. (1977) Recherches sur le polymorphisme enzymatique du complexe *Jaera albifrons*, Leach (Crustace, Isopode). pp.
- Carvalho, G. 1989a. Genetic and ecological differentiation in the *Jaera albifrons* species complex. *Report prepared for the Natural Environmental Research Council, UK*: 82.
- Carvalho, G. 1989b. Microgeographic genetic differentiation and dispersal capacity in the intertidal isopod, *Jaera albifrons* Leach.
- Carvalho, G. R. & Piertney, S. B. 1997. Interspecific comparisons of genetic population structure in members of the *Jaera albifrons* species complex. *The Journal of the Marine Biological Association of the United Kingdom* **77**: 77-93.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol* **22**: 3124-40.
- Chang, A. S. & Noor, M. A. 2007. The genetics of hybrid male sterility between the allopatric species pair *Drosophila persimilis* and *D. pseudoobscura bogotana*: dominant sterility alleles in collinear autosomal regions. *Genetics* **176**: 343-349.
- Charlesworth, B., Nordborg, M. & Charlesworth, D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided

- populations. *Genetics Research* **70**: 155-174.
- Clarkson, C. S., Weetman, D., Essandoh, J., Yawson, A. E., Maslen, G., Manske, M., Field, S. G., Webster, M., Antao, T., MacInnis, B., Kwiatkowski, D. & Donnelly, M. J. 2014. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nat Commun* **5**: 4248.
- Cooper, B. S., Ginsberg, P. S., Turelli, M. & Matute, D. R. 2017. *Wolbachia* in the *Drosophila yakuba* complex: pervasive frequency variation and weak cytoplasmic incompatibility, but no apparent effect on reproductive isolation. *Genetics* **205**: 333-351.
- Cordaux, R., Michel-Salzat, A. & Bouchon, D. 2001. *Wolbachia* infection in crustaceans: novel hosts and potential routes for horizontal transmission. *Journal of Evolutionary Biology* **14**: 237-243.
- Cordaux, R., Pichon, S., Hatira, H. B., Doublet, V., Greve, P., Marcade, I., Braquart-Varnier, C., Souty-Grosset, C., Charfi-Cheikhrouha, F. & Bouchon, D. 2012. Widespread *Wolbachia* infection in terrestrial isopods and other crustaceans. *Zookeys*: 123-31.
- Core Team 2016. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.
- Coyne, J. A. & Orr, H. A. 2004. *Speciation*. Sinauer Associates Sunderland, MA.
- Cruickshank, T. E. & Hahn, M. W. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* **23**: 3133-57.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. & Genomes Project Analysis, G. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156-8.
- Deagle, B. E., Jones, F. C., Chan, Y. F., Absher, D. M., Kingsley, D. M. & Reimchen, T. E. 2012. Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc Biol Sci* **279**: 1277-86.
- Dobzhansky, T. 1937. *Genetics and Origin of Species* **Columbia: Columbia University Press**.
- Dobzhansky, T. 1970. *Genetics of the evolutionary process*. Columbia University Press.
- Dunn, L. & Bennett, D. 1967. Sex differences in recombination of linked genes in animals. *Genetics Research* **9**: 211-220.
- Dupuis, J. R., Brunet, B. M. T., Bird, H. M., Lumley, L. M., Fagua, G., Boyle, B., Levesque, R., Cusson, M., Powell, J. A. & Sperling, F. A. H. 2017. Genome-wide SNPs resolve phylogenetic relationships in the North American spruce budworm (*Choristoneura fumiferana*) species complex. *Mol Phylogenet Evol* **111**: 158-168.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K. & Qvarnström, A. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**: 756.

D

E

- Endler, J. A. 1977. *Geographic variation, speciation, and clines*. Princeton University Press.
- Ewels, P., Magnusson, M., Lundin, S. & Källér, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047-3048.
- Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C. & Foll, M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet* **9**: e1003905.
- Excoffier, L. & Ray, N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol* **23**: 347-51.
- F**
- Faria, R. & Navarro, A. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol* **25**: 660-9.
- Feder, J. L., Egan, S. P. & Nosil, P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet* **28**: 342-50.
- Feder, J. L. & Nosil, P. 2009. Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* **63**: 3061-75.
- Feder, J. L. & Nosil, P. 2010. The Efficacy of Divergence Hitchhiking in Generating Genomic Islands during Ecological Speciation. *Evolution* **64**: 1729-1747.
- Feder, J. L., Nosil, P. & Flaxman, S. M. 2014. Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. *Frontiers in genetics* **5**.
- Feder, J. L., Xie, X., Rull, J., Velez, S., Forbes, A., Leung, B., Dambroski, H., Filchak, K. E. & Aluja, M. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proceedings of the National Academy of Sciences* **102**: 6573-6580.
- Ferveur, J.-F. 2005. Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. *Behavior genetics* **35**: 279.
- Fierst, J. L. & Hansen, T. F. 2010. Genetic architecture and postzygotic reproductive isolation: evolution of Bateson–Dobzhansky–Muller incompatibilities in a polygenic model. *Evolution* **64**: 675-693.
- Fisher, R. A. 1999. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.
- Fitzpatrick, B. M., Fordyce, J. A. & Gavrilets, S. 2008. What, if anything, is sympatric speciation? *J Evol Biol* **21**: 1452-9.
- Fobes, A. 1974. Osmotic and sodium regulation in *Jaera albifrons* Leach (Crustacea: Isopoda). *Comparative Biochemistry and Physiology* **47**: 109-116.
- G**
- Gagnaire, P.-A., Pavey, S. A., Normandeau, E. & Bernatchez, L. 2013. The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD-sequencing. *Evolution* **67**: 2483–2497.
- Gavrilets, S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* **403**: 886-889.
- Giraud, T. 2006. Speciation in parasites: host switching does not automatically lead to allopatry. *Trends in parasitology* **22**: 151-152.
- Gompert, Z., Lucas, L. K., Fordyce, J. A., Forister, M. L. & Nice, C. C. 2010. Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive

- admixture and a patchy hybrid zone. *Mol Ecol* **19**: 3171-92.
- Gompert, Z., Mandeville, E. G. & Buerkle, C. A. 2017. Analysis of Population Genomic Data from Hybrid Zones. *Annual Review of Ecology, Evolution, and Systematics*.
- Gonen, S., Lowe, N. R., Cezard, T., Gharbi, K., Bishop, S. C. & Houston, R. D. 2014. Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *Bmc Genomics* **15**: 166.
- Gow, J. L., Peichel, C. L. & Taylor, E. B. 2007. Ecological selection against hybrids in natural populations of sympatric threespine sticklebacks. *Journal of evolutionary biology* **20**: 2173-2180.
- Gregorio, O., Berdan, E. L., Kozak, G. M. & Fuller, R. C. 2012. Reinforcement of male mate preferences in sympatric killifish species *Lucania goodei* and *Lucania parva*. *Behavioral Ecology and Sociobiology* **66**: 1429-1436.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5**: e1000695.
- Haahtela, I. 1965. Morphology, habitats and distribution of species of the *Jaera albifrons* group (Isopoda, Janiridae) in Finland. *Annales Zoologici Fennici*: 309-314.
- Haldane, J. B. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of genetics* **12**: 101-109.
- Hardy, I. C. (2002) Sex ratios. In: *Concepts and research methods*. pp. Cambridge University Press, UK.
- Harr, B. 2006. Genomic islands of differentiation between house mouse subspecies. *Genome Research* **16**: 730-737.
- Harrison, R. G. & Larson, E. L. 2014. Hybridization, introgression, and the nature of species boundaries. *J Hered* **105 Suppl 1**: 795-809.
- Harrison, R. G. & Larson, E. L. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Mol Ecol*.
- Harrison, R. G. & Rand, D. M. 1989. Mosaic hybrid zone and the nature of species boundaries. In: *Speciation and its Consequences (edited by Otte D, Endler J)*: 111-133.
- Harvey, C. E. & Naylor, E. 1968. The taxonomic status of *Jaera albifrons* *syei* Bocquet (Isopoda, Asellota). *Crustaceana* **14**: 185-193.
- Hauffe, H. C. & Searle, J. B. 1993. Extreme Karyotypic Variation in a *Mus musculus domesticus* Hybrid Zone: The Tobacco Mouse Story Revisited. *Evolution* **47**: 1374-1395.
- Hedrick, P. W. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol Ecol* **22**: 4606-18.
- Hegde, S. G., Nason, J. D., Clegg, J. M. & Ellstrand, N. C. 2006. The evolution of California's wild radish has resulted in the extinction of its progenitors. *Evolution* **60**: 1187-1197.
- Herrera, S. & Shank, T. M. 2016. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Molecular phylogenetics and evolution* **100**: 70-79.
- Hertig, M. & Wolbach, S. B. 1924. Studies on rickettsia-like micro-organisms in insects. *The Journal of medical research* **44**: 329.
- Hewitt, G. M. 2001. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Molecular ecology* **10**: 537-549.

H

- Hewitt, G. M. 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**: 617-38.
- Hudson, R. R. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337-338.
- Humphreys, D. P., Rundle, H. D. & Dyer, K. A. 2016. Patterns of reproductive isolation in the *Drosophila* subquinarium complex: can reinforced premating isolation cascade to other species? *Current Zoology* **62**: 183-191.
- Hurst, G. D., Jiggins, F. M., von der Schulenburg, J. H. G., Bertrand, D., West, S. A., Goriacheva, I. I., Zakharov, I. A., Werren, J. H., Stouthamer, R. & Majerus, M. E. 1999. Male-killing *Wolbachia* in two species of insect. *Proceedings of the Royal Society of London B: Biological Sciences* **266**: 735-740.
- Iyengar, V. K., Reeve, H. K. & Eisner, T. 2002. Paternal inheritance of a female moth's mating preference. *Nature* **419**: 830.
- Jaenike, J. 2001. Sex chromosome meiotic drive. *Annual Review of Ecology and Systematics* **32**: 25-49.
- Jaenike, J., Dyer, K. A., Cornish, C. & Minhas, M. S. 2006. Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. *PLoS Biol* **4**: e325.
- Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskočilová, M. M., Nachman, M. W., Munclinger, P., Macholán, M., Piálek, J. & Tucker, P. K. 2012. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology* **21**: 3032-3047.
- Jeyaprakash, A. & Hoy, M. A. 2000. Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect molecular biology* **9**: 393-405.
- Johannesson, K. 2001. Parallel speciation: a key to sympatric divergence. *Trends in Ecology & Evolution* **16**: 148-153.
- Johnston, S. E., Huisman, J., Ellis, P. A. & Pemberton, J. M. 2017. A High-Density Linkage Map Reveals Sexual Dimorphism in Recombination Landscapes in Red Deer (*Cervus elaphus*). *G3: Genes, Genomes, Genetics* **7**: 2859-2870.
- Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403-1405.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., Brady, S. D., Zhang, H., Pollen, A. A., Howes, T., Amemiya, C., Broad Institute Genome Sequencing, P., Whole Genome Assembly, T., Baldwin, J., Bloom, T., Jaffe, D. B., Nicol, R., Wilkinson, J., Lander, E. S., Di Palma, F., Lindblad-Toh, K. & Kingsley, D. M. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55-61.
- Jones, M. 1972a. Effects of salinity on the survival of the *Jaera albifrons* Leach group of species (Crustacea: Isopoda). *Journal of Experimental Marine Biology and Ecology* **9**: 231-237.
- Jones, M. 1972b. The mouthparts of the members of the *Jaera albifrons* group of species (Crustacea: Isopoda). *Marine Biology* **14**: 264-270.

- Jones, M. & Naylor, E. 1971a. Breeding and bionomics of the British members of the *Jaera albifrons* group of species (Isopoda: Asellota). *J Zool* **165**: 183-199.
- Jones, M. B. & Fordy, M. R. 1971. A Stereoscan electron microscope study of male reproductive characters in the *Jaera albifrons* group of species (Crustacea : Isopoda). *Marine Biology* **10**: 265-271.
- Jones, M. B. & Naylor, E. 1971b. Breeding and bionomics of the British members of the *Jaera albifrons* group of species (Isopoda: Asellota). *Journal of Zoology* **165**: 183-199.
- Jónsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M., Albrechtsen, A., Petersen, B., Korneliussen, T. S. & Vilstrup, J. T. 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences* **111**: 18655-18660.
- K**
- Kakioka, R., Kokita, T., Kumada, H., Watanabe, K. & Okuda, N. 2013. A RAD-based linkage map and comparative genomics in the gudgeons (genus *Gnathopogon*, Cyprinidae). *BMC genomics* **14**: 32.
- Khaitov, V. M., Kuzmin, A. A. & Terovskaya, E. V. 2006. Morphological differences between females of different *Jaera* species (Crustacea: Asellota: Isopoda) in the White Sea: a possible solution to an old problem. *Marine Biology* **150**: 1205-1214.
- King, M. 1993. Species Evolution. The role of chromosome change. *Cambridge University Press, Cambridge, MA*.
- Kirkpatrick, M. & Servedio, M. R. 1999. The reinforcement of mating preferences on an island. *Genetics* **151**: 865-884.
- Kitano, J., Ross, J. A., Mori, S., Kume, M., Jones, F. C., Chan, Y. F., Absher, D. M., Grimwood, J., Schmutz, J. & Myers, R. M. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**: 1079.
- Klibansky, L. K. & McCartney, M. A. 2014. Conspecific sperm precedence is a reproductive barrier between free-spawning marine mussels in the northwest atlantic mytilus hybrid zone. *PLoS One* **9**: e108433.
- Knaus, B. J. & Grünwald, N. J. 2017. VCFR: a package to manipulate and visualize variant call format data in R. *Molecular ecology resources* **17**: 44-53.
- Kondrashov, A. S. & Mina, M. V. 1986. Sympatric speciation: when is it possible? *Biological Journal of the Linnean society* **27**: 201-223.
- Kulathinal, R. J., Stevison, L. S. & Noor, M. A. 2009. The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS genetics* **5**: e1000550.
- L**
- Larson, E. L., Hume, G. L., Andres, J. A. & Harrison, R. G. 2012. Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. *J Evol Biol* **25**: 174-86.
- Le Moan, A., Gagnaire, P. A. & Bonhomme, F. 2016. Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Mol Ecol* **25**: 3187-202.
- Lécher, P. 1968. Polymorphisme chromosomique dans les populations baltes et scandinaves de l'isopode *Jaera* (albifrons) syei Bocquet. *Archives de Zoologie*

- Exeprimentale et Generale* **109**: 211-227.
- Lécher, P. & Prunus, G. 1971. Caryologie et taxinomie de *Jaera (albifrons) albifrons* (Crustacé, Isopode), Populations des Côtes Bretonnes. *Archives of Zoological Experimental Genetics* **112**: 715-730.
- Lemmon, A. R. & Kirkpatrick, M. 2006. Reinforcement and the genetics of hybrid incompatibilities. *Genetics* **173**: 1145-1155.
- Lepais, O., Petit, R., Guichoux, E., Lavabre, J., Alberto, F., Kremer, A. & Gerber, S. 2009. Species relative abundance and direction of introgression in oaks. *Molecular Ecology* **18**: 2228-2242.
- 
- Maheshwari, S. & Barbash, D. A. 2011. The genetics of hybrid incompatibilities. *Annual review of genetics* **45**: 331-355.
- Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**: 229-237.
- Mallet, J. 2007. Hybrid speciation. *Nature* **446**: 279.
- Mandeville, E. G., Parchman, T. L., McDonald, D. B. & Buerkle, C. A. 2015. Highly variable reproductive isolation among pairs of *Catostomus* species. *Molecular ecology* **24**: 1856-1872.
- Mank, J. E., Axelsson, E. & Ellegren, H. 2007. Fast-X on the Z: rapid evolution of sex-linked genes in birds. *Genome Research* **17**: 618-624.
- Manthey, J. D., Geiger, M. & Moyle, R. G. 2017. Relationships of morphological groups in the northern flicker superspecies complex (*Colaptes auratus* & *C. chrysoides*). *Systematics and Biodiversity* **15**: 183-191.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**: pp. 10-12.
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J. & Jiggins, C. D. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res* **23**: 1817-28.
- Matute, D. R. 2010. Reinforcement can overcome gene flow during speciation in *Drosophila*. *Current Biology* **20**: 2229-2233.
- Mayr, E. 1963. Animal species and evolution. *Harvard University Press, Cambridge, Massachusetts*.
- Megléczy, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N. & Martin, J. F. 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**: 403-404.
- Meier, J. I., Sousa, V. C., Marques, D. A., Selz, O. M., Wagner, C. E., Excoffier, L. & Seehausen, O. 2017. Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol Ecol* **26**: 123-141.
- Merçot, H. & Charlat, S. 2004. Wolbachia infections in *Drosophila melanogaster* and *D. simulans*: polymorphism and levels of cytoplasmic incompatibility. *Genetica* **120**: 51-59.
- Metzker, M. L. 2010. Sequencing technologies--the next generation. *Nature reviews. Genetics* **11**: 31.
- Mifsud, D. V. 2011. The genetic basis of speciation in the *Jaera albifrons* species group of intertidal isopods. *PhD Thesis (University of Aberdeen, Scotland)*.
- Monteiro, C. A., Serrão, E. A. & Pearson, G. A. 2012. Prezygotic barriers to hybridization in marine broadcast spawners: reproductive timing and

mating system variation. *PLoS One* **7**: e35978.



Nadachowska-Brzyska, K., Burri, R., Olason, P. I., Kawakami, T., Smeds, L. & Ellegren, H. 2013. Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. *PLoS Genet* **9**: e1003942.

Navarro, A. & Barton, N. H. 2003. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* **57**: 447-459.

Naylor, E. 1972. *British marine isopods*. Academic Press for the Linnean Society of London.

Naylor, E. & Haahtela, I. 1966. Habitat Preferences and Interspersion of Species within the Superspecies *Jaera albifrons* Leach (Crustacea: Isopoda). *Journal of Animal Ecology* **35**: 209-216.

Nolte, A. W., Gompert, Z. & Buerkle, C. A. 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Mol Ecol* **18**: 2615-27.

Noor, M. A. & Bennett, S. M. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity (Edinb)* **103**: 439-44.

Noor, M. A., Grams, K. L., Bertucci, L. A. & Reiland, J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci U S A* **98**: 12084-8.

Nosil, P., Funk, D. J. & ORTIZ-BARRIENTOS, D. 2009. Divergent selection and heterogeneous genomic divergence. *Molecular ecology* **18**: 375-402.



Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**: 1805-1813.

Orr, H. A. & Turelli, M. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* **55**: 1085-1094.

Ortiz-Barrientos, D. & James, M. E. 2017. Evolution of recombination rates and the genomic landscape of speciation. *J Evol Biol* **30**: 1519-1521.



Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O. & Jiggins, C. D. 2012. Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics* **8**: e1002752.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS one* **7**: e37135.

Phadnis, N. & Orr, H. A. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *science* **323**: 376-379.

Pickrell, J. K. & Pritchard, J. K. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics* **8**: e1002967.

Piertney, S. B. & Carvalho, G. R. 1996. Sex ratio variation in the intertidal isopod, *Jaera albifrons*. *Journal of the Marine Biological Association of the United Kingdom* **76**: 825-828.

Prunus, G. 1968. Etude systématique des populations chez l'Isopode *Jaera* (*albifrons*) *albifrons* Forsman.

Archives de zoologie expérimentale et générale **109**: 643-702.



- Rasanen, K. & Hendry, A. P. 2014. Asymmetric reproductive barriers and mosaic reproductive isolation: insights from Misty lake-stream stickleback. *Ecol Evol* **4**: 1166-75.
- Rastas, P., Paulin, L., Hanski, I., Lehtonen, R. & Auvinen, P. 2013. Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics* **29**: 3128-34.
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlovic, M., Noor, M. A. F., Mehlig, B. & Westram, A. M. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol* **30**: 1450-1477.
- Ravinet, M., Westram, A., Johannesson, K., Butlin, R., Andre, C. & Panova, M. 2016. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol Ecol* **25**: 287-305.
- Reitzel, A., Herrera, S., Layden, M., Martindale, M. & Shank, T. 2013. Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular ecology* **22**: 2953-2970.
- Ribardière, A., Daguin-Thiébaud, C., Houbin, C., Coudret, J., Broudin, C., Timsit, O. & Broquet, T. 2017. Geographically distinct patterns of reproductive isolation and hybridization in two sympatric species of the *Jaera albifrons* complex (marine isopods). *Ecology and Evolution*.
- Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. *Trends in ecology & evolution* **16**: 351-358.
- Rigaud, T. & Juchault, P. 1995. Success and failure of horizontal transfers of feminizing *Wolbachia* endosymbionts in woodlice. *Journal of Evolutionary Biology* **8**: 249-255.
- Rigaud, T., Moreau, J. & Juchault, P. 1999. *Wolbachia* infection in the terrestrial isopod *Oniscus asellus*: sex ratio distortion and effect on fecundity. *Heredity* **83**: 469-475.
- Riquet, F., Liautard-Haag, C., Woodall, L., Bouza, C., Louisy, P., Hamer, B., Otero-Ferrer, F., Aublanc, P., Beduneau, V. & Briard, O. 2017. Parallel use of a shared genomic island of speciation in clinal and mosaic hybrid zones between cryptic seahorse lineages. *bioRxiv*: 161786.
- Ritter, S., Michalski, S. G., Settele, J., Wiemers, M., Fric, Z. F., Sielezniew, M., Sasic, M., Rozier, Y. & Durka, W. 2013. *Wolbachia* infections mimic cryptic speciation in two parasitic butterfly species, *Phengaris teleius* and *P. nausithous* (Lepidoptera: Lycaenidae). *PLoS One* **8**: e78107.
- Robinson, T., Johnson, N. A. & Wade, M. J. 1994. Postcopulatory, prezygotic isolation: intraspecific and interspecific sperm precedence in *Tribolium* spp., flour beetles. *Heredity* **73**: 155-159.
- Roesti, M., Hendry, A. P., Salzburger, W. & Berner, D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology* **21**: 2852-2862.
- Rolán-Alvarez, E., Austin, C. & Boulding, E. 2015. The Contribution of the Genus *Littorina* to the Field of Evolutionary Ecology. 157-214.
- Rougemont, Q., Gagnaire, P. A., Perrier, C., Genthon, C., Besnard, A. L., Launey, S. & Evanno, G. 2016a. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol Ecol*.

- Rougemont, Q., Gaigher, A., Lasne, E., Cote, J., Coke, M., Besnard, A. L., Launey, S. & Evanno, G. 2015. Low reproductive isolation and highly variable levels of gene flow reveal limited progress towards speciation between European river and brook lampreys. *J Evol Biol* **28**: 2248-63.
- Rougemont, Q., Roux, C., Neuenschwander, S., Goudet, J., Launey, S. & Evanno, G. 2016b. Reconstructing the demographic history of divergence between European river and brook lampreys using approximate Bayesian computations. *PeerJ* **4**: e1910.
- Rougeux, C., Bernatchez, L. & Gagnaire, P.-A. 2016. Modeling the multiple facets of speciation-with-gene-flow towards improving divergence history inference of a recent fish adaptive radiation. *bioRxiv*: 068932.
- Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources* **8**: 103-106.
- Rousset, F., Bouchon, D., Pintureau, B., Juchault, P. & Solignac, M. 1992. Wolbachiae endosymbionts responsible for various alterations of sexuality in arthropods. *Proceedings of the Royal Society of London B* **250**: 91-98.
- Roux, C., Fraise, C., Castric, V., Vekemans, X., Pogson, G. H. & Bierne, N. 2014. Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *J Evol Biol* **27**: 1662-75.
- Roux, C., Fraise, C., Romiguier, J., Anciaux, Y., Galtier, N. & Bierne, N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS biology* **14**: e2000234.
- Roux, C. & Pannell, J. R. 2015. Inferring the mode of origin of polyploid species from next-generation sequence data. *Mol Ecol* **24**: 1047-59.
- Ruegg, K., Anderson, E. C., Boone, J., Pouls, J. & Smith, T. B. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**: 4757-69.




- Sæther, S. A., Sætre, G.-P., Borge, T., Wiley, C., Svedin, N., Andersson, G., Veen, T., Haavie, J., Servedio, M. R. & Bureš, S. 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *science* **318**: 95-97.
- Saetre, G. P., Kral, K., Bures, S. & Ims, R. A. 1999. Dynamics of a clinal hybrid zone and a comparison with island hybrid zones of flycatchers (*Ficedula hypoleuca* and *F. albicollis*). *Journal of Zoology* **247**: 53-64.
- Sazama, E. J., Bosch, M. J., Shouldis, C. S., Ouellette, S. P. & Wesner, J. S. 2017. Incidence of *Wolbachia* in aquatic insects. *Ecol Evol* **7**: 1165-1169.
- Schluter, D. 2001. Ecology and the origin of species. *Trends in ecology & evolution* **16**: 372-380.
- Schluter, D. & Nagel, L. M. 1995. Parallel speciation by natural selection. *The American Naturalist* **146**: 292-301.
- Seehausen, O. 2009. Progressive levels of trait divergence along a 'speciation transect' in the Lake Victoria cichlid fish *Pundamilia*. *Ecological Reviews: Speciation and Patterns of Diversity*: 155-176.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G. P., Bank, C., Brannstrom, A., Brelsford, A., Clarkson, C. S., Eroukmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., Lindholm, A. K., Lucek, K., Maan, M. E., Marques, D. A., Martin, S. H., Matthews, B., Meier, J. I., Most, M.,

- Nachman, M. W., Nonaka, E., Rennison, D. J., Schwarzer, J., Watson, E. T., Westram, A. M. & Widmer, A. 2014. Genomics and the origin of species. *Nat Rev Genet* **15**: 176-92.
- Servedio, M. R. & Noor, M. A. 2003. The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics* **34**: 339-364.
- Shoemaker, D. D., Dyer, K. A., Ahrens, M., McAbee, K. & Jaenike, J. 2004. Decreased diversity but increased substitution rate in host mtDNA as a consequence of Wolbachia endosymbiont infection. *Genetics* **168**: 2049-58.
- Shuster, S. M. & Sassaman, C. 1997. Genetic interaction between male mating strategy and sex ratio in a marine isopod. *Nature* **388**: 373.
- Siegismund, H. R. 2002. Disparity in population differentiation of sex-linked and autosomal variation in sibling species of the *Jaera albifrons* (Isopoda) complex. *The Journal of Heredity* **93**: 432-439.
- Simoes, P. M., Mialdea, G., Reiss, D., Sagot, M. F. & Charlat, S. 2011. Wolbachia detection: an assessment of standard PCR protocols. *Mol Ecol Resour* **11**: 567-72.
- Smadja, C. M. & Butlin, R. K. 2011. A framework for comparing processes of speciation in the presence of gene flow. *Mol Ecol* **20**: 5123-40.
- Smadja, C. M., Canbäck, B., Vitalis, R., Gautier, M., Ferrari, J., Zhou, J. J. & Butlin, R. K. 2012. Large-Scale Candidate Gene Scan Reveals The Role Of Chemoreceptor Genes In Host Plant Specialization And Speciation In The Pea Aphid. *Evolution* **66**: 2723-2738.
- Sobel, J. M. & Chen, G. F. 2014. Unification of methods for estimating the strength of reproductive isolation. *Evolution* **68**: 1511-22.
- Solignac, M. 1969a. Contiguous distributions sympatry and natural hybridization in a super-species *Jaera albifrons* Leach (isopods asellidae). *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences SERIE D* **268**: 1610-&.
- Solignac, M. 1969b. Hybridation introgressive dans la population complexe des *Jaera albifrons* de Luc sur Mer. *Arch Zool Exp Gén* **110**: 629-652.
- Solignac, M. 1976. Demographic Aspects of Interspecific Hybridization. *Oecologia* **26**: 33-52.
- Solignac, M. (1978a) Genetics of ethological isolating mechanisms in the species complex *Jaera albifrons* (Crustacea, Isopoda). In: *NATO Conf Series*, Vol. 2. pp. 637-664.
- Solignac, M. (1978b) Nature, déterminisme et origine des mécanismes d'isolement dans le complexe *Jaera albifrons* (Isopodes, Asellotes). In: *Phd Thesis, Paris 6*. pp. 232.
- Solignac, M. 1981. Isolating mechanisms and modalities of speciation in the *Jaera albifrons* species complex (Crustacea, Isopoda). *Systematic Zoology* **30**: 387-405.
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., Buerkle, C. A., Feder, J. L., Bast, J., Schwander, T., Egan, S. P., Crespi, B. J. & Nosil, P. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**: 738-42.
- Sousa, V. & Hey, J. 2013. Understanding the origin of species with genome-scale data: modelling gene flow. *Nat Rev Genet* **14**: 404-14.
- Southcott, L. & Kronforst, M. R. 2017. A neutral view of the evolving genomic architecture of speciation. *Ecology and Evolution* **7**: 6358-6366.
- Staiger, H. & Bocquet, C. 1954. Cytological demonstration of female heterogamety in isopods. *Experientia* **10**: 64-66.

- Stephan, W. 2010. Genetic hitchhiking versus background selection: the controversy and its implications. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **365**: 1245-1253.
- Sun, C., Niu, Y., Ye, X., Dong, J., Hu, W., Zeng, Q., Chen, Z., Tian, Y., Zhang, J. & Lu, M. 2017. Construction of a high-density linkage map and mapping of sex determination and growth-related loci in the mandarin fish (*Siniperca chuatsi*). *BMC genomics* **18**: 446.
- Suzuki, G. & Fukami, H. 2012. Evidence of genetic and reproductive isolation between two morphs of subtropical-dominant coral *Acropora solitaryensis* in the non-reef region of Japan. *Zoological science* **29**: 134-140.
- Taylor, D. R. 1990. Evolutionary consequences of cytoplasmic sex ratio distorters. *Evolutionary Ecology* **4**: 235-248.
- Taylor, M. J., Bandi, C. & Hoerauf, A. 2005. Wolbachia. Bacterial endosymbionts of filarial nematodes. *Advances in parasitology* **60**: 245-284.
- Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W. & Tucker, P. K. 2010. The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* **64**: 472-85.
- Telschow, A., Flor, M., Kobayashi, Y., Hammerstein, P. & Werren, J. H. 2007. Wolbachia-induced unidirectional cytoplasmic incompatibility and speciation: mainland-island model. *PLoS One* **2**: e701.
- Telschow, A., Hammerstein, P. & Werren, J. H. 2005. The effect of Wolbachia versus genetic incompatibilities on reinforcement and speciation. *Evolution* **59**: 1607-1619.
- Thomas, F., Lefevre, T. & Raymond, M. 2010. *Biologie évolutive*.
- Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A., Bargelloni, L., Power, D. M., Bonhomme, F., Canario, A. V. & Reinhardt, R. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun* **5**: 5770.
- Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS biology* **3**: e285.
- Valette, V., Durand, S., Bech, N., Grandjean, F. & Beltran-Bech, S. 2017. Multiple paternity in a wild population of *Armadillidium vulgare*: influence of infection with Wolbachia? *J Evol Biol* **30**: 235-243.
- van Doorn, G. S., Edelaar, P. & Weissing, F. J. 2009. On the origin of species by natural and sexual selection. *Science* **326**: 1704-1707.
- Van Ooijen, J. 2006. JoinMap 4. *Software for the calculation of genetic linkage maps in experimental populations*. Kyazma BV, Wageningen, Netherlands **33**.
- Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. 2004. micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**: 535-538.
- Vavre, F., Fleury, F., Lepetit, D., Fouillet, P. & Boulétreau, M. 1999. Phylogenetic evidence for horizontal transmission



- of Wolbachia in host-parasitoid associations. *Molecular Biology and Evolution* **16**: 1711-1723.
- Veltsos, P., Gregson, E., Morrissey, B., Slate, J., Hoikkala, A., Butlin, R. K. & Ritchie, M. G. 2015. The genetic architecture of sexually selected traits in two natural populations of *Drosophila montana*. *Heredity (Edinb)* **115**: 565-72.
- Veuille, M. 1979. L'évolution du genre *Jaera* Leach (Isopodes; Asellotes) et ses rapports avec l'histoire de la Méditerranée. *Bijdragen tot de Dierkunde* **49**: 195-217.
- Veuille, M. 1980. Sexual behaviour and evolution of sexual dimorphism in body size in *Jaera* (Isopoda Asellota). *Biological Journal of the Linnean Society* **13**: 89-100.
- Via, S. & West, J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular ecology* **17**: 4334-4345.
- 
- Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A. & Seehausen, O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol Ecol* **22**: 787-98.
- Wagner, C. E. & Mandeville, E. G. 2017. Speciation, species persistence and the goals of studying genomic barriers to gene flow. *J Evol Biol* **30**: 1512-1515.
- Welch, J. J. & Jiggins, C. D. 2014. Standing and flowing: the complex origins of adaptive variation. *Molecular ecology* **23**: 3935-3937.
- Werren, J. H. 1983. Sex ratio evolution under local mate competition in a parasitic wasp. *Evolution* **37**: 116-124.
- Werren, J. H. 1997. Biology of wolbachia. *Annual review of entomology* **42**: 587-609.
- Werren, J. H., Baldo, L. & Clark, M. E. 2008. Wolbachia: master manipulators of invertebrate biology. *Nature Reviews Microbiology* **6**: 741-751.
- Werren, J. H. & Windsor, D. M. 2000. Wolbachia infection frequencies in insects: evidence of a global equilibrium? *Proc Biol Sci* **267**: 1277-85.
- Werren, J. H., Zhang, W. & Guo, L. R. 1995. Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. *Proceedings of the Royal Society of London B: Biological Sciences* **261**: 55-63.
- West, S. A. & Sheldon, B. C. 2002. Constraints in the evolution of sex ratio adjustment. *Science* **295**: 1685-1688.
- Westram, A. M., Galindo, J., Alm Rosenblad, M., Grahame, J. W., Panova, M. & Butlin, R. K. 2014. Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Mol Ecol* **23**: 4603-16.
- Westram, A. M., Panova, M., Galindo, J. & Butlin, R. K. 2016. Targeted resequencing reveals geographical patterns of differentiation for loci implicated in parallel evolution. *Molecular ecology* **25**: 3169-3186.
- Wielstra, B., Burke, T., Butlin, R. K., Avci, A., Üzümlü, N., Bozkurt, E., Olgun, K. & Arntzen, J. W. 2017. A genomic footprint of hybrid zone movement in crested newts. *Evolution Letters* **1**: 93-101.
- Wolf, J. B. & Ellegren, H. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet* **18**: 87-100.
- Wood, T., Burke, J. & Rieseberg, L. 2005. Parallel genotypic adaptation: when evolution repeats itself. *Genetics of Adaptation*: 157-170.

Wu, C. I. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851-865.

x

Xu, Y., Bi, C., Wu, G., Wei, S., Dai, X., Yin, T. & Ye, N. 2016. VGSC: a web-based Vector Graph toolkit of genome Synteny and Collinearity. *BioMed research international* **2016**.

y

Yeates, S. E., Diamond, S. E., Einum, S., Emerson, B. C., Holt, W. V. & Gage, M. J. 2013. Cryptic choice of conspecific sperm controlled by the impact of ovarian fluid on sperm swimming behavior. *Evolution* **67**: 3523-3536.

Yukilevich, R. 2012. Asymmetrical patterns of speciation uniquely support

reinforcement in *Drosophila*. *Evolution* **66**: 1430-1446.

2

Zardi, G. I., Nicastro, K. R., Canovas, F., Costa, J. F., Serrão, E. A. & Pearson, G. A. 2011. Adaptive traits are maintained on steep selective gradients despite gene flow and hybridization in the intertidal zone. *PloS one* **6**: e19402.

Zhou, W., Rousset, F. & O'Neill, S. 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. *Proceedings of the Royal Society of London B: Biological Sciences* **265**: 509-515.

Zug, R. & Hammerstein, P. 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PloS one* **7**: e38544.

RESUME

Dans l'étude de la spéciation, l'existence de zones dans lesquelles deux espèces en sympatrie montrent différents niveaux de différenciation génétique constitue une bonne opportunité pour comprendre la nature et l'évolution des barrières à la reproduction entre espèces.

Jaera (albifrons) albifrons et *Jaera (albifrons) prae-hirsuta* sont deux espèces d'isopodes marins qui coexistent en sympatrie le long des côtes nord-européennes. Des barrières écologiques, comportementales, et génétiques cloisonnent efficacement ces deux espèces, à l'exception d'une région unique (Normandie, France) dans laquelle des morphotypes intermédiaires avaient été décrits dans les années 60.

Au cours de cette thèse, l'utilisation de microsatellites et de SNPs obtenus par séquençage RAD ont permis de démontrer que le niveau d'isolement interspécifique varie fortement entre sites, de l'hybridation introgressive à l'isolement quasi-complet. Une analyse comparative de ces sites combinant échantillonnage en populations naturelles, croisements expérimentaux et analyses génomiques a ensuite permis de : i) mettre en avant le rôle prépondérant de l'isolement sexuel (qui reste fort dans les populations introgressées) accompagné d'une barrière post-zygotique relativement faible, ii) découvrir la présence de bactéries *Wolbachia* au sein des deux espèces, iii) démontrer que la coexistence des deux espèces résulte d'une spéciation allopatrique suivie de contacts secondaires avec reprise de flux de gènes d'intensité variable, et iv) montrer que ces flux de gènes varient également fortement au sein du génome, les chromosomes sexuels et des chromosomes réarrangés semblant limiter fortement l'introgression.

Mots-clés : différenciation interspécifique, hybridation, barrières à la reproduction, analyses comparatives, zones hybrides

ABSTRACT

Within the field of speciation, sympatric areas with different levels of interspecific genetic differentiation offer a good opportunity to understand the nature and evolution of reproductive barriers between species.

Jaera (albifrons) albifrons and *Jaera (albifrons) prae-hirsuta* are two species of marine isopods that coexist in sympatry along the northern European coasts. Ecological, behavioral and genetic barriers efficiently isolate these two species, except in a unique region (Normandy, France) where morphological phenotypes were described in the 60's.

In this thesis, microsatellites and SNPs obtained from RAD-sequencing allowed me to demonstrate that the level of interspecific isolation varies widely between sites, from introgressive hybridization to quasi-complete isolation. A comparative analysis of these sites combining sampling from natural populations, experimental crosses, and genomic analyses then allowed me to: i) demonstrate the predominant role of sexual isolation (which remains strong in introgressed populations), together with a relatively weaker post-zygotic barrier, ii) discover the presence of *Wolbachia* bacteria within the two species, iii) demonstrate that the coexistence of these species originate from an allopatric speciation followed by secondary contacts with varying levels of heterospecific gene flow renewal, and iv) show that gene flow varies also strongly along the genome, with an effect of sex chromosomes and rearranged chromosomes apparently limiting introgression.

Key words : interspecific differentiation, hybridization, reproductive barriers, comparative analyses, hybrid areas