



**HAL**  
open science

# Adaptation of symbiotic coral micro-algae to environmental changes in the Pacific Ocean

Julie Lê-Hoang

► **To cite this version:**

Julie Lê-Hoang. Adaptation of symbiotic coral micro-algae to environmental changes in the Pacific Ocean. Ecology, environment. Université Paris-Saclay, 2021. English. NNT : 2021UPASL091 . tel-03627907

**HAL Id: tel-03627907**

**<https://theses.hal.science/tel-03627907>**

Submitted on 1 Apr 2022

**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.

Adaptation des micro-algues  
symbiotiques du corail aux  
changements environnementaux  
dans l'Océan Pacifique

*Adaptation of symbiotic coral  
micro-algae to environmental changes  
in the Pacific Ocean*

**Thèse de doctorat de l'Université Paris-Saclay**

École doctorale n° 577, Structure et Dynamique des  
Systèmes Vivants (SDSV)

Spécialité de doctorat : Sciences de la Vie et de la Santé

Unité de recherche : Université Paris-Saclay, Univ Evry, CNRS, CEA,  
Génomique métabolique, 91057, Evry-Courcouronnes, France.

Référent : Université d'Évry-Val-d'Essonne

**Thèse présentée et soutenue à Évry, le 15/12/2021, par**

**Julie LÊ-HOANG**

**Composition du jury**

<b>Daniel GAUTHERET</b> PR, Université Paris-Saclay/I2BC, UMR - 9198	Président du Jury
<b>Hélène MAGALON</b> MC, Université de la Réunion/Ifremer/CNRS, UMR - EN- TROPIC	Rapporteuse
<b>Cécile SABOURAULT</b> PR, Université Côte d'Azur/ECOSEAS, UMR - 7035	Rapporteuse
<b>Guillaume MITTA</b> PR, Université de Perpignan/IHPE, UMR - 5244	Examinateur

**Direction de la thèse**

<b>Patrick WINCKER</b> DR CEA, CEA/Genoscope, Évry - UMR - 8030	Directeur de Thèse
<b>Quentin CARRADEC</b> Chercheur CEA, CEA/Genoscope, Évry - UMR - 8030	Co-Encadrant



*"Les espèces qui survivent ne sont pas les espèces les plus fortes, ni les plus intelligentes,  
mais celles qui s'adaptent le mieux aux changements."*

*Charles Darwin (1809 - 1882)*



# Remerciements

Mes sincères remerciements à toutes les personnes qui ont participé de près ou de loin à l'élaboration et à la concrétisation de cette thèse.

En premier lieu, je tiens à remercier mon directeur de thèse Patrick Wincker, qui m'a aidée, soutenue et accueillie dans son laboratoire, pendant ces années de thèse. Un grand merci pour ses conseils.

Je remercie Quentin Carradec mon encadrant de thèse, pour son encadrement, pendant la thèse.

Je remercie particulièrement Eric Armstrong, ancien post-doctorant au LAGE, qui m'a aidée, soutenue et beaucoup appris sur la science des coraux. Toujours bienveillant envers moi, merci pour le temps précieux que tu m'as accordée, pour la transmission de ton savoir scientifique et pour ces moments partagés, pendant la thèse.

Je remercie particulièrement également Julie Poulain, membre du LAGE qui m'a aidée et soutenue. Un grand merci pour nos échanges scientifiques et personnelles, merci pour ton implication dans ma thèse.

Je tiens à remercier particulièrement mes confrères Romuald Laso-Jadart et Kévin Sugier, pour leur aide, leur soutien et tous ces grands moments passés dans le "bureau des thésards". Des débats animés sur la frontière du Sud et du Nord de la France, nos études sur la prise de poids chez les Géno-thésards et tous ces moments passés avec vous, dans et en dehors du Genoscope.

Un grand merci aux autres thésards du LAGE, en particulier merci à Nina Guérin, Marie Burel et Paul Frémont, ainsi qu'à Janaina Rigonato post-doctorante au LAGE, avec qui les soirées au Genoscope, les Genopubs et les conférences ont été des meilleures. Merci pour leur soutien et pour leur compagnie pendant ces quelques années passées à vos côtés.

Merci à mon laboratoire d'accueil le LAGE, merci à Olivier Jaillon, Eric Pelletier, Betina Porcel, Amin Madoui, Adrien Thurotte, Evangelia Eleftheriou, Céline Orvain, Laurie Bertrand, Sophie Layac, pour leur accueil, leur conseil et leur bienveillance envers moi. Merci également aux nouveaux arrivants et aux anciens du LAGE, Jade Leconte, Arnaud Meng, Sarah Farhat, Alexei Vorobev et Adriana Alberti.

Merci à toute l'équipe LBGB, qui m'a beaucoup aidée pendant cette thèse et avec qui j'ai passé de très bons moments également. Merci pour vos conseils et votre accueil, en particulier merci à Benjamin Noel, Jean Marc Aury, Corinne Da Silva, France Denoeud, Caroline Belser, Stefan Engelen, Arnaud Couloux, Nachida Tadrent, Aimeric Bruno, Benjamin Istace, Marion Dubarry, Fred Gavory, Simone Duprat, Salima Kritli et Paul Mielle.

Un grand merci à l'équipe Système du Genoscope, sans qui cette thèse aurait pu être plus complexe, merci à Claude Scarpelli, Franck Anière et Sylvain Bonneval pour leur aide précieuse.

Merci à toute l'équipe de Production pour leur implication dans ma thèse, merci à toute l'équipe pour la production des données Tara qui m'ont permises de réaliser cette thèse. Un grand merci en particulier à Karine Labadie, Emmanuelle Petit, Corinne Cruaud, Eric Mathieu, Maria Dubois, Muriel Ronsin, Aude Perdereau, Pedro Oliveira, Mélanie Beraud et Peggy Perroud.

Je remercie particulièrement les secrétaires Catherine Sarlande, Catherine Contrefois et Nancy Delpech, toujours bienveillantes et d'une gentillesse à toute épreuve. Merci à vous.

Merci aux membres du LABGEM, David Vallenet, Mathieu Dubois et Adelme Bazin. Merci pour les soirées Genopubs animées et organisées grâce à vous.

Merci à tous les autres thésards, en particulier merci à Oriane Della-Negra, Chloé Lelièvre, Laurine Ducrot et Déborah Martins. De supers moments partagés avec vous et de bons souvenirs en mémoire.

Merci aux enseignants de l'Université d'Évry pour m'avoir fait confiance et pour m'avoir transmis la passion de l'enseignement pendant cette thèse. En particulier, merci à Valérie Chaudru, Elisabeth Petit-Teixeira, Carène Rizzon, Nathalie Boudet et Anne Zaparucha.

Merci aux personnels de l'école doctorale, au consortium Tara, aux membres de mon comité de thèse, aux membres de mon jury de thèse, aux restaurants (Chez Serge, Le Palais, Le Bistrot Evryen, Picard, et tous les autres)..

Pour finir, un grand merci à mes proches et à ma famille, en particulier à Esteban, ma mère, mon père, ma soeur, mon frère, Christophe, Indiana et Jones, mais également merci à ma belle famille etc. Sans vous la thèse aurait été plus compliquée, merci d'avoir été là, de m'avoir soutenue et de m'avoir écoutée et épaulée pendant cette thèse.





# Contents

<b>List of figures</b>	<b>8</b>
<b>List of tables</b>	<b>9</b>
<b>Abbreviations</b>	<b>12</b>
<b>I State of the art</b>	<b>15</b>
I.1 Ecology of coral reefs . . . . .	15
I.1.1 Coral Taxonomy . . . . .	15
I.1.2 Coral life cycle and morphological diversity . . . . .	17
I.1.2.1 Coral life cycle . . . . .	17
I.1.2.2 Coral morphological diversity . . . . .	19
I.1.3 Geographic distribution of coral reefs . . . . .	20
I.1.4 Biodiversity of coral reefs . . . . .	22
I.1.5 Anthropogenic importance of coral reefs . . . . .	24
I.1.6 Coral reefs threats . . . . .	24
I.1.6.1 Bleaching events . . . . .	24
I.1.6.2 A multitude of threats . . . . .	26
I.1.7 Coral studies . . . . .	28
I.2 Corals in symbiosis . . . . .	31
I.2.1 Obligatory symbiosis with the micro-algae Symbiodiniaceae . . . . .	31
I.2.2 Symbiosis breakdowns . . . . .	33
I.3 Biology of Symbiodiniaceae . . . . .	38
I.3.1 Description of Symbiodiniaceae . . . . .	38
I.3.2 Symbiodiniaceae life cycle . . . . .	39
I.3.3 Symbiodiniaceae in symbiosis and heritability . . . . .	40
I.3.4 Free-living form of Symbiodiniaceae . . . . .	40
I.4 Symbiodiniaceae diversity . . . . .	43
I.4.1 Symbiodiniaceae identification . . . . .	43
I.4.2 Symbiodiniaceae hosts . . . . .	44
I.4.3 Symbiodiniaceae phylogeny . . . . .	45
I.5 Genomes and transcriptomes of Symbiodiniaceae . . . . .	49
I.5.1 Symbiodiniaceae references . . . . .	49
I.5.2 Gene regulation in Symbiodiniaceae . . . . .	50
I.5.3 Symbiodiniaceae under heat stress . . . . .	50
I.5.4 Symbiodiniaceae adaptation to light intensity . . . . .	52
I.5.5 Symbiodiniaceae nutrient transport . . . . .	53
I.5.6 Symbiodiniaceae adaptation in different location . . . . .	54
I.6 The <i>Tara</i> Pacific expedition . . . . .	55

I.7	Aim of the thesis . . . . .	58
<b>II</b>	<b>Symbiodiniaceae diversity study in the <i>Tara</i> Pacific project</b>	<b>61</b>
II.1	Materials and Methods . . . . .	62
II.1.1	Availability of Symbiodiniaceae transcriptomic and genomic references . . . . .	62
II.1.1.1	SymPortal: a database dedicated to Symbiodiniaceae ITS2 sequences . . . . .	62
II.1.1.2	Symbiodiniaceae transcriptome references . . . . .	63
II.1.1.3	Symbiodiniaceae genome references . . . . .	65
II.1.2	GC content estimation and cleaning of Symbiodiniaceae references . . . . .	65
II.1.3	Detection of potential contamination in Symbiodiniaceae references . . . . .	65
II.1.4	Coral sampling for RNA sequencing . . . . .	69
II.1.4.1	Coral sampling . . . . .	69
II.1.4.2	DNA and RNA extractions . . . . .	71
II.1.4.3	mRNA selection and libraries preparation . . . . .	71
II.1.4.4	Illumina sequencing . . . . .	71
II.1.5	Bioinformatics methods to align dual-transcriptomic reads on Symbiodiniaceae references . . . . .	74
II.1.6	Population structure of <i>Cladocopium C1</i> . . . . .	75
II.1.6.1	Metagenomic samples used to understand <i>Cladocopium C1</i> diversity . . . . .	76
II.1.6.2	Metatranscriptomic samples: a new kind of data to do population structure . . . . .	76
II.2	Results . . . . .	79
II.2.1	Detection of the contaminations in Symbiodiniaceae references . . . . .	79
II.2.2	Estimating Symbiodiniaceae diversity in the <i>Tara</i> Pacific coral samples using all Symbiodiniaceae reference transcriptomes . . . . .	82
II.2.3	Population structure of <i>Cladocopium C1</i> in <i>Pocillopora sp.</i> . . . . .	93
II.2.3.1	Analyses of ITS2 sequences using metabarcoding data . . . . .	93
II.2.3.2	Revealing sub-variants of the ITS2 sequences in <i>Cladocopium C1</i> . . . . .	94
II.2.3.3	Genomic vertical coverage to identify the population structure of <i>Cladocopium C1</i> . . . . .	96
II.2.3.4	Population structure of <i>Cladocopium-C1</i> using metatranscriptomic samples . . . . .	98
II.3	Discussion and Conclusions . . . . .	101
II.3.1	A multitude of Symbiodiniaceae genomes and transcriptomes . . . . .	101
II.3.2	Diversity of Symbiodiniaceae in the three coral species studied . . . . .	102
II.3.2.1	A limited Symbiodiniaceae diversity in <i>Porites sp.</i> colonies . . . . .	102
II.3.2.2	Two Symbiodiniaceae genera in <i>Millepora sp.</i> colonies . . . . .	103
II.3.2.3	High host/symbiont fidelity in <i>Pocillopora sp.</i> colonies. . . . .	104
II.3.3	Classical bioinformatics methods to construct <i>Cladocopium C1</i> population structure . . . . .	104
II.3.4	A new definition of Symbiodiniaceae diversity . . . . .	105
<b>III</b>	<b>Adaptation of Symbiodiniaceae in symbiosis across the Pacific Ocean</b>	<b>107</b>
III.1	Materials and Methods . . . . .	108
III.1.1	Differential expression analyses - Determination of biological replicates . . . . .	108
III.1.1.1	Expression profiles of the <i>Tara</i> transcriptomic datasets . . . . .	108
III.1.1.2	Determination of biological replicates - Environmental conditions . . . . .	108
III.1.1.3	Determination of biological replicates - Genetic conditions . . . . .	109

III.1.2	bioinformatics methods to perform differential expression analyses . . . . .	110
III.1.3	bioinformatics methods to explain the Symbiodiniaceae expression profiles with associated functions . . . . .	111
III.1.3.1	Gene Ontology or GO terms . . . . .	111
III.1.3.2	Protein families or PFAM . . . . .	112
III.1.3.3	Kyoto Encyclopedia of Genes and Genomes or KEGG . . . . .	113
III.1.3.4	<i>Symbiodinium A3</i> and <i>Cladocopium C1</i> annotations . . . . .	114
III.2	Results . . . . .	115
III.2.1	Expression level of Symbiodiniaceae in dual-transcriptomic samples . . . . .	115
III.2.1.1	Expression level of Symbiodiniaceae in dual-transcriptomic samples - PCA . . . . .	115
III.2.1.2	Expression level of Symbiodiniaceae in dual-transcriptomic samples - Pearson correlations . . . . .	120
III.2.2	Environmental adaptation of Symbiodiniaceae across Pacific Ocean . . . . .	124
III.2.2.1	Environmental adaptation of <i>Symbiodinium A3</i> in the Pacific Ocean	124
III.2.2.2	Environmental adaptation of <i>Cladocopium C15</i> in the Pacific Ocean . . . . .	125
III.2.2.3	Environmental adaptation of <i>Cladocopium C1</i> in the Pacific Ocean	127
III.2.2.4	Environmental adaptation of Symbiodiniaceae across the Pacific Ocean . . . . .	129
III.2.2.5	Genetic adaptation of <i>Cladocopium C1</i> across the Pacific Ocean	133
III.2.3	Differentially expressed functions for the adaptation of Symbiodiniaceae in the Pacific Ocean . . . . .	134
III.2.3.1	Expression and associated functions for the adaptation of <i>Symbiodinium A3</i> in <i>Millepora sp.</i> samples across the Pacific Ocean	134
III.2.3.2	Differentially expressed functions of <i>Cladocopium C1</i> in <i>Pocillopora sp.</i> samples across the Pacific Ocean . . . . .	139
III.3	Transcriptomic plasticity and symbiont shuffling underpin <i>Pocillopora</i> acclimatization across heat-stress regimes in the Pacific Ocean . . . . .	142
III.4	Discussion and Conclusions . . . . .	182
III.4.1	Convergent expression patterns of Symbiodiniaceae across the Pacific Ocean in three different corals . . . . .	182
III.4.2	<i>In situ</i> transcriptomic adaptation of Symbiodiniaceae at island scale . . . . .	183
III.4.3	Gene functions characterization of <i>Symbiodinium A3</i> in different environmental conditions . . . . .	184
III.4.4	Expression and associated functions for the adaptation of <i>Cladocopium C1</i> in <i>Pocillopora sp.</i> . . . . .	185
III.4.5	Collaborations . . . . .	186
<b>IV</b>	<b>Adaptation of Symbiodiniaceae in surrounding water across the Pacific Ocean</b>	<b>189</b>
IV.1	Materials and Methods . . . . .	191
IV.1.1	Coral surrounding water sampling to RNA sequencing . . . . .	191
IV.1.1.1	Surrounding water sampling . . . . .	191
IV.1.1.2	Illumina sequencing . . . . .	192
IV.1.2	bioinformatics methods to align transcriptomic reads on genome and transcriptome references . . . . .	193
IV.1.3	Population structure of <i>Cladocopium C1</i> in coral surrounding water . . . . .	194
IV.1.4	bioinformatics methods to determine the genetic and the environmental adaptation of Symbiodiniaceae, in the seawater compared to the coral host	194

IV.2	Results	196
IV.2.1	Symbiodiniaceae diversity in different environmental samples across the Pacific Ocean	196
IV.2.1.1	Symbiodiniaceae diversity in coral surrounding water using Symbiodiniaceae reference transcriptomes	196
IV.2.1.2	Potential coral contamination in coral surrounding water samples	199
IV.2.1.3	Symbiodiniaceae diversity in the <i>Tara</i> Pacific surface samples using Symbiodiniaceae references	200
IV.2.1.4	Population structure of <i>Cladocopium C1</i> in coral surrounding water samples	202
IV.2.2	Gene expression variations of <i>Cladocopium C1</i> between CSW and coral samples	204
IV.2.2.1	Expression level of <i>Cladocopium C1</i> genes in coral surrounding water samples	204
IV.2.2.2	Differentially expressed genes of <i>Cladocopium C1</i> between coral and water samples	207
IV.3	Discussion and Conclusions	211
IV.3.1	bioinformatics methods to understand the acclimation of Symbiodiniaceae hospite versus free-living.	211
IV.3.2	Symbiodiniaceae diversity hospite versus free-living	211
IV.3.3	<i>Cladocopium C1</i> adaptation in seawater versus coral host	212
<b>V</b>	<b>Discussion and Conclusions</b>	<b>215</b>
V.1	Symbiodiniaceae diversity study in the <i>Tara</i> Pacific project	215
V.2	Adaptation of Symbiodiniaceae in symbiosis across the Pacific Ocean	216
V.3	Adaptation of Symbiodiniaceae in surrounding water across the Pacific Ocean	217
V.4	Prospects	218
	<i>BIBLIOGRAPHY</i>	219
	<i>APPENDICES</i>	231
	<i>APPENDIX I : Résumé de la thèse en français</i>	232
	<i>APPENDIX II : Illumina sequencing</i>	243
	<i>APPENDIX III : Detection of the contaminations in Symbiodiniaceae references</i>	244
	<i>APPENDIX IV : Estimating Symbiodiniaceae diversity in the Tara Pacific coral samples using all Symbiodiniaceae reference transcriptomes</i>	248
	<i>APPENDIX V : Estimating Symbiodiniaceae diversity in the Tara Pacific coral samples using all Symbiodiniaceae reference transcriptomes</i>	252
	<i>APPENDIX VI : Estimating Symbiodiniaceae diversity in the Tara Pacific coral samples using all Symbiodiniaceae reference transcriptomes</i>	255
	<i>APPENDIX VII : Revealing sub-variants of the ITS2 sequences in Cladocopium C1</i>	256
	<i>APPENDIX VIII: Genomic vertical coverage to identify the population structure of Cladocopium C1</i>	258
	<i>APPENDIX IX: Population structure of Cladocopium C1 using metatranscriptomic samples</i>	259
	<i>APPENDIX X: Differentially expressed functions of Cladocopium C1 in Pocillopora sp. samples across the Pacific Ocean</i>	261
	<i>APPENDIX XI: Differentially expressed functions of Cladocopium C1 in Pocillopora sp. samples across the Pacific Ocean</i>	262

# List of Figures

I.1	Cnidarian phylogenetic tree based on mitochondrial protein genes ( <b>Kayal E. et al. 2013</b> ).	16
I.2	Polyp schema of coral. . . . .	17
I.3	Coral life cycle schema, inspired by <b>Thompson J. et al. 2015</b> . . . . .	18
I.4	<i>In situ</i> images of several <i>Millepora</i> specimens examined ( <b>Takama et al. 2018</b> ). . . . .	19
I.5	Images of <i>Pocillopora ligulata</i> colonies, <b>A-E</b> ; <i>P. meandrina</i> colonies, <b>F-J</b> ; and <i>P. eydouxi</i> colonies, <b>K-O</b> from O’ahu, Hawai’i from <b>Johnston et al. 2018</b> . . . . .	20
I.6	Global diversity of coral reefs at the scale of 141 reef eco-regions ( <b>Huang et al. 2015</b> ). . . . .	21
I.7	The four main coral reef types with their descriptions and some examples of their locations ( <b>Moberg et al. 1999</b> ). . . . .	21
I.8	Schema of the different coral reef types, inspired from <b>Darwin et al. 1842</b> . . . . .	21
I.9	Coral reefs as the base of the marine food alimentary chain. . . . .	22
I.10	Interactions in the tropical seascape, showing the connections between mangroves, sea-grass meadows and coral reefs ( <b>Moberg et al. 1999</b> ). . . . .	23
I.11	Estimated expenditures by tourists within the ten jurisdictions with the highest total values. ( <b>Spalding et al. 2017</b> ). . . . .	24
I.12	Recurrent severe coral bleaching ( <b>Hughes et al. 2017</b> ). . . . .	25
I.13	The global extent of mass bleaching events of coral reefs in 2015 and in 2016 ( <b>Hughes et al. 2018</b> ). . . . .	25
I.14	List of some of the more comprehensively documented field assessments on the effects of enhanced terrestrial runoff and other forms of pollution, on the ecology of coral reefs ( <b>Fabricius et al. 2005</b> ). . . . .	27
I.15	View of unchanged (A) and changed (B, C) colonies at low-flow and high-flow sites south of the Gulf of California, March 2011–November 2014 ( <b>Paz-García et al. 2015</b> ). . . . .	28
I.16	DNA analysis has helped to recognize that many species of fish and invertebrates that were once thought to have widespread distributions are in fact made up of multiple distinct genetic lineages ( <b>Knowlton et al. 2015</b> ). . . . .	30
I.17	Location of Symbiodiniaceae in coral polyps ( <b>Fournier et al. 2013</b> ). . . . .	31
I.18	Metabolic interactions and calcification processes between Symbiodiniaceae and coral polyps ( <b>Fournier et al. 2013</b> ). . . . .	32
I.19	Five different types of cellular mechanisms of symbiont loss from cnidarian host tissues ( <b>Weis et al. 2008</b> ). . . . .	33
I.20	Reactions leading to bleaching ( <b>Fournier A. et al. 2013</b> ). . . . .	35
I.21	Ecological feedback processes on a coral reef showing pathways of disturbance caused by climate change ( <b>Hoegh-Guldberg et al. 2007</b> ). . . . .	37
I.22	Transmission electron micrograph of cross-section through a dividing cell ( <b>Trench et al. 1981</b> ). . . . .	38
I.23	The life stages of dinoflagellates in the Symbiodiniaceae family ( <b>Trench et al. 1981</b> ). . . . .	38
I.24	Schema of the Symbiodiniaceae production ( <b>Fitt and Trench 1983</b> ). . . . .	39
I.25	Biogeography and diversity of Symbiodiniaceae in the open ocean ( <b>Decelle et al. 2018</b> ). . . . .	41

I.26	Relative proportions of free-living Symbiodiniaceae found in each habitat during the different sampling periods ( <b>Granados-Cifuentes et al. 2013</b> ). . . . .	42
I.27	A schematic representation of the multiple levels of diversity found within Symbiodiniaceae ( <b>Coffroth M.A et al. 2005</b> ). . . . .	43
I.28	Phylogenetic relationships between the major clades of Symbiodiniaceae ( <b>Fournier et al. 2013</b> ). . . . .	44
I.29	Proposed new genera in the family Symbiodiniaceae and LSU rDNA phylogeny portraying their evolutionary relationships, including remaining lineages that require generic names ( <b>LaJeunesse et al. 2018</b> ). . . . .	45
I.30	An unrooted phylogenetic reconstruction of clade A based on maximum parsimony analysis of ITS2 sequences ( <b>LaJeunesse et al. 2009</b> ). . . . .	46
I.31	Independent radiations of Symbiodiniaceae clade C in the (a) Indo-Pacific and (b) Atlantic-Caribbean ( <b>LaJeunesse et al. 2004</b> ). . . . .	47
I.32	Morphological differences within and among Symbiodiniaceae clades ( <b>LaJeunesse et al. 2018</b> ). . . . .	48
I.33	Breakthroughs in NGS of Symbiodiniaceae ( <b>Levin et al. 2017</b> ). . . . .	49
I.34	Differential responses of <i>Porites lobata</i> to experimental heat stress treatments ( <b>Hume et al. 2012</b> ). . . . .	51
I.35	Scanning electron microscopy (SEM) of Symbiodiniaceae showing surface structures under different light conditions ( <b>Xiang et al. 2015</b> ). . . . .	52
I.36	Schematic summary of host recognition mechanisms and cargo transport ( <b>Lin et al. 2015</b> ). . . . .	53
I.37	The <i>Tara</i> Pacific expedition . . . . .	55
I.38	Fishes and corals species sampled during the <i>Tara</i> Pacific expedition. . . . .	56
I.39	Sampling processes during the <i>Tara</i> Pacific expedition. . . . .	56
I.40	<i>Tara</i> Pacific sampling plan for each island with metatranscriptomic samples for each coral colonies. . . . .	57
I.41	<i>Tara</i> Pacific expedition map with islands sequenced in metatranscriptomic during the thesis indicated in pink. . . . .	58
II.1	Phylogenetic tree of the taxonomic lineages selected, to summarize the taxonomic affiliation results. . . . .	68
II.2	Corals species sampled during the <i>Tara</i> Pacific expedition. . . . .	69
II.3	The <i>Tara</i> Pacific sampling plan for each island and dual-transcriptomic samples using the three different corals. . . . .	70
II.4	The <i>Tara</i> Pacific expedition map with islands sequenced in metatranscriptomic during the thesis (indicated in pink). . . . .	70
II.5	Workflow developed by the LBGB team at the Genoscope and applied on dual-transcriptomic samples from the <i>Tara</i> Pacific expedition. . . . .	73
II.6	Taxonomic affiliation of transcripts from the 28 Symbiodiniaceae reference transcriptomes selected, using Uniref90 database. . . . .	80
II.7	Taxonomic affiliation of transcripts from the 28 Symbiodiniaceae reference transcriptomes, using a database created at the Genoscope. . . . .	81
II.8	Taxonomic affiliation of CDS from 6 Symbiodiniaceae reference genomes, using a database created at the Genoscope. . . . .	82
II.9	Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes. . . . .	84
II.10	Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes. . . . .	85
II.11	Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes. . . . .	86
II.12	Barplots of the number of reads aligned on specific Symbiodiniaceae transcriptome and genome references. . . . .	89

II.13	Barplots of the number of reads aligned on specific Symbiodiniaceae transcriptome and genome references. . . . .	90
II.14	Barplots of the number of reads aligned on specific Symbiodiniaceae transcriptome and genome references. . . . .	91
II.15	Relative abundance of Symbiodiniaceae ITS2 profiles in <i>Pocillopora sp.</i> samples. . . . .	94
II.16	Comparisons of some ITS2 nucleotide sequences of different <i>Cladocopium C1</i> obtained in ITS2 profiles analyses. . . . .	95
II.17	Barplots of the mean coverage by contig (A) and CDS (B) from <i>Cladocopium C1</i> reference genome and CDS aligned on dual-genomic (A) and dual-transcriptomic (B) samples of <i>Pocillopora sp.</i> . . . . .	97
II.18	Hierarchical clustering of <i>Cladocopium C1</i> populations in metatranscriptomic samples of <i>Pocillopora sp.</i> . . . . .	100
III.1	Schema of the two methods applied on dual-transcriptomic samples from the <i>Tara Pacific</i> expedition, to define biological replicates in different environmental conditions. . . . .	109
III.2	PCA of gene expression normalized in TPM on each dual-transcriptomic samples. . . . .	117
III.3	PCA of gene expression normalized in TPM on each dual-transcriptomic samples. . . . .	118
III.4	PCA of gene expression normalized in TPM on each dual-transcriptomic samples. . . . .	119
III.5	Pearson correlations on gene expression levels normalized in TPM on <i>Symbiodinium A3</i> in <i>Millepora sp.</i> . . . . .	121
III.6	Pearson correlations on gene expression levels normalized in TPM on <i>Cladocopium C15</i> in <i>Porites sp.</i> . . . . .	122
III.7	Pearson correlations on gene expression levels normalized in TPM on <i>Cladocopium C1</i> in <i>Pocillopora sp.</i> . . . . .	123
III.8	Number of differentially expressed genes in <i>Symbiodinium A3</i> , from <i>Millepora sp.</i> samples. . . . .	124
III.9	Number of differentially expressed genes in <i>Cladocopium C15</i> , from <i>Porites sp.</i> samples. . . . .	126
III.10	Number of differentially expressed genes in <i>Cladocopium C1</i> , from <i>Pocillopora sp.</i> . . . . .	128
III.11	CCA of gene expression from DEG normalized in TPM. . . . .	130
III.12	CCA of gene expression from DEG normalized in TPM. . . . .	131
III.13	CCA of gene expression from DEG normalized in TPM. . . . .	132
III.14	Number of DEG between each <i>Cladocopium C1</i> lineages defined, in <i>Pocillopora sp.</i> samples. . . . .	133
III.15	The most enriched PFAM2GO in DEG of <i>Symbiodinium A3</i> , after selection of PFAM involved in photosynthesis metabolic pathway for down-regulated genes in Gambier compared to Moorea. . . . .	137
III.16	The most enriched PFAM2GO in DEG of <i>Symbiodinium A3</i> , after selection of PFAM involved in photosynthesis metabolic pathway for Down-regulated genes in Moorea compared to Aitutaki. . . . .	138
III.17	KEGG enriched in <i>Cladocopium C1</i> genome, using iPath3 visualization. . . . .	139
IV.1	The <i>Tara Pacific</i> sampling plan for coral surrounding water samples in each sampling site. . . . .	191
IV.2	Barplots of the number of CSW reads aligned on Symbiodiniaceae reference transcriptomes. . . . .	197
IV.3	Barplots of the number of CSW (A) and dual-transcriptomic (B) reads aligned on specific Symbiodiniaceae references. . . . .	198
IV.4	Barplots of the number of CSW reads aligned on <i>Pocillopora meandrina.</i> reference genome. . . . .	199
IV.5	Barplots of the number of Surface samples reads aligned on Symbiodiniaceae reference transcriptomes. . . . .	201



IV.6	Hierarchical clustering of <i>Cladocopium C1</i> populations in CSW samples from metatranscriptomic reads. . . . .	203
IV.7	PCA of gene expression normalized in TPM of each CSW samples. . . . .	205
IV.8	Pearson correlations on gene expression normalized in TPM of <i>Cladocopium C1</i> in CSW samples. . . . .	206
IV.9	Number of differentially expressed genes in <i>Cladocopium C1</i> , from the CSW samples, using island as biological replicates. . . . .	207
IV.10	Number of differentially expressed genes in <i>Cladocopium C1</i> , using samples from the same island and environment as biological replicates. . . . .	208
IV.11	Number of differentially expressed genes of <i>Cladocopium C1</i> taking in account the genetic and the environment, in and out of <i>Pocillopora sp.</i> colonies. . . . .	210
1	Paired-End sequencing protocol applied on polyA RNA sequences from coral samples of the <i>Tara</i> Pacific expedition. . . . .	243
2	GC content of each Symbiodiniaceae reference transcriptomes selected and studied during the thesis. . . . .	245
3	GC content of each Symbiodiniaceae reference genomes selected and studied during the thesis. . . . .	246
4	GC Content of 19 Symbiodiniaceae reference transcriptomes shared and studied with <b>Gonzales et al. in 2017</b> . . . . .	247
5	Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes filtered at 70% identity and 30% coverage. . . . .	249
6	Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes filtered at 70% identity and 30% coverage. . . . .	250
7	Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes filtered at 70% identity and 30% coverage. . . . .	251
8	Boxplots of mean identity by contig from <i>Cladocopium C1</i> reference genome aligned on dual-genomic samples from <i>Pocillopora</i> samples on the first 11 islands. . . . .	258
9	KEGG enriched in each <i>Cladocopium C1</i> lineages aligned on dual-transcriptomic samples, using iPath3 visualization. . . . .	261

# List of Tables

II.1	Symbiodiniaceae transcriptomes selected and used to analyse the <i>Tara</i> Pacific samples.	64
II.2	Symbiodiniaceae genomes selected and used to analyse the <i>Tara</i> Pacific samples. . . .	65
II.3	Number of dual-transcriptomic samples per island, site and coral sequenced at the Genoscope, using in the thesis. . . . .	73
II.4	Number of metagenomic samples per island and site from <i>Pocillopora sp.</i> samples sequenced at the Genoscope. . . . .	76
II.5	Number of remaining SNPs and number of genes covered by these SNPs in the different filtering steps of the GATK pipeline. . . . .	98
III.1	Table used to perform Fisher test applied on PFAM on the different datasets . . . . .	113
III.2	Number of DEG associated to GO terms involved in transport of nutrients, selected in <i>Symbiodinium A3</i> . . . . .	139
IV.1	Number of metatranscriptomic samples per island, site and environment from/around <i>Pocillopora sp.</i> colonies, sequenced at the Genoscope. . . . .	192
IV.2	Number of remaining SNPs and number of genes covered by these SNPs in the different filtering steps of GATK pipeline for population structure of <i>Cladocopium C1</i> from CSW samples. . . . .	202
1	Some calculations on <i>Millepora sp.</i> reads aligned on Symbiodiniaceae reference transcriptomes. . . . .	252
2	Some calculations on <i>Porites sp.</i> reads aligned on Symbiodiniaceae reference transcriptomes. . . . .	253
3	Some calculations on <i>Pocillopora sp.</i> reads aligned on Symbiodiniaceae reference transcriptomes. . . . .	254
4	Some calculs on dual-transcriptomic reads aligned on specific Symbiodiniaceae reference transcriptomes or genomes. . . . .	255
5	Number of variations and type of mutations in different <i>Cladocopium</i> ITS2 sequences .	256
6	Comparison bewteen ITS2 profiles and population structure of <i>Pocillopora sp.</i> and <i>Cladocopium C1</i> , obtained using dual-genomic and dual-transcriptomic respectively, in each sample . . . . .	259
7	PFAM enriched in DEG using genetic of <i>Cladocopium C1</i> and Fisher test exact .	262



# Abbreviations

- **BLAST**: Basic Local Alignment Search Tool
- **BP**: Biological Process
- **BWA**: Burrows-Wheeler Aligner
- **CC**: Cellular Component
- **CCA**: Canonical Correlation Analysis
- **CDS**: CoDing Sequence
- **CEA**: Commissariat à l'Énergie Atomique
- **CNRS**: Centre National de la Recherche Scientifique
- **COI**: Cytochrome Oxidase I
- **CSW**: Coral Surrounding Water samples
- **DEG**: Differentially Expressed Genes
- **DET**: Differentially Expressed Transcripts
- **DNA**: DesoxyriboNucleic Acid
- **DUALG**: Dual-Genomic
- **DUALT**: Dual-Transcriptomic
- **EMBL**: European Molecular Biology Laboratory
- **GATK**: Genome Analysis ToolKit
- **GO**: Gene Ontology
- **HMM**: Hidden Markov Models
- **iPath3**: interactive Pathways explorer 3
- **ITS2**: Internal Transcribed Spacer 2
- **IUPAC**: the International Union of Pure and Applied Chemistry
- **KEGG**: Kyoto Encyclopedia of Genes and Genomes
- **LAGE**: Laboratoire d'Analyses Génomiques des Eucaryotes

- **LBGB**: Laboratoire de Bio-informatique pour la Génomique et la Biodiversité
- **MF**: Molecular Functions
- **miRNA**: micro-RNA
- **MMA**: Micosporin like Amino Acid
- **MMETSP**: the Marine Microbial Eukaryote Transcriptome Sequencing Project (**Keeling et al. 2014**)
- **mRNA**: messenger RNA
- **MVIEW**: Multiple alignment VIEWer
- **NCBI**: National Center for Biotechnology Information
- **NGS**: Next Generation Sequencing
- **PCA**: Principal Component Analysis
- **Pfam**: Protein families
- **POC**: Particulate Organic Carbon
- **rDNA**: ribosomal DNA
- **RNA**: RiboNucleic Acid
- **RNA-Seq**: RNA Sequencing
- **ROS**: Reactive Oxygen Species
- **rRNA**: ribosomal RNA
- **SAR**: Stramenopiles, Alveolates and Rhizaria super group
- **smRNA**: small RNA
- **SNP**: Single Nucleotide Polymorphism
- **SRF**: SuRFace samples
- **TPM**: Transcripts Per kilobase Millions
- **UMR**: Unité Mixte de Recherches
- **U.S.**: United States
- **U.S.A.**: United States of America
- **U.V.**: Ultra-Violet
- **WGCNA**: WeiGhted Correlation Network Analysis





# Chapter I

## State of the art

### I.1 Ecology of coral reefs

#### I.1.1 Coral Taxonomy

A long time ago "coral" was described as a stone present in the marine environment. However, observations by fishermen of its capacity for growth, led to the attribution of a living origin to these marine organisms. Redescribed as petrified aquatic plants, they were subsequently identified as belonging to the animal realm because of their reactivity to human touch.

Today the word coral is a generic term used to describe organisms with a hard stony substance secreted, as an internal skeleton found in the *Cnidaria* phylum (**Figure I.1**). This phylum contains marine organisms commonly known as anemones, medusas and corals, which represent a large diversity of forms and lifestyles and encompass marine organisms that are swimming in the sea water or fixed on a solid support. In this branch, corals are found in the classes *Anthozoa* and *Hydrozoa*.

Christian Gottfried Ehrenberg described for the first time the class of *Anthozoa* in 1831 (**Cevasco G. A. et al. 1997**) and a few years later in 1865, the zoologist Addison Emery Verrill created the *Cnidaria* phylum, which groups the classes of *Anthozoa* and *Hydrozoa* where corals are represented.



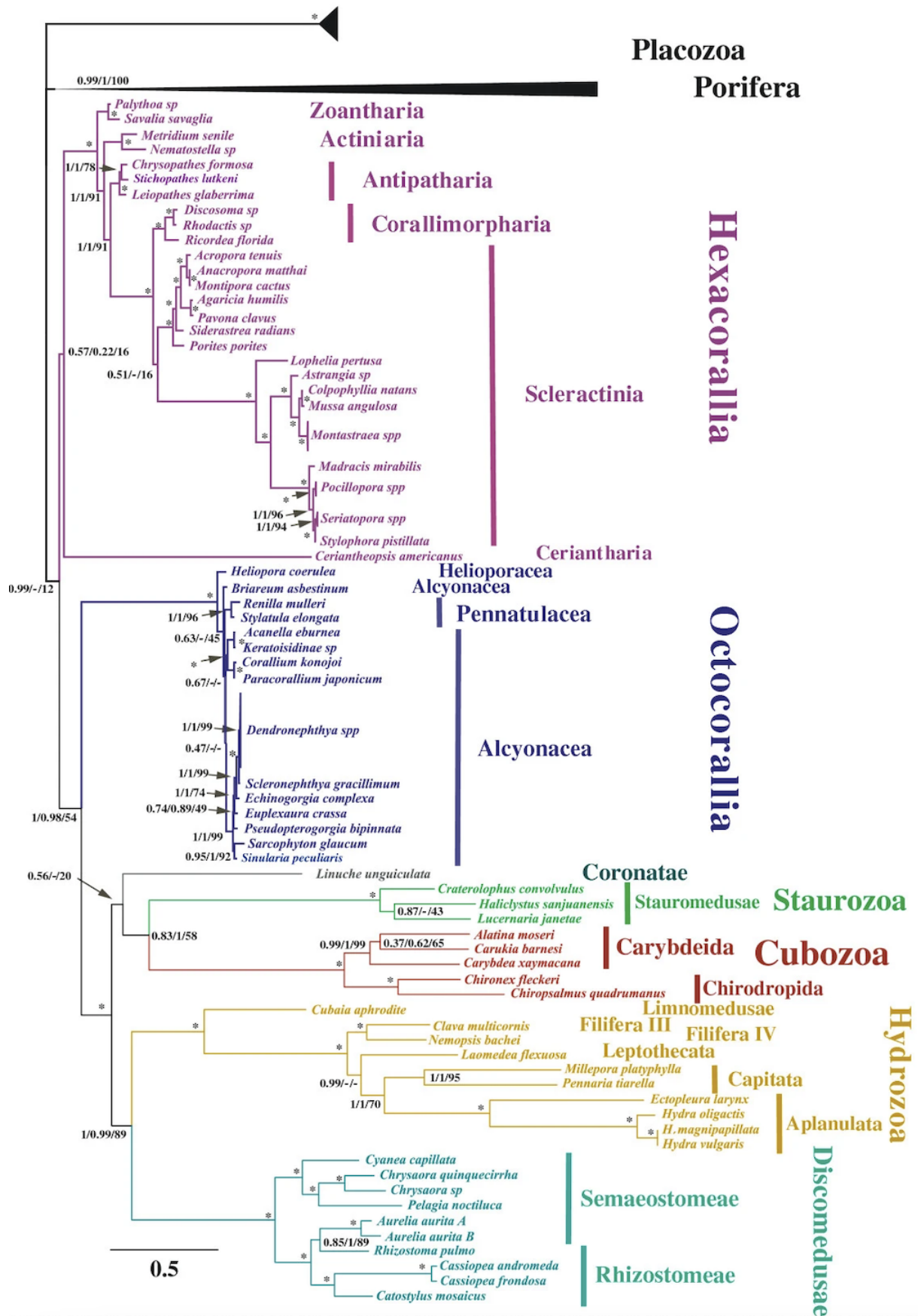


Figure I.1: Cnidarian phylogenetic tree based on mitochondrial protein genes (Kayal E. et al. 2013).

The *Anthozoa* class groups organisms that can be colonial, clonal or solitary and contains approximately 7.500 species (Daly M. *et al.* 2007). The first subclass within *Anthozoa*, the *Hexacorallia* consists of species which have a rigid mineralized skeleton and it is composed of a wide variety of marine organisms, with diverse forms and species. The second subclass of *Anthozoa* is *Octocorallaries* and in this class, corals are called "gorgonaceae" like *Corallium rubrum* and *Tubipora musica*.

The *Hydrozoa* class includes species more commonly known as "soft corals". This class contains approximately 3.500 species and it is distinct from the other classes because the hydrozoans have gametogenic tissue with an ectodermal origin, compared to the other classes which have gametogenic tissue with endodermal origin (Daly M. *et al.* 2007).

## I.1.2 Coral life cycle and morphological diversity

### I.1.2.1 Coral life cycle

Corals have a generally sedentary mode of life and tend to form colonies composed of interconnected individual polyps (Figure I.2). During the polyp phase, the coral feeds on *plankton* to grow, relying on a variety of prey to meet its physiological needs (Ferrier-Pagès *et al.* 2003).

Coral colonies exhibit different modes of reproduction: most of colonies are hermaphroditic and the remainder are gonochoric with separate male and female colonies (Bikerland *et al.* 1996). Corals can also reproduce asexually, by fragmentation or by generation of a larva from the parent colony. This mode of reproduction can result in reduced diversity of coral populations (Mollica *et al.* 2018).

Coral colonies emit gametes in the marine environment, then male and female gametes merged to constitute a coral zygote. After multiple cell divisions the zygote develops into a free-swimming planula larva, that can live between 3-21 days in the marine environment

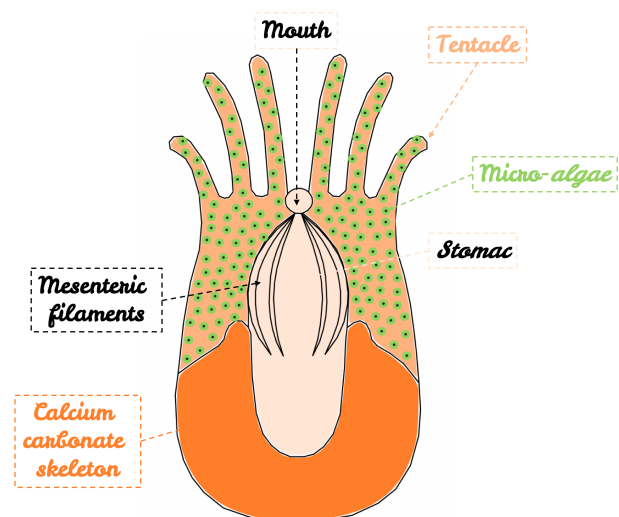


Figure I.2: Polyp schema of coral.

and moving with the currents. This allows the colonization of new areas, from a few meters to miles away from the site of gamete release (**Birkeland et al. 1996**). Then, the larva affixes itself to the substrate (i.e. settles) forming a polyp which will bud and then build a new coral colony (**Figure I.3**).

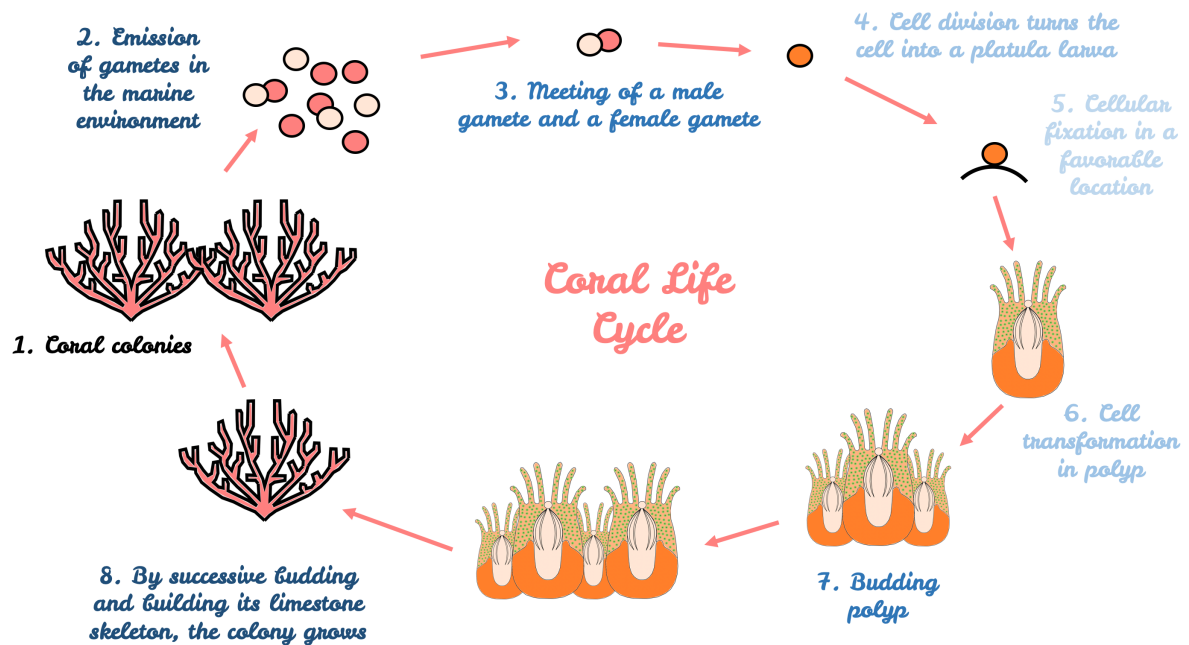


Figure I.3: Coral life cycle schema, inspired by **Thompson J. et al. 2015**.

Coral colonies emit gametes in the marine environment (2). Then, the male gamete and female gamete meet and constitute the first coral cell (3). After cell division, the cell turns into a planula larva (4) and will settle down in a favorable location (5). After multiple divisions, cells constitute a polyp (6), which will bud (7). Then, by successive budding and building its limestone skeleton polyps constitute a coral colony (8).

Reproduction is often linked to favorable environmental conditions like the abundance of sunlight, the space available for colonization and also by ecological features like the absence of coral predators such as sea urchins, some fishes and crown-of-thorns starfishes (**Mollica et al. 2018**).

There exist two distinct processes of coral skeletal growth which are extension (upward growth) and densification (lateral thickening), both of which depend on environmental conditions (**Mollica et al. 2018**).

Optimal growth and reproduction also depend on the water temperatures, that can vary depending on the geographical location and among the coral species. For example, optimal temperatures should be between 22°C and 26°C in the Pacific Ocean, and below a water temperature of 18°C, growth is severely affected in this ocean. There is also a low survival rate

of corals with water temperature above 30°C, except in some areas like the Red Sea. Coral adaptation to temperature is linked to their respective environment. Indeed, the same species of coral in a tropical region will be more resistant than its counterpart in a sub-tropical region (Dubinsky, 1990).

Coral growth depends also on seawater salt content and their growth is optimal in salinity conditions between 32-36%. A low salinity of 20‰ limits their growth and reproduction, and the same is true for a higher salinity beyond 38‰ (Dubinsky in 1990).

### I.1.2.2 Coral morphological diversity

There is a very high morphological diversity among coral species, which is characterized by diverse forms and colors of corals. This is true even among a single genus. For example, the *Millepora* genus of the *Hydrozoa* class contains many species with diverse forms such as dense branching (A), spaced branching (B), fan-like branching (C), encrusting morphology (D), plate-like morphology (E) and branching morphology (F) (Figure I.4).

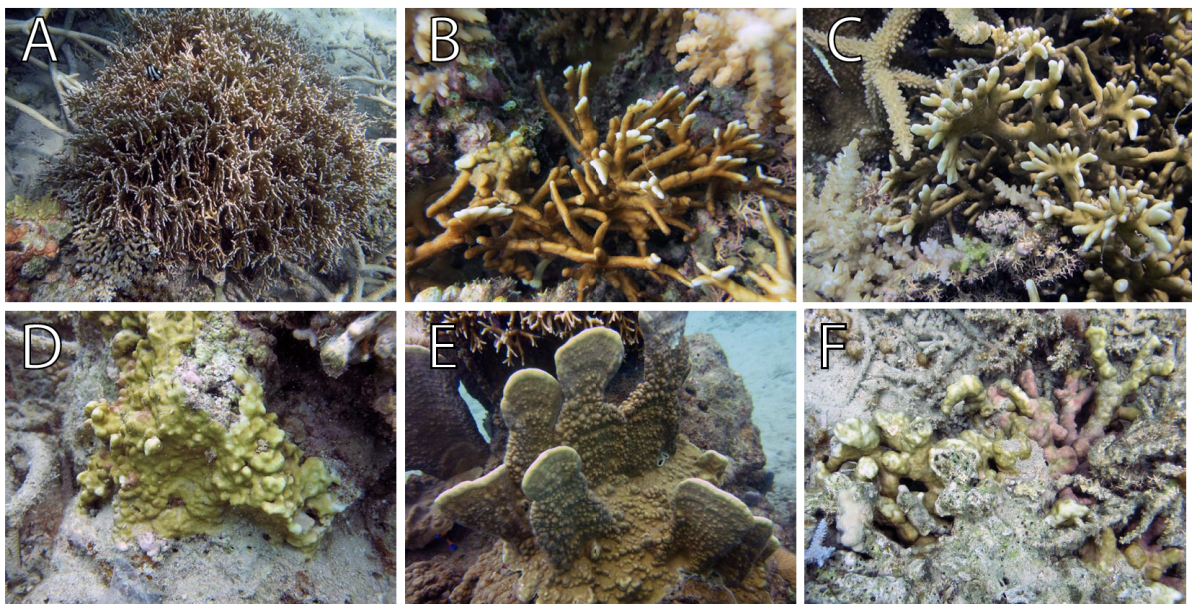


Figure I.4: *In situ* images of several *Millepora* specimens examined (Takama *et al.* 2018). (A) *Millepora intricata*; (B) *Millepora dichotoma*; (C) *Millepora intricata*; (D) *Millepora exaesa*; (E) *Millepora platyphylla*; (F) *Millepora tenera*.

The identification of corals at the species level can also be challenging because they present both intra-species dissimilarities and inter-species similarities (Huang *et al.* 2015). For example, pictures of three coral species of *Pocillopora* genus collected from Hawaii illustrate this

challenge (**Figure I.5**). These pictures show how the same coral species can be similar or different visually, and also confirm the difficulties of taxonomically identifying coral to the species level solely from a morphological point of view (**Johnston et al. 2018**). Whereas the identification of different genera can be accomplished using morphology, distinguishing between species remains a challenge.

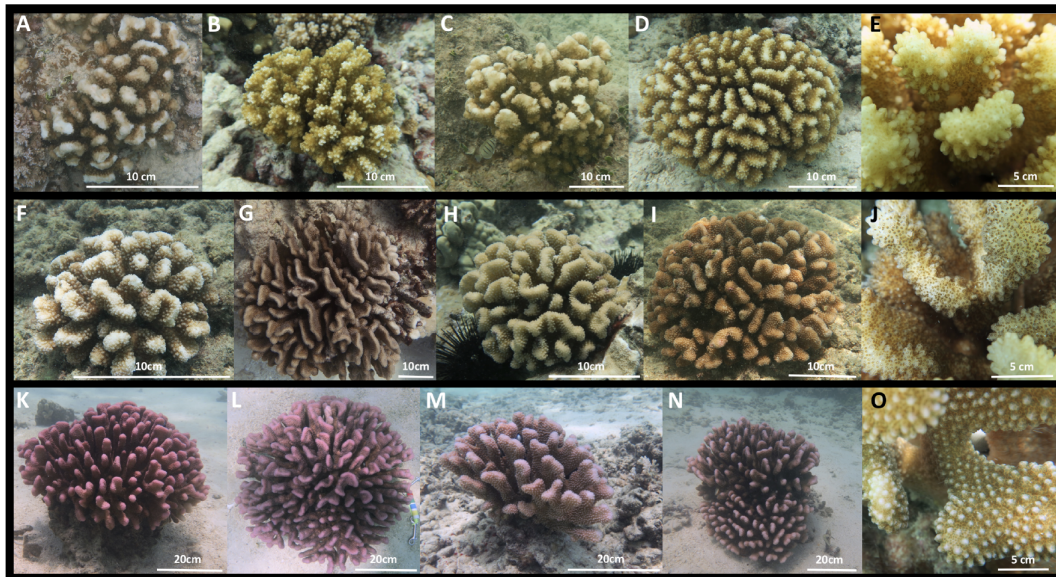


Figure I.5: Images of *Pocillopora ligulata* colonies, **A-E**; *P. meandrina* colonies, **F-J**; and *P. eydouxi* colonies, **K-O** from O'ahu, Hawai'i from **Johnston et al. 2018**.

### I.1.3 Geographic distribution of coral reefs

Coral reefs cover around 0.1% of the planet (**Porter et al. 2001**) and are scattered across oceans and seas (**Figure I.6**). The highest coral species richness is found in the Indo-Pacific Ocean, called the "coral triangle", which represents the oldest and richest geographic region in terms of coral species richness (**Huang et al. 2015**).

There exist four main reef types which present high diversity in terms of constitution, geographical location and also characteristics (**Figure I.7**). There are: platform reefs, fringing reefs, barrier reefs and atolls.

These different coral reefs are constructed after many years of geologic movements (**Figure I.8**). In fact, the different types of reefs originate in most cases from the formation of a volcanic island, which after many years transforms successively into a fringing reef, followed by a barrier reef and then an atoll.

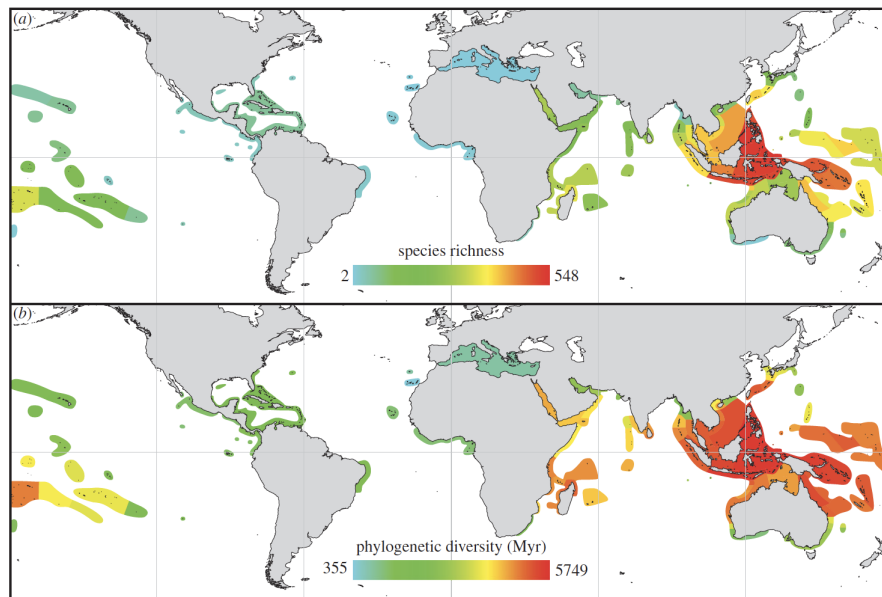


Figure I.6: Global diversity of coral reefs at the scale of 141 reef eco-regions (Huang *et al.* 2015). (a) Species richness (b) Phylogenetic diversity in millions of years (Myr), ranked to facilitate visualization of the colour gradient.

Platform reefs	Fringing reefs	Barrier reefs	Atolls
Frequently found in the lagoons created by atolls and barrier reefs	Closely follow shorelines, narrow shallow lagoon	Separated from land by a relatively wide, deep lagoon	Horseshoe shaped or circular reef surrounding a central lagoon (often far from land in the open ocean)
In the Great Barrier Reef lagoon, Belize, Red Sea, Bahamas	Red sea, East Africa, Seychelles and other Indo-Pacific islands, most Caribbean reefs	The Great Barrier Reef in Australia, Belize Barrier Reef, off Mayotte in the Western Indian Ocean	>95% of the atolls are in the Indo-Pacific, others are found outside Belize and in Western Atlantic

Figure I.7: The four main coral reef types with their descriptions and some examples of their locations (Moberg *et al.* 1999).

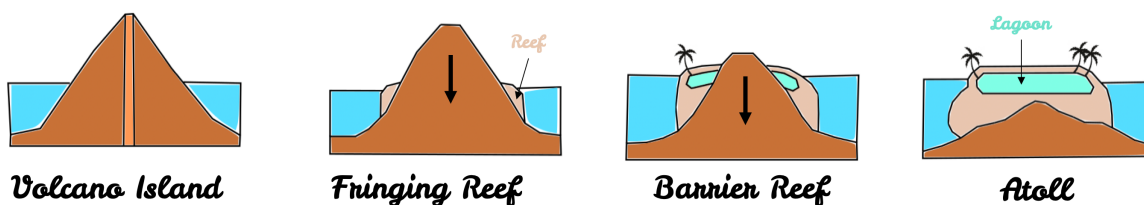


Figure I.8: Schema of the different coral reef types, inspired from Darwin *et al.* 1842. After the extinction of the volcano, a fringing reef is formed around the inactive cone. After some time and with the movement of the tectonic plates, the remains of the volcano form an island whose erosion and sinking into the oceanic crust forms the barrier reefs. After submersion of the remains of erosion it forms an atoll forms.

### I.1.4 Biodiversity of coral reefs

Coral reefs are the base of the food chain in the marine environment (**Figure I.9**). They serve as a refuge for marine organisms like fishes and crustaceans, which feed on zooplankton, who themselves feed on photosynthetic organisms known as phytoplankton, and feed predators in turn. All of these organisms are found in the coral reef ecosystem.

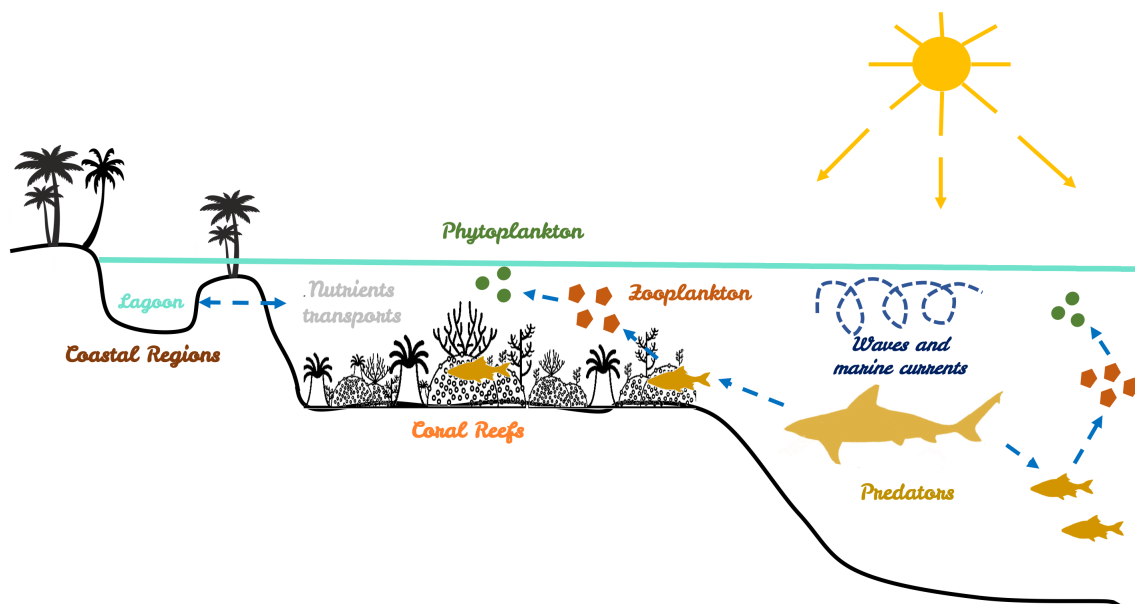
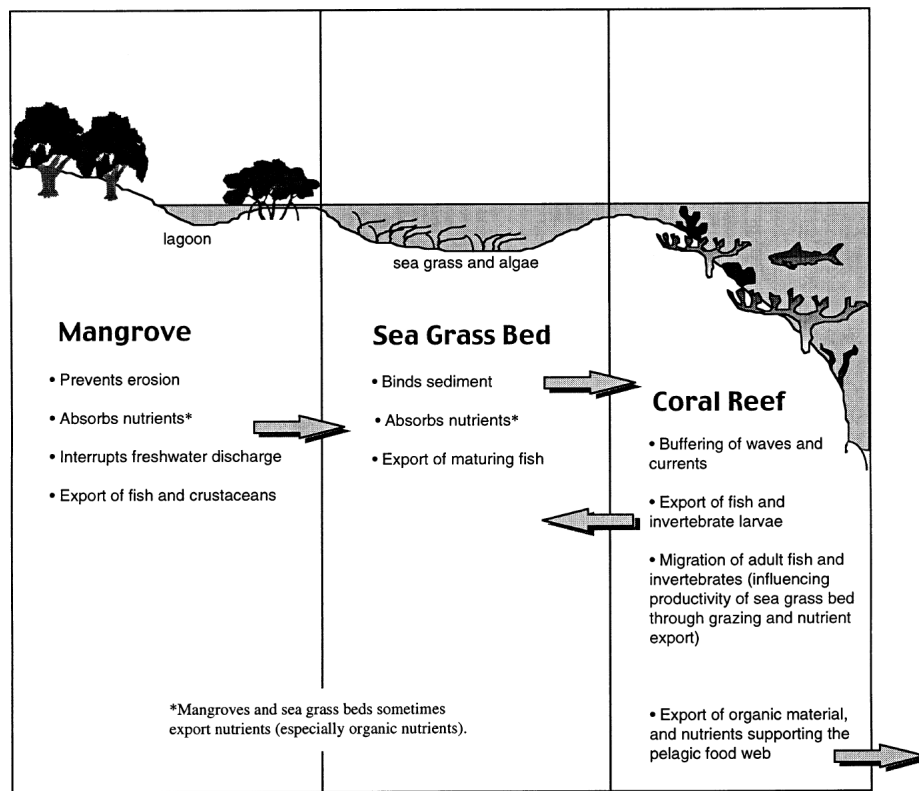


Figure I.9: Coral reefs as the base of the marine food alimentary chain.

From their constitution, coral reefs harbor a phylogenetic biodiversity that has no terrestrial equivalent, with more than 34 different classes of various marine organisms living in coral reefs (**Porter et al. 2001**). Through this wealth of organisms, they constitute an important reservoir of marine biodiversity, which is essential for the survival of many marine species.

Reef ecosystems interact with other marine ecosystems such as seagrass meadows, mangroves and open-sea environments (**Figure I.10**). They export different organisms such as larvae and micro-organisms like copepods, as well as nutrients. Nutrients can be both exported and imported from coral reef to seagrass beds and/or mangrove habitats. It appears that coral reefs have an ecological impact on marine biodiversity and are essential to its maintenance (**Moberg et al. 1999**).



Adopted from White (1986) and Ogden (1988).

Figure I.10: Interactions in the tropical seascape, showing the connections between mangroves, sea-grass meadows and coral reefs (Moberg *et al.* 1999).



### I.1.5 Anthropogenic importance of coral reefs

Coral reefs have core anthropogenic importances, for example they act as barriers to protect coasts against different meteorological events (**Moberg et al. 1999**). Indeed, their location near to coastal regions (**Figure I.9**) turns them into a wall of protection and they therefore attenuate sea currents and protect coastal regions against storm damage.

They have a considerable impact on the economy of many countries which are nearby coral reefs (**Figure I.11**). Indeed, they allowed the development of a touristic industry representing an important resource for many countries, from \$ 451.2 Millions (USD) per year in the Philippines to \$ 1657.4 Millions in Mexico. Other countries around the world like Egypt, Indonesia, Mexico, Thailand, Australia, China, the Philippines, the U.S.A. and Japan are interesting examples of regions economically relying on tourism linked to coral reefs.

	Proportion of tourism which is coastal, non-urban	Reef-adjacent tourism value (Million US\$ per year)	Proportion of reef-coast tourism assigned as on-reef	On-reef tourism value (Million US\$ per year)	Total tourism value (Million US\$ per year)
Egypt	44%	947.5	53%	4519.8	5467.3
Indonesia	29%	1106.2	20%	1991.2	3097.5
Mexico	27%	1657.4	9%	1342.5	2999.9
Thailand	34%	1331.6	9%	1078.6	2410.2
Australia	24%	473.1	40%	1703.0	2176.1
China	13%	1348.1	2%	88.1	1435.1
Philippines	30%	451.2	23%	934.0	1385.1
USA (Hawaii)	58%	680.1	9%	550.8	1230.9
Japan	10%	542.6	13%	634.9	1177.5
USA (Florida)	11%	850.6	4%	306.2	1156.8

Figure I.11: Estimated expenditures by tourists within the ten jurisdictions with the highest total values. (**Spalding et al. 2017**).

Through their biodiversity, coral reefs are also the main source of food for local populations living near coral reefs. It is therefore vital for these populations to preserve this ecosystem (**Hawkins et al. 2004**).

### I.1.6 Coral reefs threats

#### I.1.6.1 Bleaching events

Coral reefs are under multiple threats since the two past decades, due to global environmental changes and local threats. Several of these threats induce a distinctive coral loss of color named coral bleaching. This phenomenon indicates that the coral is under stress, and this coral bleaching was observed all around coastal regions in different coral reefs (**Figure I.12**). This phenomenon can affect small or large coral colonies, as well as an entire coral reef or only a part of it.

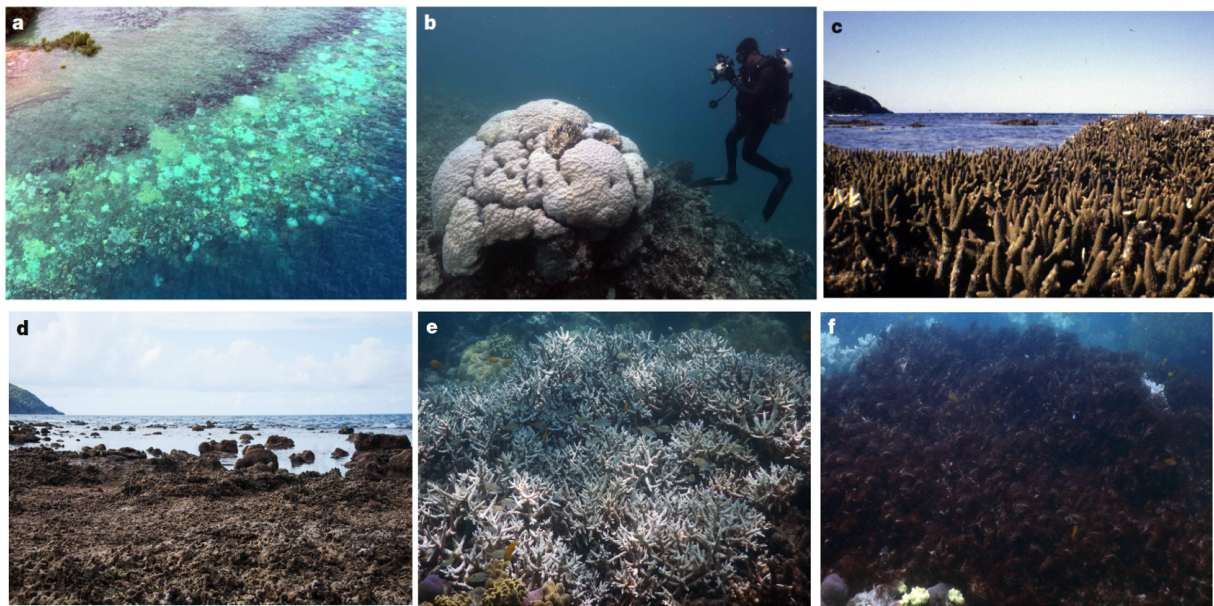


Figure I.12: Recurrent severe coral bleaching (Hughes *et al.* 2017).

**a:** Severe bleaching in Princess Charlotte Bay, North-East Australia, in March 2016. **b:** Severe bleaching in 2016 on the northern Great Barrier Reef. **c:** Large, old beds of corals, on Orpheus Island, Queensland photographed in 1997. **d:** Eighteen years later in May 2016, corals at this site have never recovered. **e & f:** Mature stands of corals after heat stress and colonized by algae over a period of just a few weeks in 2016 on Lizard Island, Great Barrier Reef, Before (**e**) and After (**f**).

Globally, these coral bleaching events affect all oceans (Figure I.13) and in 2015 and in 2016, these events affected more than 30% of the reefs across all the oceans.

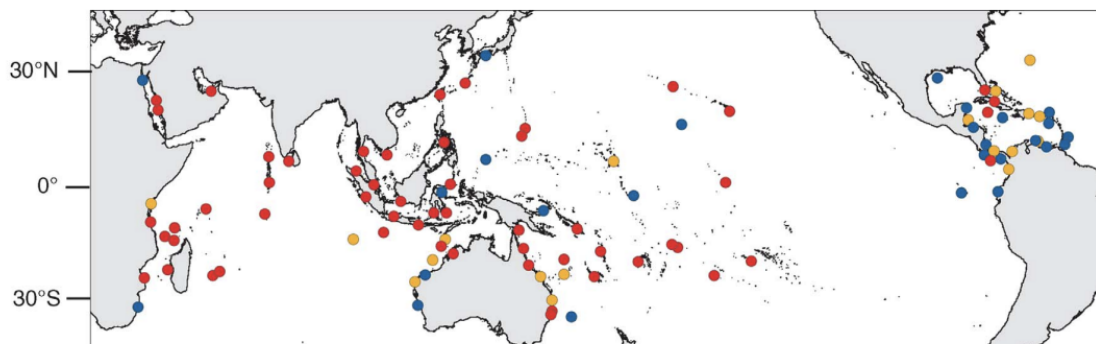


Figure I.13: The global extent of mass bleaching events of coral reefs in 2015 and in 2016 (Hughes *et al.* 2018).

Symbols show 100 reef locations that were assessed: red circles, severe bleaching affecting >30% corals; orange circles, moderate bleaching affecting <30% corals; and blue circles, no substantial bleaching recorded.

### I.1.6.2 A multitude of threats

Global warming is the major cause of coral bleaching events. A thermal stress of  $+1^{\circ}\text{C}$  above the highest monthly mean sea surface temperature is enough to initiate the bleaching process (**Kumagai *et al.* 2018**). This threat correlates with anthropogenic activities, which induce disturbances in the environment (**Hughes *et al.* 2003; Hughes *et al.* 2017**).

Acidification of the oceans, which is caused by anthropogenic carbon emissions, is the second major cause of coral death. In fact, anthropogenic carbon emissions in the atmosphere is partially absorbed, and then dissolved in the oceans inducing acidification. The main consequences for corals are the decrease of skeleton calcification through inhibition of calcium carbonate deposition (**Phinney *et al.* 2006**).

Other coral mortality causes are observed in different regions and an anthropogenic origin has been identified for most of them (**Fabricius *et al.* 2005 and Figure I.14**). These local pollutants are found throughout the different regions near the coral reefs and exacerbate local stress on coral reefs.

Added to local pollution, there is also coastal urbanisation and sedimentation that disrupt this fragile ecosystem (**Figure I.14**).

Overfishing and excessive nutrient input, interact with increasing temperatures to disrupt coral reefs down to microbial scales. Combined with elevated temperatures these effects are exacerbated and can result in coral death by weakening and exposing them to increased predation and also to bacterial opportunism (**Zaneveld *et al.* 2016**).

The consequences of all these different threats include less diverse communities of reefs and difficulties in maintaining reef structures (**Hoegh-Guldberg *et al.* 2006; Hoegh-Guldberg *et al.* 2007**). However, some coral species are more tolerant than others to climate changes and coral bleaching events (**Hughes *et al.* 2003**). In fact, some coral species can fully recover after the return of favorable environmental conditions (**Hughes *et al.* 2017**).

Location	Agent	Response	Source
Northern Gulf of Aqaba (Eilat), Red Sea	50% increase in nutrients from floating fish farms	50% coral mortality from benthic algal blooms, 3–4 fold reduced reef calcification, 50% increased P/R ratio	Loya (2004)
	Sewage discharge, spillage of phosphate dust	Increased algal growth trapping sediment; fourfold increased mortality in <i>Stylophora pistillata</i> , possibly from reduced light, inhibition of calcification, and increased sedimentation	Walker and Ormond (1982)
Reunion Island, Indian Ocean	Coastal urbanisation, groundwater enriched with nutrients from untreated sewage	Higher coral cover, coral diversity, fish diversity and density of sea urchins, and lower macroalgal density on reefs away from nutrient enrichment and in the 1970s before nutrient enrichment, than on nutrient-enriched reefs. High bioerosion, calcification slower than reef erosion on nutrient enriched reefs	Cuet et al. (1988), Montaggioni et al. (1993), Naim (1993) and Chazottes et al. (2002)
Hong Kong	Excess pollutants, nutrients, sediment dredging	Low coral recruitment, few zooxanthellate octocorals, disappearance of giant clams ( <i>Tridacna</i> spp.), high bioerosion	Morton (1994) and Hodgson and Yau (1997)
Japan	Eutrophication and sedimentation	Declining coral cover	Shimoda et al. (1998)
	Gradients away from rivers	Change in coral community composition away from source	West and Van Woesik (2001)
Philippines	Excess sedimentation from logging	Declining coral cover, declining biodiversity due to disappearance of sediment-sensitive species over 12 months, inhibition of coral settlement	Hodgson (1990a) and Hodgson and Walton Smith (1993)
Indonesia	Excess nutrients and sedimentation	Low coral cover, reduced coral diversity, unaltered vertical extension but low skeletal density in massive corals, increased bioerosion	Edinger et al. (2000), Tomascik et al. (1997), Edinger et al. (1998) and Holmes et al. (2000)
Great Barrier Reef	Gradient in nutrients and turbidity	Increased macroalgal cover and richness (esp. red and green macroalgae), reduced octocoral richness	Fabricius et al. (in press) and Fabricius and De'ath (2004)
	Gradient away from river	Reduced coral cover, richness; increased filter feeders and macroalgae near source	van Woesik et al. (1999)
	Turbidity	Decreasing richness of zooxanthellate octocorals	Fabricius and De'ath (2001b)
	Inshore-offshore gradient, terrestrial runoff Sedimentation gradient	Increasing density of internal macrobioeroders towards the coast Decreasing cover of crustose coralline algae	Hutchings et al. (in press) Fabricius and De'ath (2001a)
Kanehoe Bay, Hawaii	Nutrients	Reduced coral cover, increased filter feeders, increased macroalgal cover	Smith et al. (1981), Hunter and Evans (1995), Stimson and Larned (2000) and Stimson et al. (2001)
Barbados	Eutrophication gradient	Photosynthetic pigments increase with increasing nutrient enrichment. Convex modal responses in gross photosynthesis, respiration, linear extension, calcification (enhanced by nutrients, depressed by turbidity)	Marubini (1996), Tomascik and Sander (1985) and Tomascik (1990)
		Reduced species diversity, probably due to differences in sediment rejection abilities, combined with feeding and reproductive strategies, altered community structure; increased bioerosion in coral rubble	Tomascik and Sander (1987b) and Holmes (2000)
		Reduced gamete formation, larval development and settlement, reduced recruit and juvenile density and diversity, juveniles larger, increased juvenile mortality	Tomascik and Sander (1987a), Tomascik (1991), Hunte and Wittenberg (1992) and Wittenberg and Hunte (1992)
Grand Cayman Island	Untreated fecal sewage, sixfold increased bacterial biomass	Fivefold increased internal bioerosion by the boring sponge <i>Cliona delitrix</i>	Rose and Risk (1985)
Costa Rica (2 sites)	Sedimentation	Low live coral cover, low species diversity, and large average colony diameters, high acid-insoluble residues incorporated in skeleton on exposed reef	Cortes and Risk (1985)
Brazil (2 sites)	Eutrophication	High macroalgal abundances, high density of heterotrophs	Costa Jr et al. (2000)

Figure I.14: List of some of the more comprehensively documented field assessments on the effects of enhanced terrestrial runoff and other forms of pollution, on the ecology of coral reefs (Fabricius *et al.* 2005)

### I.1.7 Coral studies

Today, most coral studies are focused on genetic data, because, as discussed previously, it is a challenge to identify coral species morphologically. For example, a recent study showed that coral species from the *Pocillopora* genus can change their form under different environmental conditions (**Paz-García et al. 2015 and Figure I.15**).

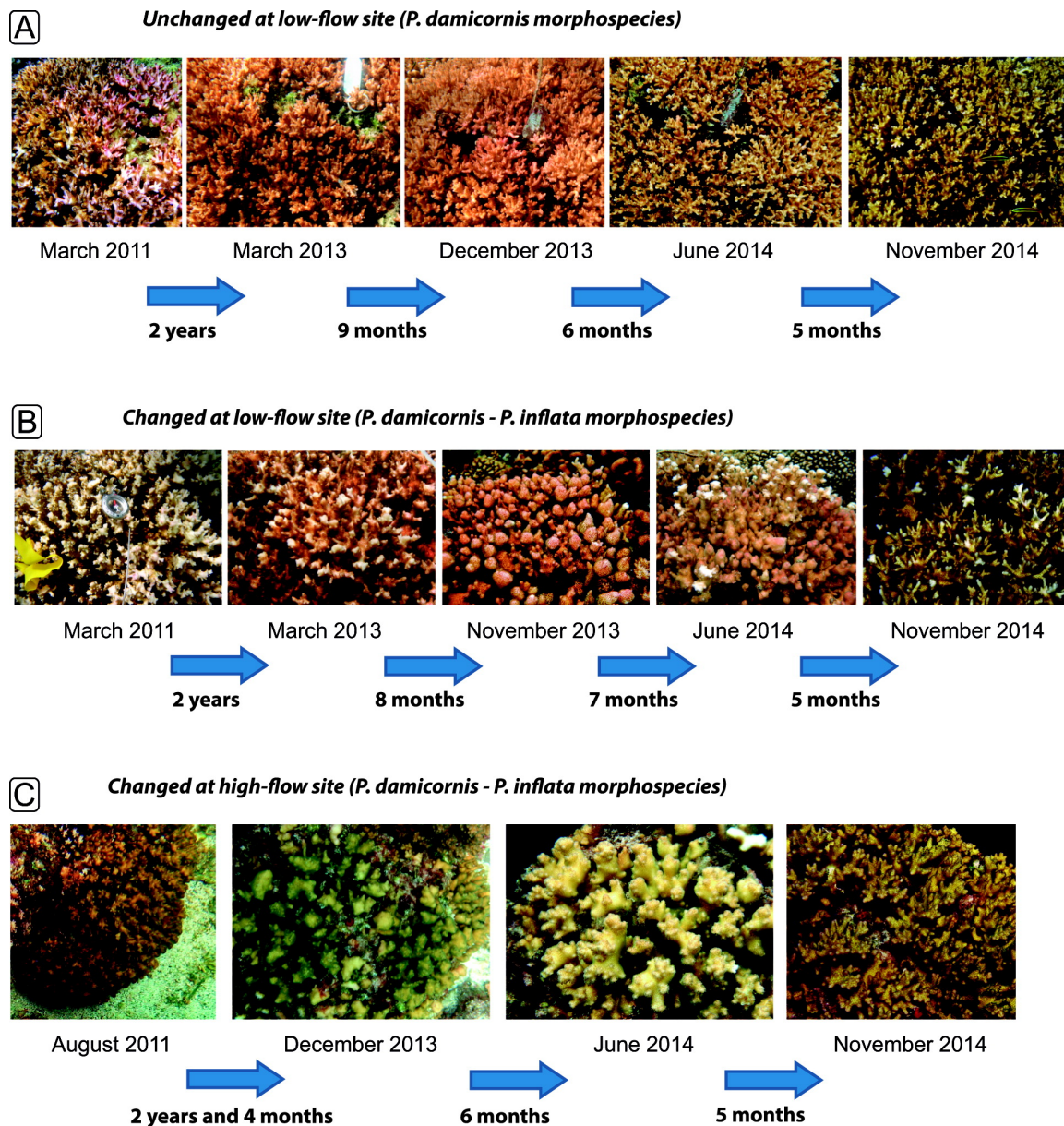


Figure I.15: View of unchanged (A) and changed (B, C) colonies at low-flow and high-flow sites south of the Gulf of California, March 2011–November 2014 (**Paz-García et al. 2015**)

*Pocillopora* corals can change morphologically after a few months or some years (**Figure I.15**). For example, during a 4 years period, *Pocillopora damicornis* in (A) remain unchanged at low flow site, whereas colonies of *Pocillopora inflata* showed altered growth forms at low flow site (B) and high flow site (C). This study illustrates the difficulties to identify coral species based on their morphology.

These problems of identification can be partially resolved by the sequencing of marker genes (**Johnston et al. 2018**). Indeed, the use of barcoding methods using marker genes has shed some light on phylogenetic points. The mitochondrial Cytochrome Oxidase I (COI) gene is frequently used to define coral colonies (**Johnston et al. 2018**).

However, it is difficult to find a discriminant marker gene to identify the different species of a coral genus. These marker gene sequences exhibit few variations between coral species (**Coffroth et al. 2008**).

Thus, the use of genetic methods and the observations of physical traits make it possible to obtain more details on the different species of corals. However, they remain incomplete for complex cases. That is why high-throughput sequencing methods are often used to sequence whole genomes and then infer population structure using different coral genome references.

In fact, sexual reproduction of the coral leads to its high dispersal capacity from one reef to another. These sequencing methods therefore make it possible to identify an individual at a given time and at a given place. For example, a maximum of individual sequenced and identified in one place can allow for a global vision of all the individuals in one reef.

An example of this kind of DNA analysis based on genomic samples was is shown below (**Figure I.16**). They studied *Pocillopora sp.* colonies across the Indo-Pacific Ocean and grouped different lineages of *Pocillopora sp.* in 4 Clades which are considered as a lineage.

Dispersal of Clade 1 coral larvae enables them to colonize a large region (Indo-Pacific), whereas in other reefs the dispersal of Clades 2, 3 and 4 are limited (Coastal regions of East Africa and Red Sea). DNA analysis has helped to recognize that many species of fish and invertebrates that were once thought to have widespread distributions are in fact made up of multiple distinct genetic lineages (**Knowlton et al. 2015**).

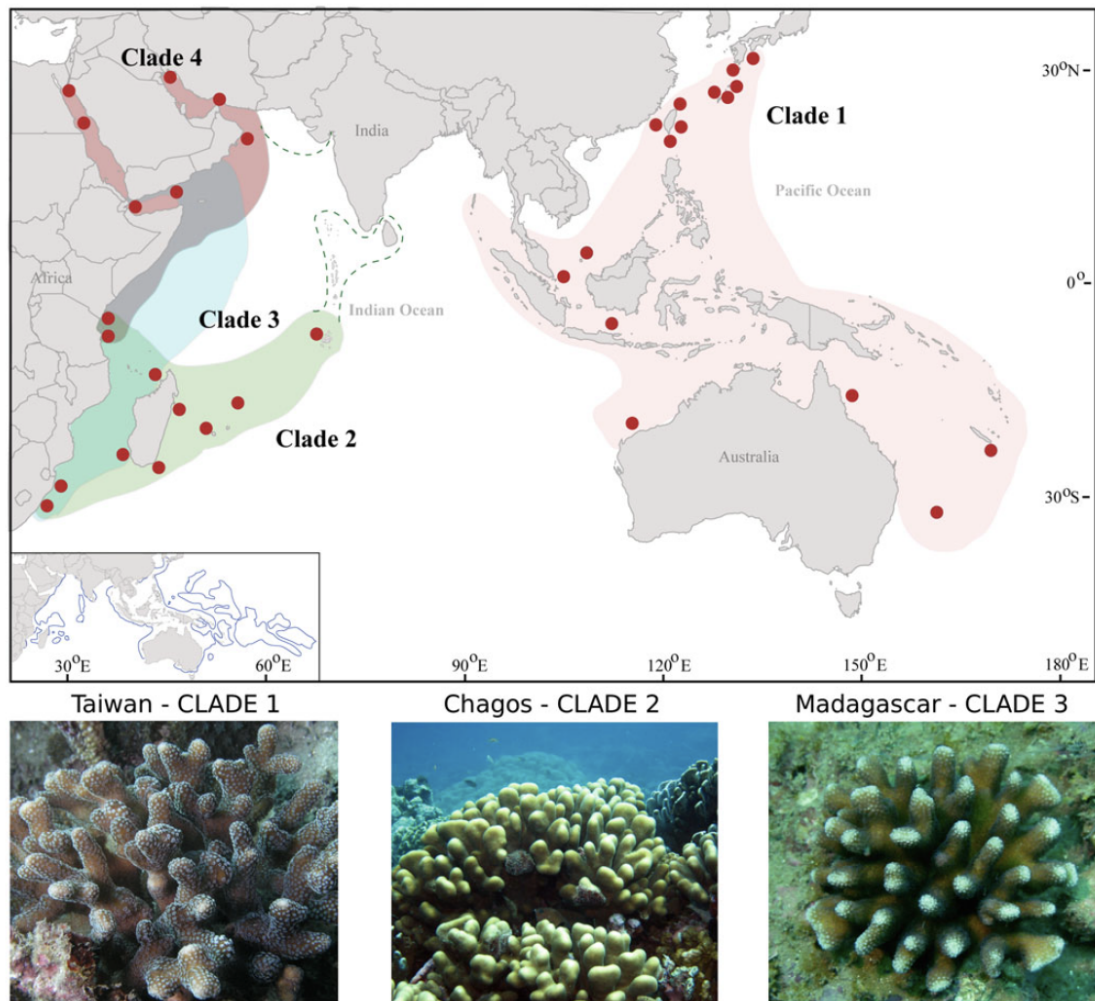


Figure I.16: DNA analysis has helped to recognize that many species of fish and invertebrates that were once thought to have widespread distributions are in fact made up of multiple distinct genetic lineages (**Knowlton *et al.* 2015**).

## I.2 Corals in symbiosis

### I.2.1 Obligatory symbiosis with the micro-algae Symbiodiniaceae

Corals live in relation with an assembly of different species and constitute a whole ecological unit. These species are bacteria, viruses and photosynthetic protists which together with the coral host, form a holobiont (**Rowan *et al.*, 1992**).

Most coral species live in obligatory endosymbiosis with micro-algae of the Symbiodiniaceae family. These micro-algae are housed in the oral endoderm of the coral, which is the second cell layer of the polyp (**Figure I.17 and Muscatine *et al.* 1977**).

Symbiodiniaceae produce nutrients via photosynthesis and these nutrients are then used for the growth and the reproduction of the coral host, allowing the production of the coral skeleton and therefore the formation of coral reefs (**Fournier *et al.* 2013**).

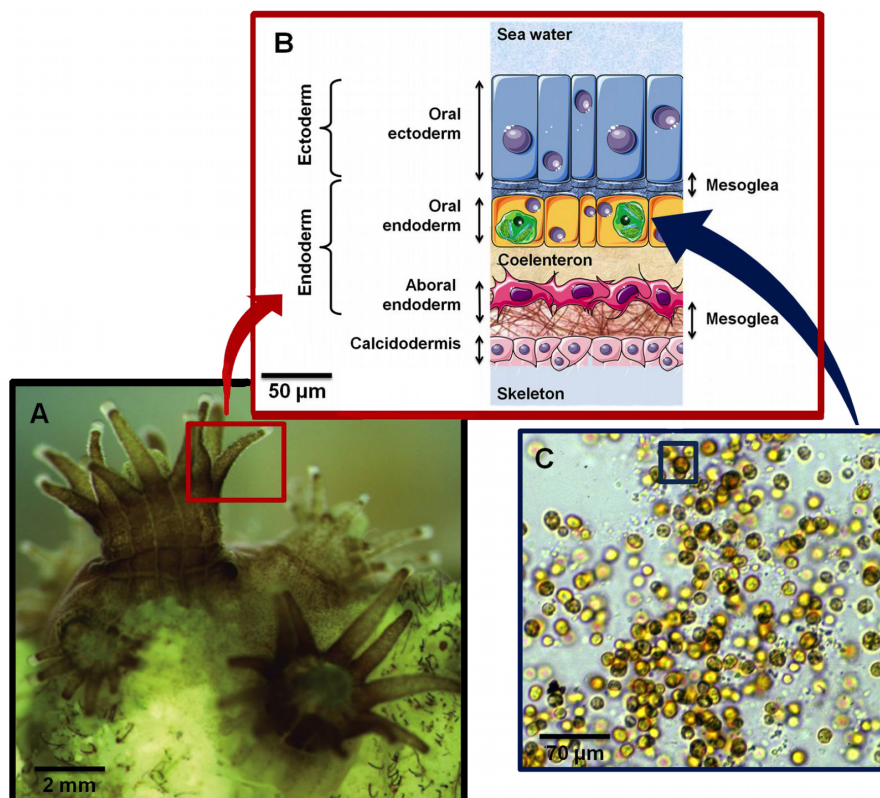


Figure I.17: Location of Symbiodiniaceae in coral polyps (**Fournier *et al.* 2013**).  
**(A)** Close up of coral polyps **(B)** Cross section showing histological organization of a symbiotic and calcifying scleratinian (stony corals). Corals are composed of two cell layers: ectoderm (epidermis) and endoderm (gastrodermis). Symbiodiniaceae are found in peri-vacuolar host membranes called symbiosomes located in the gastrodermal cells **(C)** Symbiodiniaceae under a light microscope.



The presence of this obligatory photosynthetic symbiont explains why most corals are found in bright environments in the first 50 meters of sea water (**Dubinsky et al. 1990**).

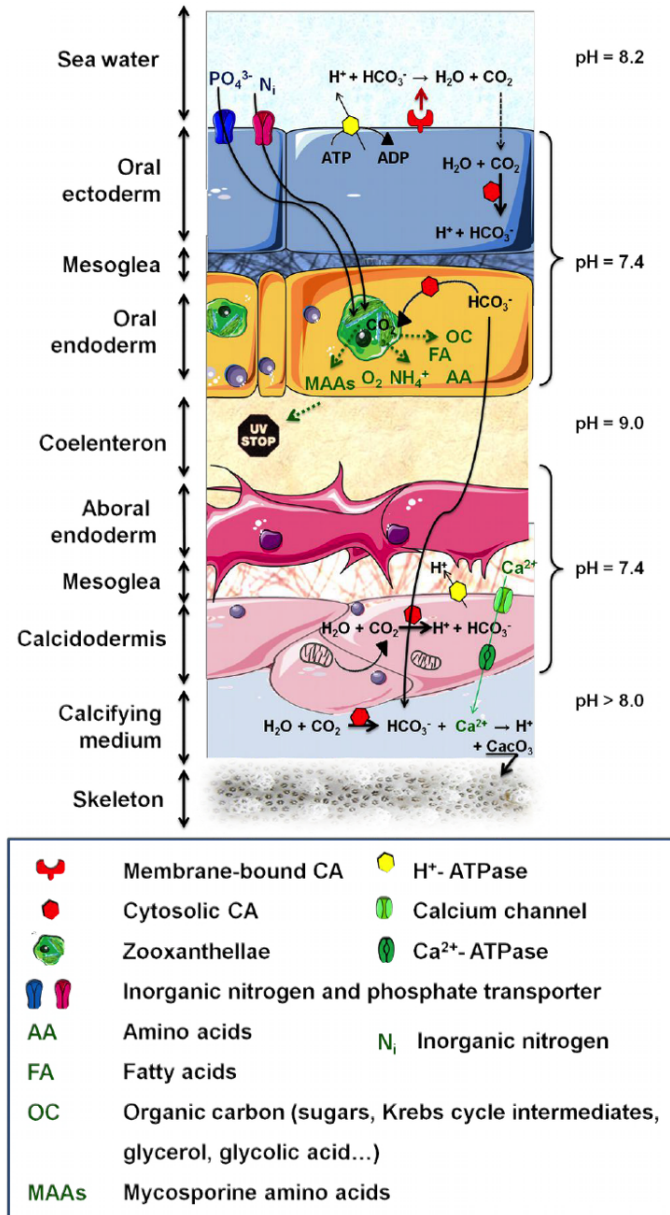


Figure I.18: Metabolic interactions and calcification processes between Symbiodiniaceae and coral polyps (**Fournier et al. 2013**).

The autotrophic algae acquires metabolic breakdown products and nutrients from the host (inorganic carbon and nitrogen ( $NH_4^+$ ,  $NO_3^-$ ,  $N_2$ ), phosphate ( $PO_4^{3-}$ ) and other inorganic compounds). The algae produce enough organic molecules to fulfill up to 95% of the host's energy demand. Carbon uptake necessitates an  $H^+$ -ATPase to acidify the boundary layer and lead to  $HCO_3^-$  formation which is then quickly and externally dehydrated into  $CO_2$  by a membrane-bound CA (Carbonic anhydrase).  $CO_2$  subsequently diffuses into the host, where it is converted to  $HCO_3^-$  by a cytosolic CA, preventing  $CO_2$  leakage. DIC (Dissolved inorganic carbon) is then supplied to (i) Rubisco for reduction or (ii) calcification sites. Biomineralization also necessitates a calcium (Ca) import from surrounding seawater and proton export, resulting from the mineralization process.

There are multiple metabolic interactions between the symbiotic micro-algae and the coral host (**Figure I.18**), and these are largely beneficial to both partners. Indeed, Symbiodiniaceae benefit from a protected environment inside host cells that contain nutrients in high concentration, which allows the symbiont to use them. Micro-algae use  $CO_2$ ,  $H_2O$ , light and other nutrients to perform photosynthesis, using nutrients conferred by the coral host. In exchange 80-95% of photosynthesis products are needed and used by the coral to grow (**Weis et al. 2008**). These products are amino acids, sugar and also fatty acids which are essential to the coral host. Symbiodiniaceae participate in the recycling of nutrients, like nitrogen for example (**Muscatine et al. 1977**). Micro-algae also produce Micosporine like Amino Acids (MAAs), which are secondary metabolites able to absorb U.V. radiation, and therefore protect the coral host against U.V. damages (**Rosic et al. 2012**).

There are physiological interactions between both organisms such as nutrient transports, which positively affect the life cycle of the holobiont unit. This association also produces of metabolites that are not formed by either of the organisms separately (**Muscatine et al. 1977**).

## I.2.2 Symbiosis breakdowns

The expulsion of Symbiodiniaceae from its coral host, is at the origin of bleaching events, that is why, Symbiodiniaceae expulsion mechanisms are extensively studied. There are several hypotheses concerning these mechanisms (**Figure I.19**).

Among these hypotheses there are:

- Degradation *in situ*: the cellular micro-algae die or are killed in the host cell. Leftovers can either be digested or eliminated.
- Exocytosis: there is an elimination of the micro-algae from the gastro-vascular cavity.
- Loss of the micro-algae by detachment from the host cell.

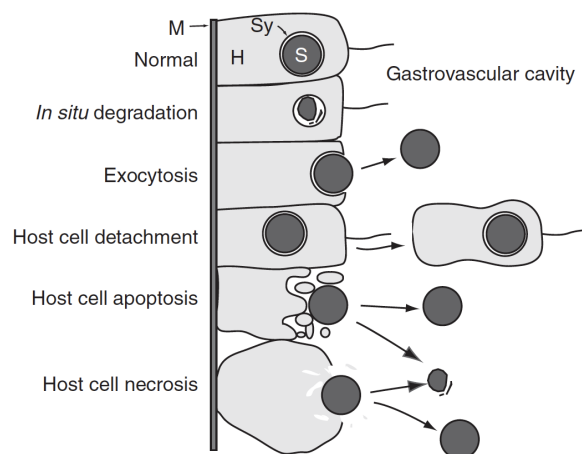


Figure I.19: Five different types of cellular mechanisms of symbiont loss from cnidarian host tissues (**Weis et al. 2008**).

- Apoptosis of the host cell, which causes the exit of the micro-algae.
- Necrosis of the host cell, which induces the Symbiodiniaceae expulsion.

The stress and death of coral host cells play a role in the breakdowns of the symbiosis, which induce coral bleaching. Symbiodiniaceae expulsion can also be due to damage in host cells (**Paxton et al. 2013**).

If environmental stressors are the main cause of Symbiodiniaceae expulsion, the replacement of a resident Symbiodiniaceae population by another population of micro-algae could also be an explanation for these mechanisms of expulsion (**Douglas et al. 2003**). Indeed, under some environmental conditions a micro-algae population can switch with another to confer higher fitness (i.e. adaptation) under the new environmental conditions within the host (**Douglas et al. 2003**).

Moreover, micro-algae survive outside from the hosts and can recolonize corals over a period of time or if the environmental conditions to establish the symbiosis are recovered. Without its symbiont, the coral dies within a few weeks (**Hume et al. 2012**).

Mass coral bleaching events have induced mass coral mortality and have increased over the last two decades. It appears that these events are a natural phenomena which contrasts with the natural stable symbiosis between coral and Symbiodiniaceae (**Wilkinson et al. 1999**).

It was shown that heat stress has an impact on the stable symbiosis and therefore on the metabolism of Symbiodiniaceae (**Figure I.20**). In consequence, there are potential breakdowns in the metabolic pathways in Symbiodiniaceae, which impact the components and the metabolisms of the micro-algae like photosynthesis, mitochondria, the Calvin cycle, etc... (**Fournier A. et al. 2013**). Acute heat and light stress can cause damage to photosystem II, for example, which induces a break in transport of electrons and then a reduction in ATP production. The accumulation of electrons, induces the formation of Reactive Oxygen Species (ROS) which are a danger for the survival of the micro-algae.

Moreover, the simulation of heat stress using nitric oxide production in Symbiodiniaceae showed a possible link between nitric oxide and the coral bleaching phenomenon. In fact nitric oxide is used in physiological functions of the micro-algae to recycle nitric compounds. A high production of nitric oxide in Symbiodiniaceae causes heat stress which also causes a decline in the photosynthetic efficiency. The increase of nitric oxide production at high temperatures

suggests that heat stress stimulates nitrate oxide production in the the micro-algae (**Bouchard et al. 2008**).

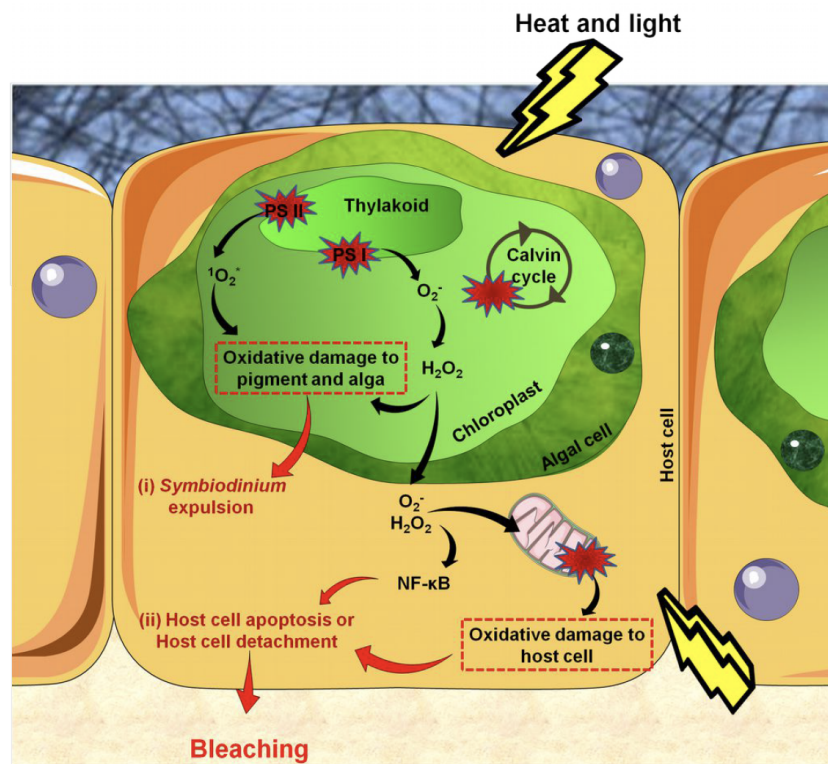


Figure I.20: Reactions leading to bleaching (**Fournier A. et al. 2013**).

Reactive Oxygen Species (ROS) production starts in the algae due to damaged proteins in photosystem II (PS II), impairments of the Calvin cycle and altered thylakoid membranes. Accumulated ROS then diffuse into host tissues causing damages to mitochondrial membranes, leading to ROS production in host tissues. This activates the innate immune response (NF-κB) which triggers apoptosis. Bleaching can be seen as the end of a controlled infection, with parasites signaling their presence by high ROS production.

A well-defined cascading response to thermal stress in corals has been identified in a multi-trait-based bleaching study (**Gardner et al. 2017**). The results showed that the cascading response to thermal stress is conserved across coral species. Moreover, bleaching is a conserved mechanism in the host, and specific adaptations linked to the corals antioxidant capacities drive differences in the sensitivity and thus tolerance of coral species to thermal stress. In fact, antioxidant capacities activate super oxide dismutase to eliminate ROS proteins. In consequence, it confers an increased resistance to thermal stress. The resistance and tolerance of corals facing environmental changes are important for maintaining the symbiosis between a coral and its micro-algae (**Gardner et al. 2017**).

Symbiotic breakdowns can also have important ecological feedback effects on coral reef

ecosystems (**Figure I.21**). For example, when ocean warming or acidification lead to a breakdown of the holobiont symbiosis, these changes can induce different relationships between corals and its environment, and then can be disturbed at different points, leading to a higher rugosity of the coral and therefore erosion. There is also ecological competition between the different micro-algae to adapt at different environmental conditions, after exposure to heat stress. This affects the micro-algae and by consequence the coral host in some points like coral growth.

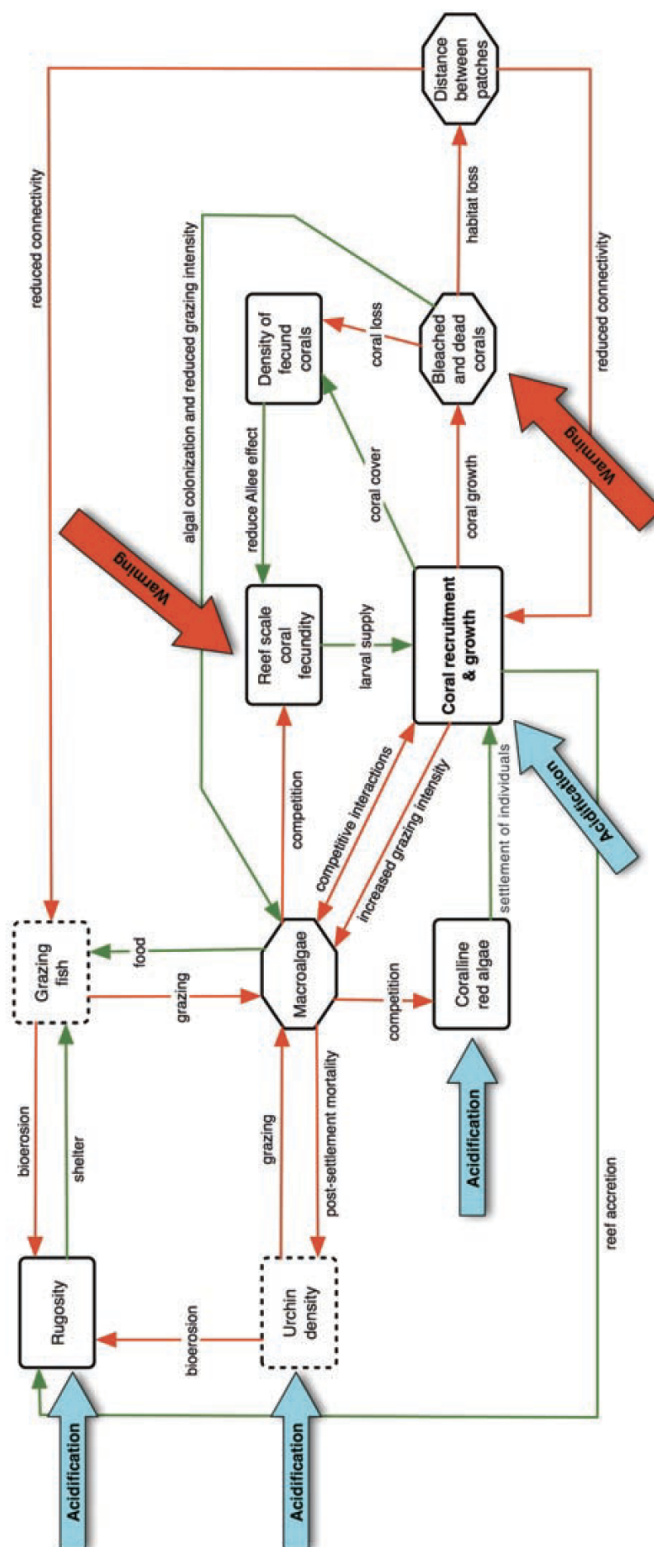


Figure I.21: Ecological feedback processes on a coral reef showing pathways of disturbance caused by climate change (Hoegh-Guldberg *et al.* 2007).

Impact points associated with ocean acidification (e.g., reduced reef rugosity, coralline algae) are indicated by the blue arrows, and impact points from global warming (e.g., bleached and dead corals) by the red arrows. Boxes joined by red arrows denote that the first factor has a negative (decreasing) influence on the box indicated. Green arrows denote positive (increasing) relationships. Over time, the levels of factors in hexagonal boxes will increase, whereas those in rectangular boxes will decline.

Boxes with dashed lines are amenable to local management intervention.

## I.3 Biology of Symbiodiniaceae

### I.3.1 Description of Symbiodiniaceae

Symbiodiniaceae is a family of unicellular eukaryote algae (**Figure I.22**). This micro-alga has a flagellum that it uses to move in the marine environment. It is found in two main forms, the coccoid form and the mastigote form. In the coccoid stage (**Figure I.23.B**), the micro-alga is spherical from 6 to 13  $\mu\text{m}$  in diameter, whereas in mastigote stage the cell is smaller with a flagellum (**Figure I.23.A**). Moreover, Symbiodiniaceae live in symbiosis or in free-living forms in marine environment.

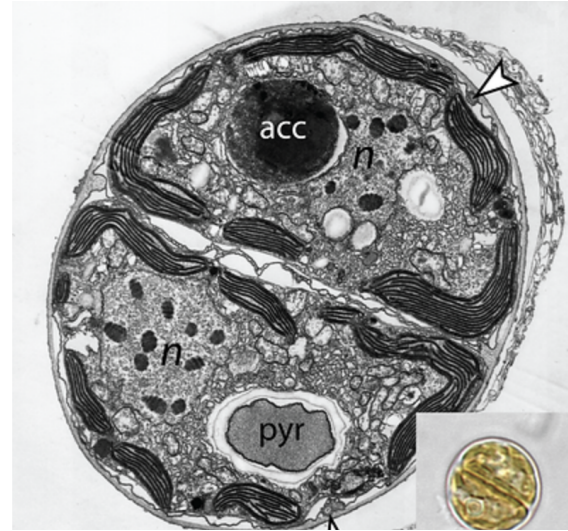


Figure I.22: Transmission electron micrograph of cross-section through a dividing cell (**Trench *et al.* 1981**).

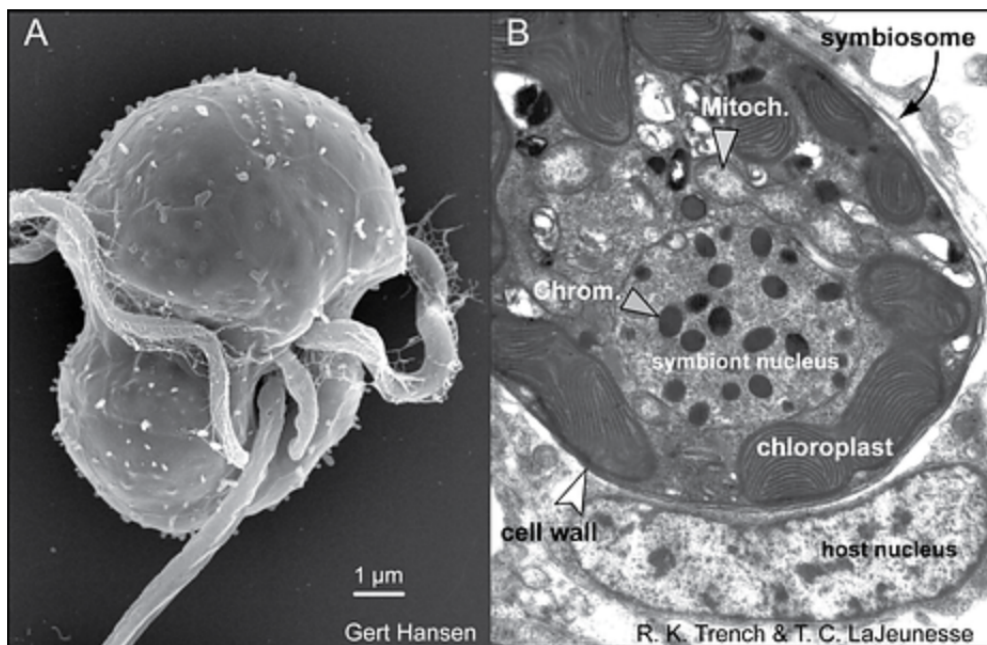


Figure I.23: The life stages of dinoflagellates in the Symbiodiniaceae family (**Trench *et al.* 1981**). (A) Electron micrographs of a micro-alga mastigote and (B) the coccoid cell in hospite.





### I.3.3 Symbiodiniaceae in symbiosis and heritability

Symbiodiniaceae species are present in a wide range of potential hosts and can live in symbiosis with different organisms from different classes like *Mollusca*, *Foraminifera*, *Scleratinia*, *Hydrozoa* etc. (Fournier *et al.* 2013). Among all these symbioses, the most famous concerns the symbiosis with corals.

The heritability of Symbiodiniaceae between coral colonies is either based on vertical or horizontal transmission according to the species. It was demonstrated that Symbiodiniaceae could be transmitted directly from the environment to the coral or maternally transmitted (Quigley *et al.* 2017).

In the case of horizontal transmission, Symbiodiniaceae acquisition is dependent on the environmental conditions, but also dependent on the requirements of the coral host. For example, after a bleaching event if the environmental conditions during recovery are favorable there is an horizontal acquisition of the micro-algae. These mechanisms are controlled by the coral host (Quigley *et al.* 2017).

In the maternal transmission pathway, Symbiodiniaceae are transmitted from coral polyps which already contain Symbiodiniaceae inside the cell. These transmission modes are influenced by the genetic lineage of the host (Quigley *et al.* 2017).

### I.3.4 Free-living form of Symbiodiniaceae

Symbiodiniaceae have the ability to live outside their host in the marine environment. This ability allows researchers to study these micro-algae in the laboratory under different environmental conditions. These experiments enable the study of Symbiodiniaceae diversity, investigations of gene expression of targeted genes through transcriptomic methods, or manipulative experiments to study the effect of different environmental conditions on the biology of Symbiodiniaceae (temperature, salinity, nitrogen or carbon dioxide concentrations, etc.).

Today the main studies are focused on Symbiodiniaceae diversity across the Pacific, the Atlantic and the Indian Oceans, but also in the Red Sea and the Caribbean Sea (Figure I.25). For example, a study of Symbiodiniaceae diversity in "free-living" form in the marine environment, across seas and oceans (A) and especially across coral reefs was performed. This study followed the *Tara* Oceans expedition which took place from 2009 to 2013. The search for the

micro-algae was made on several size fractions (B) like 0.8-5  $\mu\text{m}$ ; 20-180  $\mu\text{m}$  and 180-2000  $\mu\text{m}$ .

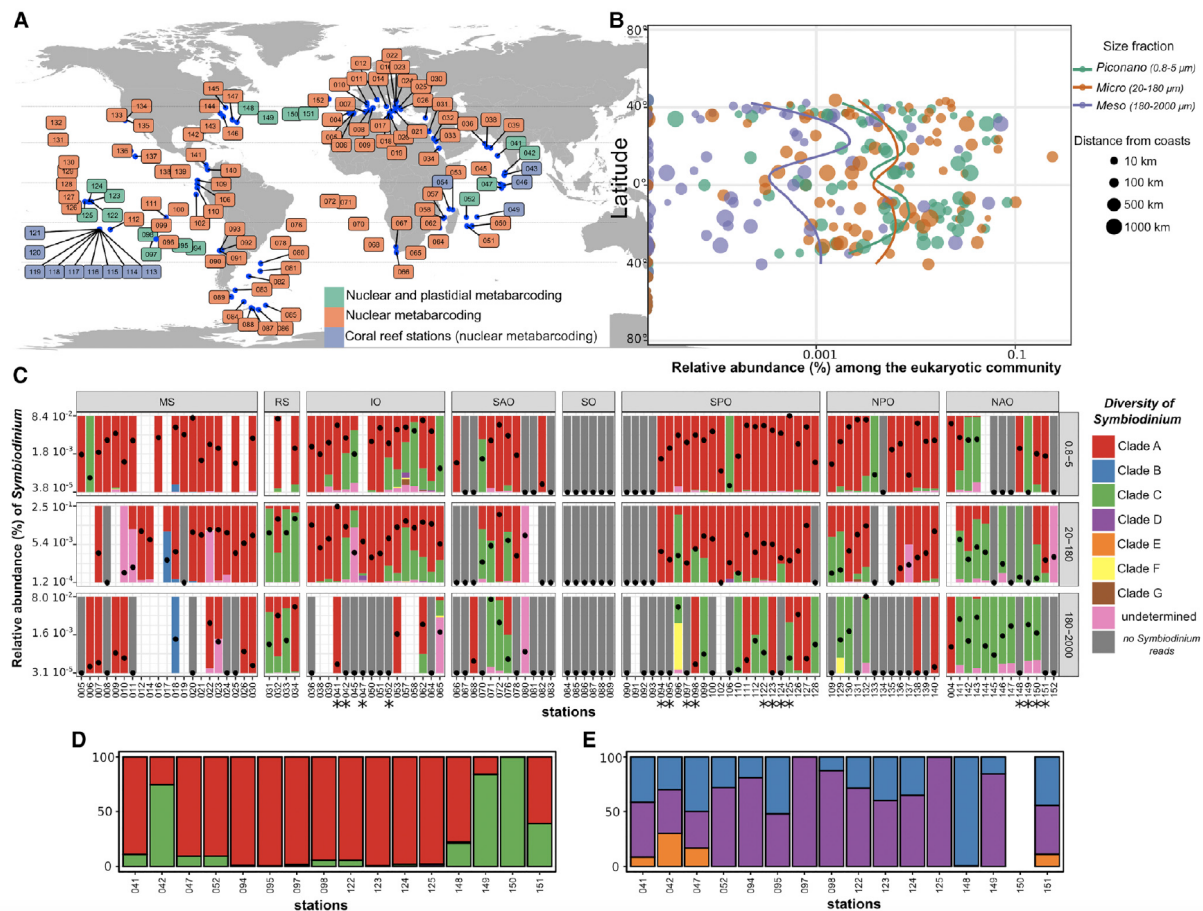


Figure I.25: Biogeography and diversity of Symbiodiniaceae in the open ocean (Decelle *et al.* 2018).

The diversity of the micro-algae in the marine environment (C) is composed of clade A, representing the major part of this diversity, followed by clade C in each size fraction. There is also a clear absence of the micro-algae in the Southern Ocean (SO). These results showed that Symbiodiniaceae can live far from coral reefs and they are mostly related to clades A and C.

In another study, researchers focused on the diversity of the free-living micro-algae at different periods and in different environments (Figure I.26). They found that the main clades A, B and C are present in the surrounding waters at different periods with relatively few F and G clades in summer. These results showed that it is complex to study the micro-algae over different periods in "free-living" form.

From these analyses it appears that Symbiodiniaceae at local scale live with a specific

host depending on environmental conditions and host organisms. At ocean scale, it can be hypothesized that Symbiodiniaceae use intermediate hosts to survive far away from coral reefs.

In this perspective, more studies on free-living Symbiodiniaceae are needed to better capture the biology of these important organisms when they are not in symbiosis with corals, which could conversely, enlighten questions on symbiosis itself.

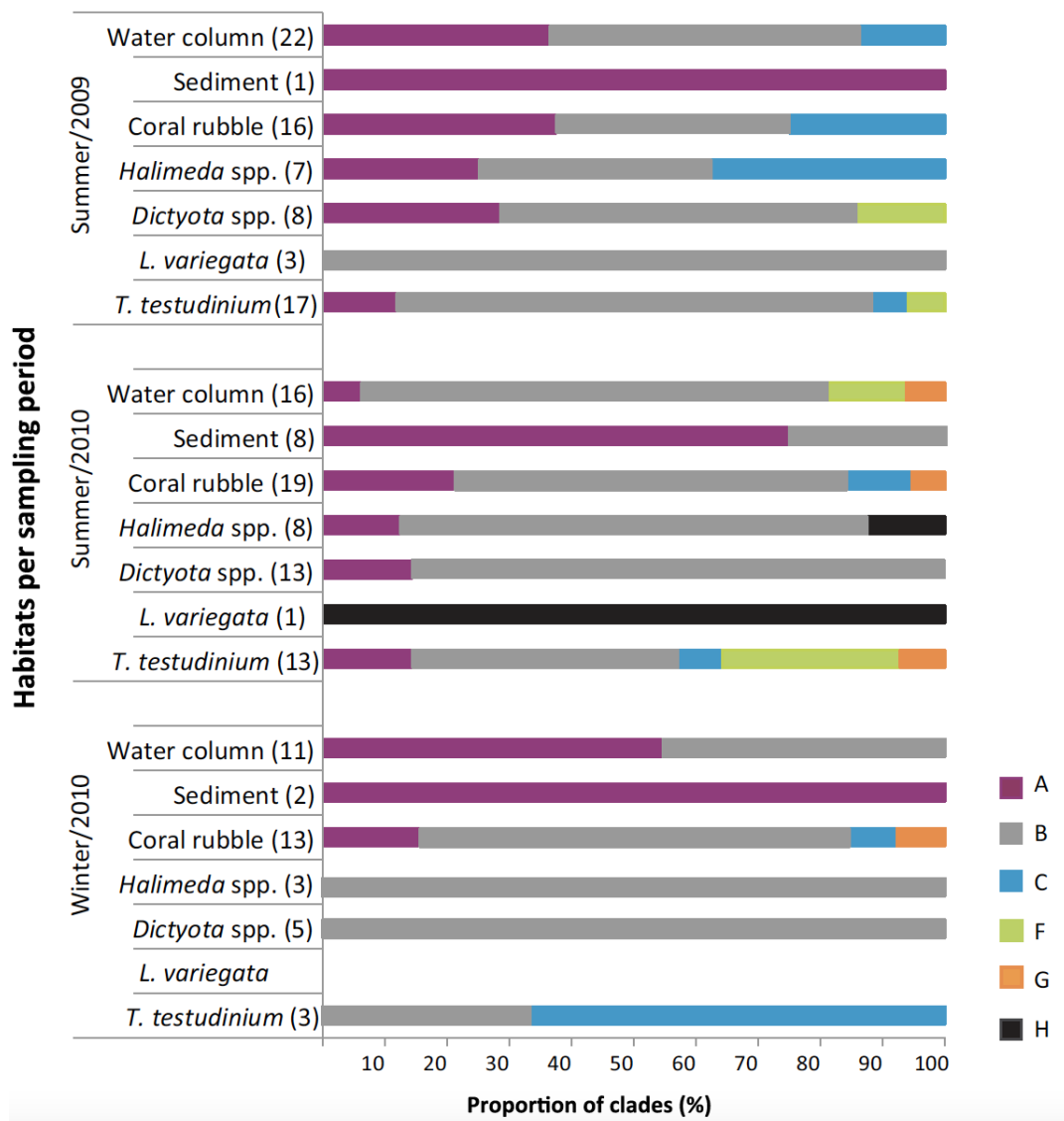


Figure I.26: Relative proportions of free-living Symbiodiniaceae found in each habitat during the different sampling periods (**Granados-Cifuentes et al. 2013**).

These proportions were determined in each sampling period by dividing the number of Symbiodiniaceae sequence counts per clade by the total number of Symbiodiniaceae sequence counts in each habitat. Numbers in parenthesis represent the sample sizes of each habitat that were successfully amplified. Symbiodiniaceae clade colored as shown on the lower right. *Lobophora variegata* was not collected during winter of 2010

## I.4 Symbiodiniaceae diversity

### I.4.1 Symbiodiniaceae identification

The family Symbiodiniaceae is found in the kingdom Eukaryota, and are related to the SAR<sup>1</sup> clade. Within this clade they belong to the phylum Alveolata, the Dinophyceae class and to the order of Suessiales.

The modernization and development of high throughput sequencing methods has enabled a major advance in deciphering species diversity, including the Symbiodiniaceae family. Indeed, rRNA<sup>2</sup> sequences help to explain the diversity of symbiotic dinoflagellates (**Rowan *et al.* 1992** and **Sadler *et al.* 1992**). It is now possible to identify the genus and species of this micro-algae using these sequences (**LaJeunesse *et al.* 2018**).

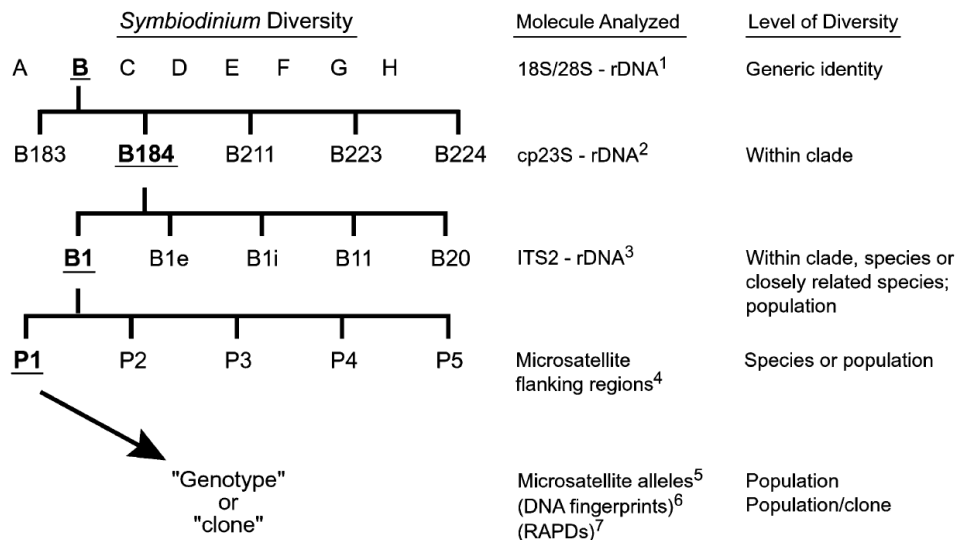


Figure I.27: A schematic representation of the multiple levels of diversity found within Symbiodiniaceae (**Coffroth M.A *et al.* 2005**).

Symbiodiniaceae diversity can be studied at different levels:

- It is possible to determine the genus of the micro-algae by sequencing the 18S rRNA or the 28S rRNA. There are 9 Symbiodiniaceae genera originally identified with a letter from A to I.
- The 300 base pair long ITS2<sup>3</sup> sequence, localised between the 5.8S rRNA and the 28S

<sup>1</sup>Stramenopiles, Alveolates and Rhizaria super group

<sup>2</sup>ribosomal RNA

<sup>3</sup>Internal Transcribed Spacer 2

rRNA, can be used to differentiate Symbiodiniaceae at the species level.

- However, the ITS2 sequence is a multi-copy marker with a high intragenomic diversity. That's why ITS2 sequences cause analytical challenges in distinguishing intra and inter genomic diversity. The species are named with a letter associated with the genus, following the letter to the ITS2 sequence.

## I.4.2 Symbiodiniaceae hosts

The main known hosts for the different Symbiodiniaceae are well known (**Figure I.28 and Fournier *et al.* 2013**).

Some clades are found in many host organisms belonging to various orders, such as clade A and C, which are thus called generalist clades. However, other clades such as Clade I for example have a more restricted host diversity, it represents a specialized clade, which means it groups species that are able to live within a restricted number of hosts. A summary of all the Symbiodiniaceae species and hosts discovery are presented in the work of **Riddle *et al.* 2016**.

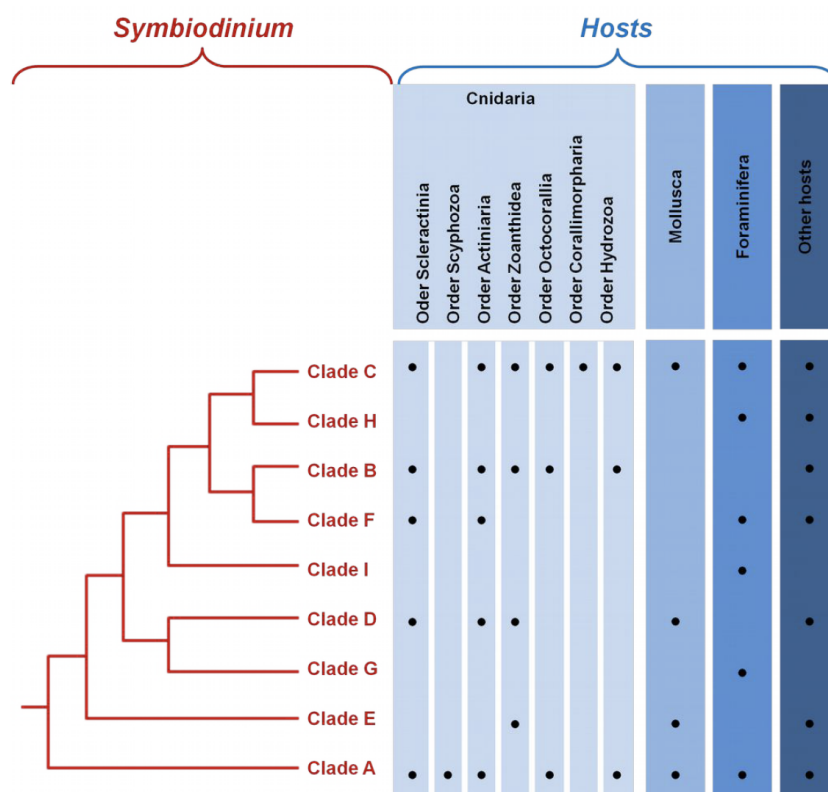


Figure I.28: Phylogenetic relationships between the major clades of Symbiodiniaceae (**Fournier *et al.* 2013**).

### I.4.3 Symbiodiniaceae phylogeny

The Symbiodiniaceae phylogeny was recently updated (**LaJeunesse *et al.* 2018**). The distinctions between Symbiodiniaceae clades made it possible to assign new names to certain genera (**Figure I.29**). The clades formerly named A, B, C, D, E, F and G are now respectively named *Symbiodinium*, *Breviolum*, *Cladocopium*, *Durusdinium*, *Effrenium*, *Fugacium* and *Gerakladium*.

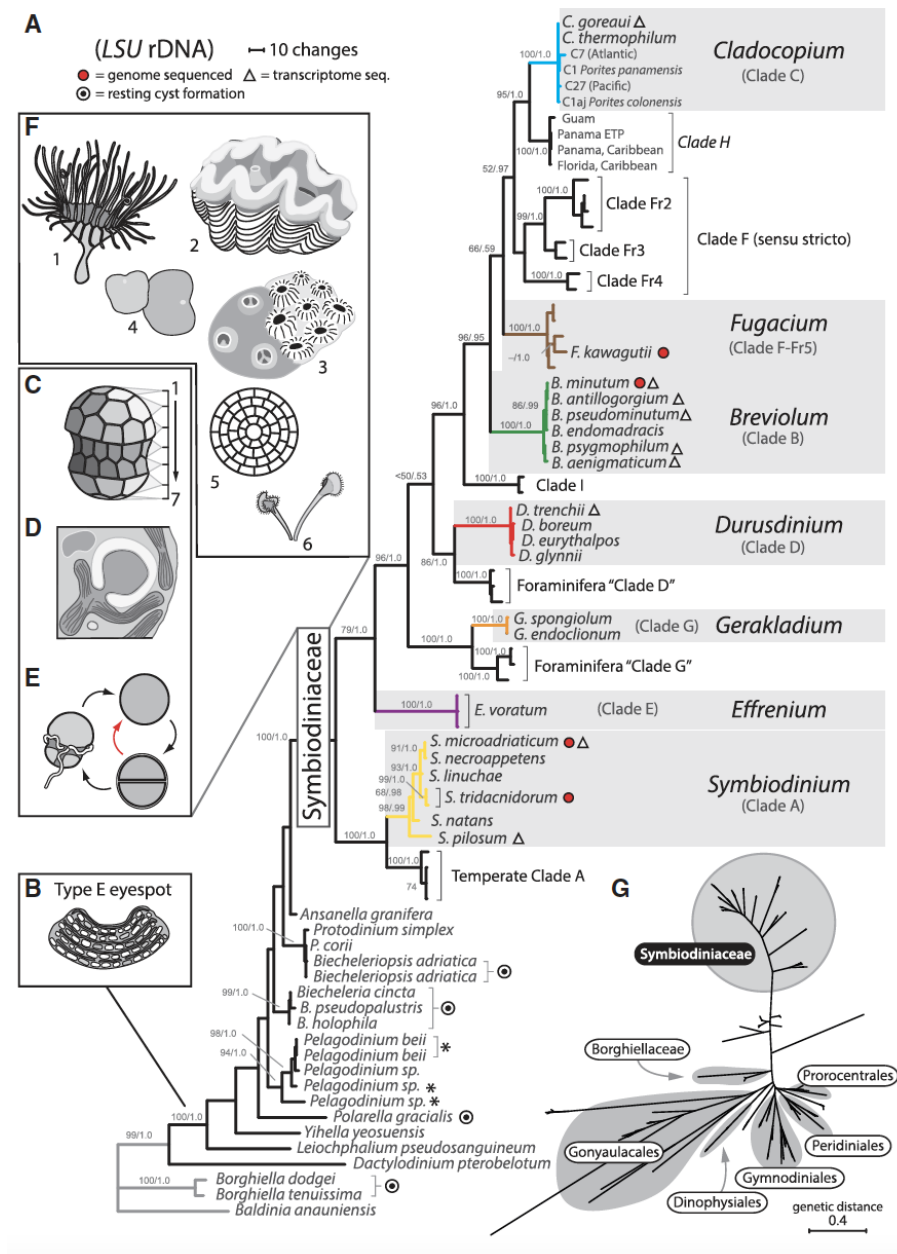


Figure I.29: Proposed new genera in the family Symbiodiniaceae and LSU rDNA phylogeny portraying their evolutionary relationships, including remaining lineages that require generic names (**LaJeunesse *et al.* 2018**).

Clade A is the ancestral clade of the Symbiodiniaceae family. An unrooted phylogenetic reconstruction of clade A based on maximum parsimony analysis of ITS2 sequences was performed (**Figure I.30**). This clade is well studied in the corals from the genus *Millepora*, located in the Indo-Pacific Ocean, for example.

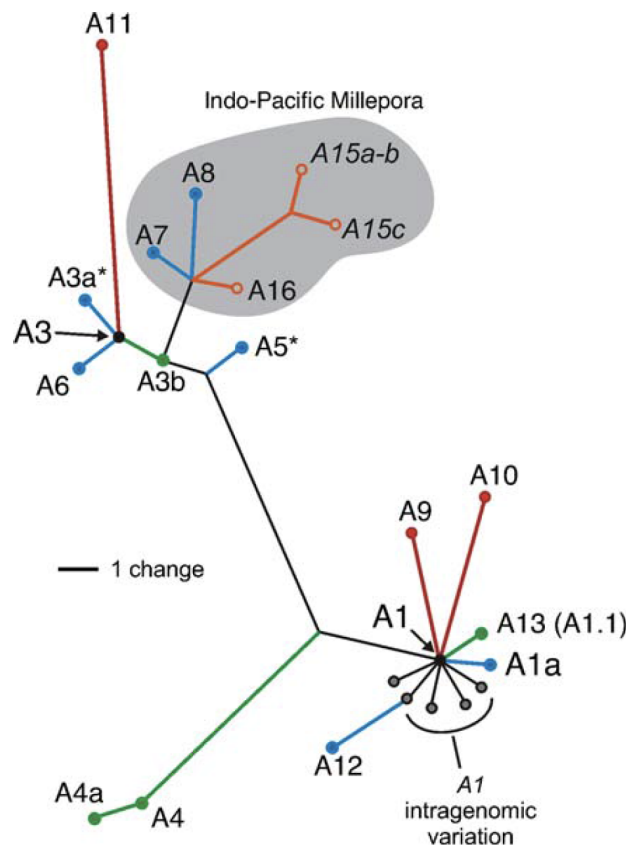


Figure I.30: An unrooted phylogenetic reconstruction of clade A based on maximum parsimony analysis of ITS2 sequences (**LaJeunesse et al. 2009**).

Branches are color-coded to signify the regional origin of the ITS2 variant. Red = Red Sea; Orange = Indian Ocean; Blue = Pacific, Green = Western Atlantic.

A phylogenetic reconstruction of the generalist clade C was also performed (**Figure I.31**). The 2 main variants in this clade are called C1 and C3, and they constitute the ancestral core of the *Cladocopium* genus. The other edges represent sub-variants with possible intragenomic variations, often found in the same coral host species. It also shows that multiple Symbiodiniaceae species can be associated to the same coral genus. For example, corals in the genus *Porites* are associated with C45 in the Atlantic Ocean and C15 in the Indo-Pacific Ocean. Moreover, this tree shows the difficulty of identifying Symbiodiniaceae species based on ITS2 sequences, because clade C is the most diverse clade.

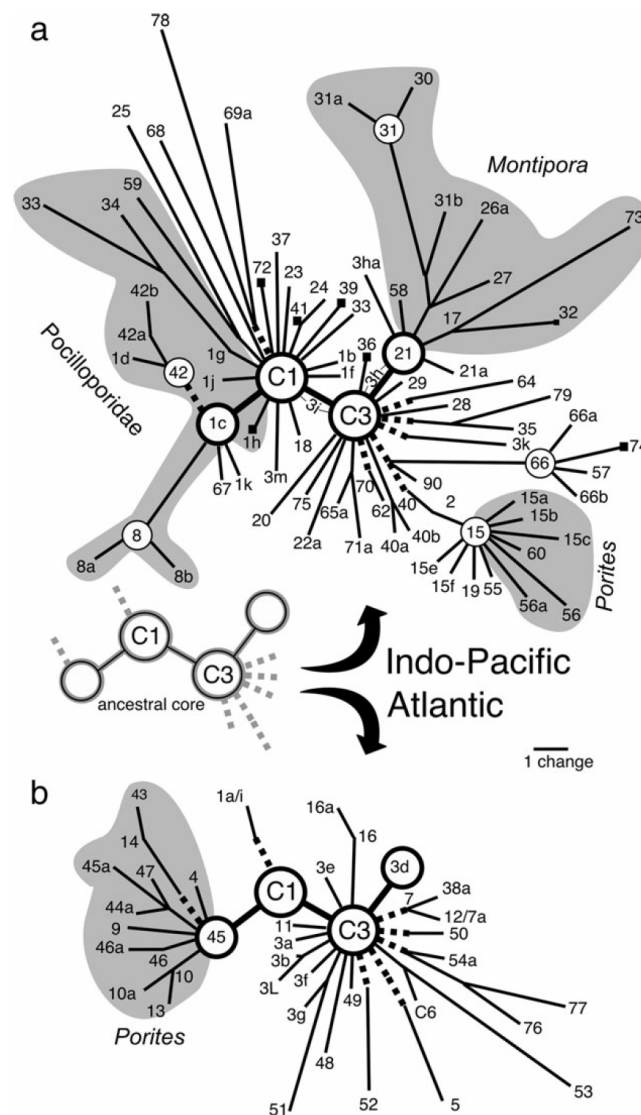


Figure I.31: Independent radiations of Symbiodiniaceae clade C in the (a) Indo-Pacific and (b) Atlantic-Caribbean (LaJeunesse *et al.* 2004).

A small number of closely related progenitor sequences constitute the "ancestral core"; Little or no sequence overlap exists between radiations from each ocean. Branches with broken lines extending from the "ancestral" core are instances of overlap but may represent homoplasy <sup>4</sup>.

By comparison with the phylogenetic tree of clade A, this tree is more diverse in terms of ITS2 sequences. As another example, the phylogenetic reconstruction of ITS2 sequences of clade D was not done, because only few species are identified in this clade.

These results show that each clade has its own specifications and complexities. They reflect a spatial and temporal genetic structure of micro-algae species, therefore the genetic diversity of ITS2 sequences reflects the biogeography of Symbiodiniaceae (Howells *et al.* 2013).



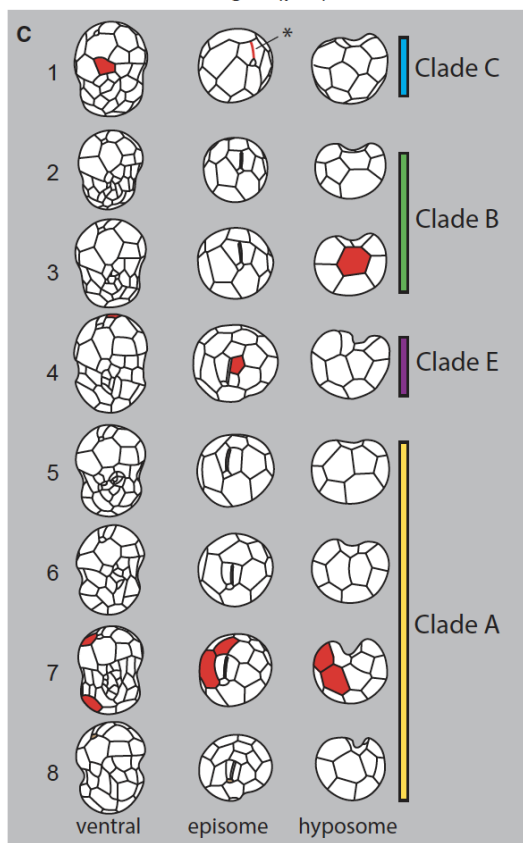


Figure I.32: Morphological differences within and among Symbiodiniaceae clades (LaJeunesse *et al.* 2018).

Amphiesmal plate arrangements and patterns of motile cells currently known for taxa within and between Symbiodiniaceae clades. Plates highlighted in red indicate unique differences in plate numbers among the seven latitudinal series of amphiesmal vesicles. The asterisk indicates absence, or reduction, of the elongate apical vesicle (EAV; "acrobases" or apical furrow apparatus). 1. *Symbiodinium goreau*, 2. *S. minutum*, 3. *S. psygmophilum*, 4. *S. voratum*, 5. *S. microadriaticum*, 6. *S. necroappetens*, 7. *S. tridacnidorum*, 8. *S. natans*.

across most clades. Differences in plate number and shape can be higher between species within a clade than they are among clades (Figure I.32).

Several Symbiodiniaceae can sometimes colonize the same coral colony (Yorifuji *et al.* 2017). Moreover, the Symbiodiniaceae diversity in the host is dependent on some criteria linked to the age of the host and also to environmental parameters (Yorifuji *et al.* 2017). For example, the micro-algae composition in juvenile *Acropora tenuis* colonies (stony corals) changes according to different environmental conditions: clade D is more adapted to high temperatures than Symbiodiniaceae in clade C (Yorifuji *et al.* 2017).

After heat stress, coral hosts tend to have a prevalence of one kind of symbiont, more adapted to its new environmental conditions. This change in the symbiont diversity, may help the coral survive after modifications in the environment. (Yorifuji *et al.* 2017). The results show also that coral growth decreases after symbiont switch in the coral host, caused by modification in environmental conditions (Yorifuji *et al.* 2017).

This kind of switch is also observed for example in *Acropora tenuis*, by a switch in the dominant symbiont type from *Symbiodinium* to *Durusdinium* after acclimation to changing environmental conditions (Yorifuji *et al.* 2017).

Some genera of Symbiodiniaceae were distinguished through morphology based on microscope observations. The number, shape and arrangement of amphiesmal plates are similar

## I.5 Genomes and transcriptomes of Symbiodiniaceae

### I.5.1 Symbiodiniaceae references

To understand mechanisms of coral bleaching and other biological aspects of Symbiodiniaceae, genomes and transcriptomes of the micro-algae were sequenced and assembled. The first Symbiodiniaceae transcriptome was sequenced in 2012 (**Figure I.33**). Today there are more than 30 transcriptome<sup>5</sup> and 8 genome<sup>6</sup> references available.

The number of studies focusing on differential expression analyses between Symbiodiniaceae genes has increased considerably. In most of these studies, Symbiodiniaceae are cultivated in their free-living form in the laboratory, enabling, for example to study their reactions under different stress conditions.

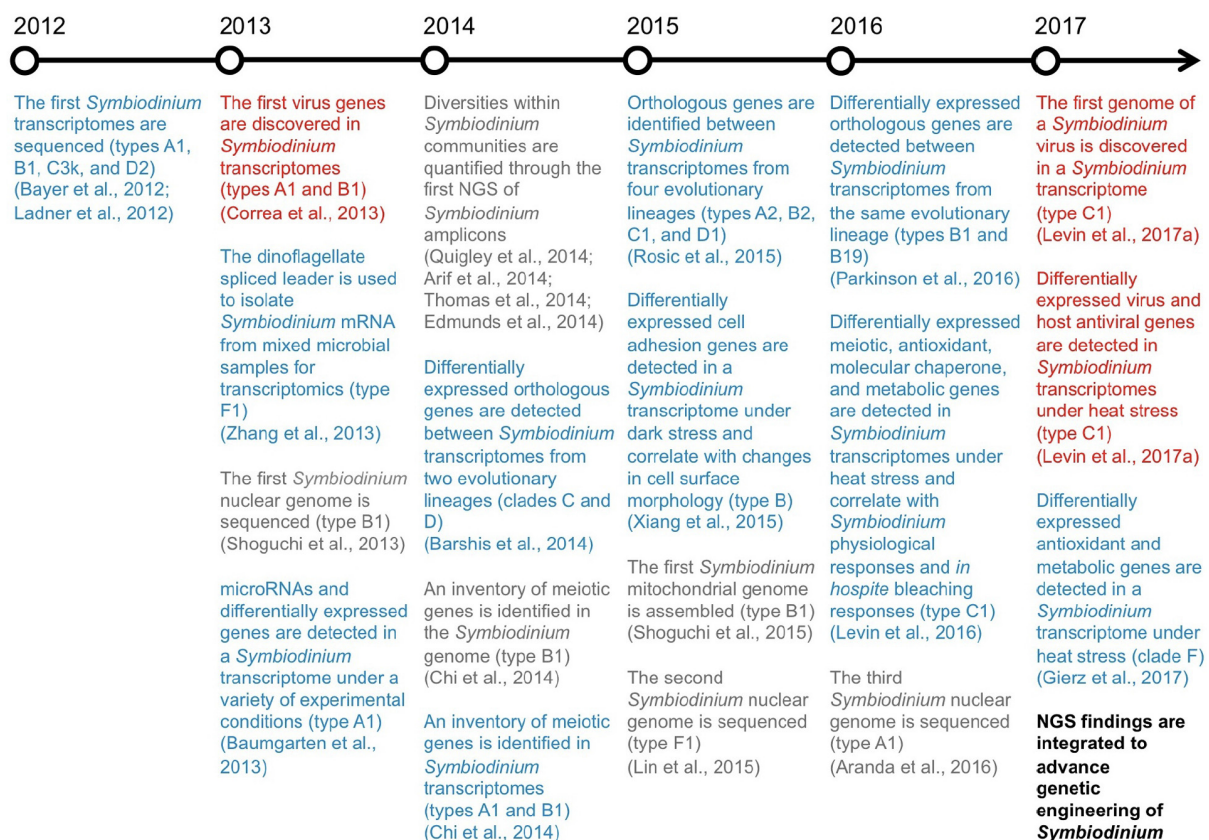


Figure I.33: Breakthroughs in NGS of Symbiodiniaceae (Levin et al. 2017). A timeline highlighting the key genomic (gray), transcriptomic (blue), and virus RNA (red) findings from recent NGS studies of Symbiodiniaceae.

<sup>5</sup>Chapter 2 - Table II.1

<sup>6</sup>Chapter 2 - Table II.2

Other analyses of Symbiodiniaceae within coral were done and they are very important for understanding the mechanisms which govern the coral symbiosis. In this study, the sequencing methods are named dual-genomic and dual-transcriptomic (**Shinzato *et al.* 2014**).

### I.5.2 Gene regulation in Symbiodiniaceae

The integration of small RNA and mRNA<sup>7</sup> sequences, in the expression profiling of Symbiodiniaceae were studied (**Baumgarten *et al.* 2013**). These small RNA are thought to be mRNA potentially involved in the control of protein modifications, signaling, gene expression and response to DNA damages.

Moreover Symbiodiniaceae contain a complete set of core histones and some of them are known to be involved in gene expression regulation (**Bayer *et al.* 2012**).

Transcriptome analyses indicate that a rather low number of transcription factors are present in Symbiodiniaceae genomes. The "cold shock domains" are predominant among transcription factors and there is also a high number of antioxidative genes (**Mihailovich *et al.* 2010**).

### I.5.3 Symbiodiniaceae under heat stress

Transcriptomic analysis of response to thermal stress in Symbiodiniaceae, revealed differentially expressed genes involved in stress response such as the antioxidant components and molecular chaperones, photosynthesis, protein complexes and enzymes such as fatty acid desaturases (**Gierz *et al.* 2017**).

The different clades of Symbiodiniaceae are affected by heat stress, and the consequences are a decrease in Symbiodiniaceae density and photosynthetic efficiency (**Gierz *et al.* 2017**). Moreover production of ROS<sup>8</sup> under heat stress is deleterious to Symbiodiniaceae cells. ROS can cause oxidative damage to lipids, proteins and DNA (**Gierz *et al.* 2017**). Fatty acids are essential to maintaining thylakoid membrane lipid composition against thermal stress and are useful (**Tchernov *et al.* 2004**) as lipid biomarkers for stress (**Kneeland *et al.* 2013**).

It also appears that photosynthesis can be used as an indicator of thermal stress in Symbiodiniaceae. In fact, under heat stress over 30°C this pathway is impaired whereas over 34°C it ceases completely (**Iglesias-Prieto *et al.* 1992**). There is also an impact of higher temperature

---

<sup>7</sup>messenger RNA

<sup>8</sup>Reactive Oxygen Species

on the pigment profiles of Symbiodiniaceae, which are altered by the photo-oxidative response (Venn *et al.* 2006).

Some other studies showed resistance of micro-algae during heat stress. This resistance is linked to the micro-algae species diversity and can confer increased resilience to heat stress, which is also beneficial for the coral host (Gierz *et al.* 2017).

The resistance can be caused by the exposure of micro-algae during a short time to heat stress. In fact, if the period of stress exposure is short, the impact on the symbiosis is less than a longer exposure. This suggests that the physiological plasticity of the host and/or symbiotic components appears to play an important role in the context of ocean warming. The physiology of the organisms which compose the symbiosis, has an importance on the mechanisms leading to rapid holobiont acclimation (Bellantuono *et al.* 2012).

These different results show that Symbiodiniaceae are capable of adapting to their hosts and to environmental conditions. For example, the adaptation of the coral *Porites lobata* which contains *Cladocopium C3* in Fiji compared to Abu Dhabi after exposure to heat stress was studied (Figure I.34). It showed that *Porites lobata* symbiosis from Abu Dhabi can recover after a long period of exposure to heat stress (15 days) (Hume *et al.* 2012). These results showed that geographical location has an impact on the resistance of the symbiosis in coral and in Symbiodiniaceae (Hume *et al.* 2012).

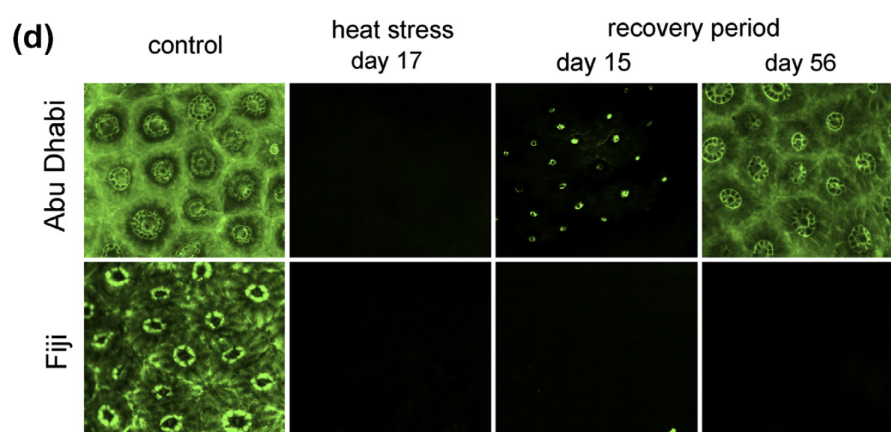


Figure I.34: Differential responses of *Porites lobata* to experimental heat stress treatments (Hume *et al.* 2012).

Fluorescence microscopic images documenting the intensity changes of cyan tissue fluorescence from control conditions, after 17 days of incubation time at 31.5°C, and during the recovery period 15 and 56 days after returning the corals to reduced temperature.

Another recent study underlines that high-frequency temperature variability influences bleaching prevalence (**Safaie *et al.* 2018**). These results also confirmed that an increase in daily temperature range could reduce the odds of more severe bleaching by a factor of 33. Again, it shows how the coral and the symbiont can adapt in the context of coral bleaching events.

#### I.5.4 Symbiodiniaceae adaptation to light intensity

Photosynthesis is essential for micro-algae growth. Variations in the light intensity induce modifications in the morphology of the Symbiodiniaceae cell (**Figure I.35**). In fact, the expression of many genes appears to be significantly impacted by light. Some transcripts of photosynthesis functions declined in high light conditions whereas some others declined in the dark. Other genes are involved in photoprotection and the dissipation of excess absorbed light energy. Moreover when micro-algae are transferred from light to darkness, the transcripts encoding several cell adhesion proteins rapidly declined, which correlated with changes in cell surface morphology (**Xiang *et al.* 2015**).

Under bleaching conditions, the photosynthetic apparatus is activated at extremely efficient non-photochemical mechanism. This mechanism is associated with a transition from an initially heterogeneous photosystem II (PSII) pool to a homogeneous pool. It results in an almost complete cessation of photosynthetic electron transport. This mechanism leads to the inhibition of organic carbon production which may provide a trigger for the expulsion of Symbiodiniaceae from the host (**Slavov *et al.* 2016**).

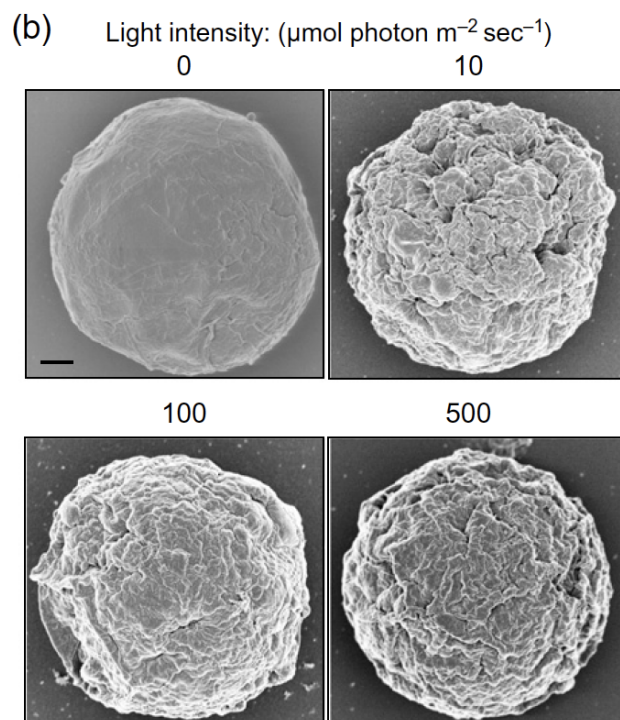


Figure I.35: Scanning electron microscopy (SEM) of Symbiodiniaceae showing surface structures under different light conditions (**Xiang *et al.* 2015**).

The surface of Symbiodiniaceae after exposure of the cells to different light intensities for 24h. Light intensity is given above each image; scale bar is 1 micrometers.

### I.5.5 Symbiodiniaceae nutrient transport

Studies of Symbiodiniaceae genomes and transcriptomes have revealed a very high diversity in functions of nutrient transport functions (**Figure I.36**). Exchanges between host and symbiont constitute the majority of studies. Nutrient transport is essential for the life of the micro-algae and in the coral symbiosis, it is essential for the survival of the coral (**Figure I.36**). These exchanges are complex and varied; there are over 32 major micro-algae related transporters that span various metabolic pathways (**Lin et al. 2015**).

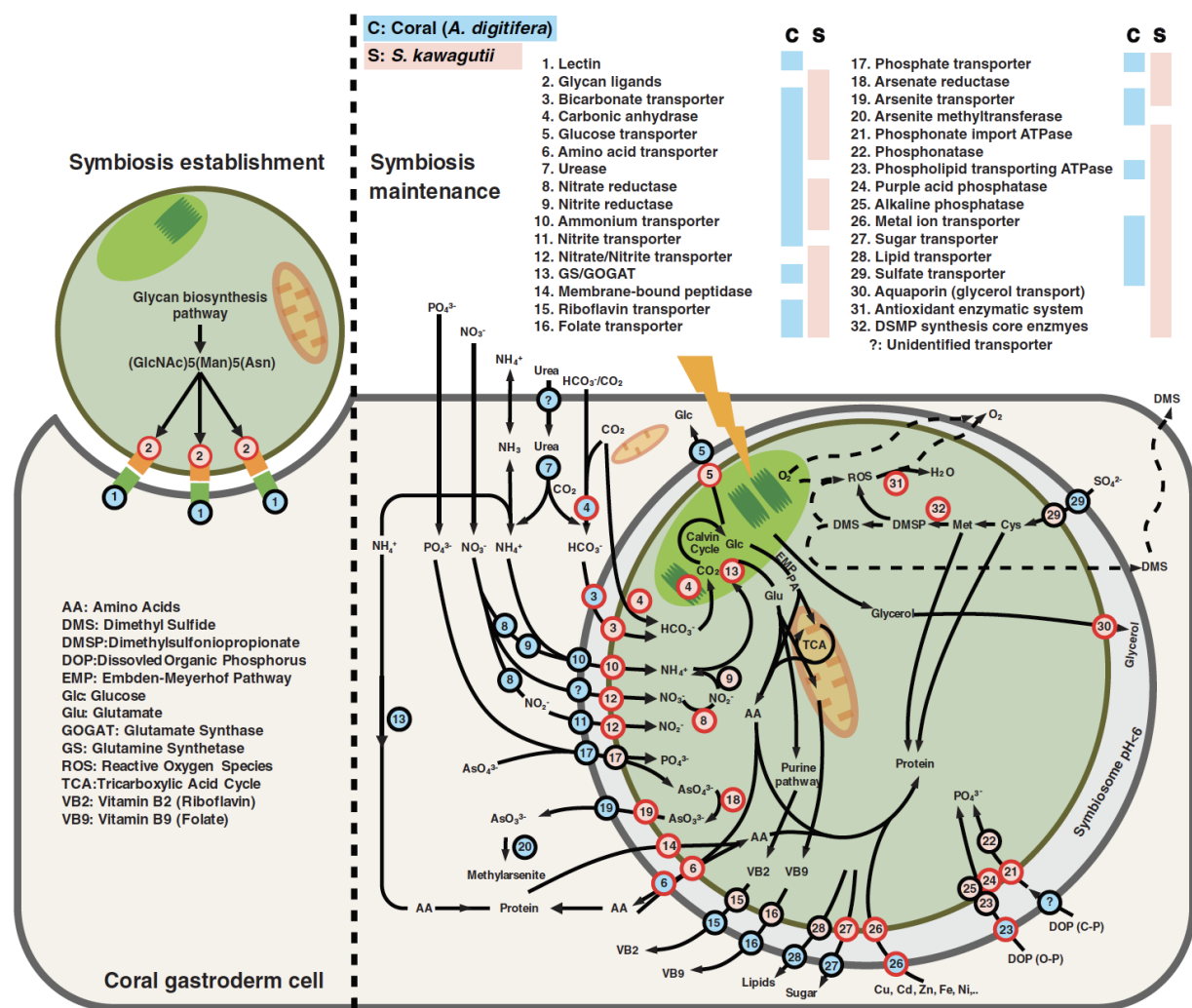


Figure I.36: Schematic summary of host recognition mechanisms and cargo transport (**Lin et al. 2015**). Small circles represent genes from the coral (sky blue fill) and from the Symbiodiniaceae (pale pink fill), with predicted miRNA target genes marked with red outlines. Only genes considered directly related to symbiosis are shown. All the Symbiodiniaceae transporter families shown have members that are computationally predicted to be plasma membrane proteins.

Among these metabolic pathways, several essential representatives can be found in different locations within the Symbiodiniaceae cell such as the thylakoid, mitochondria and plasmic membrane linked to the transports of nutrients. These essential pathways are: Bicarbonate transporter (3), Glucose transporter (5), Amino Acid transporter (6), Ammonium transporter (10), Nitrite transporter (11), Nitrate/Nitrite transporter (12), Riboflavin transporter (15), Folate transporter (16), Phosphate transporter (17), Arsenite transporter (19), Metal ion transporter (26), Sugar transporter (27), Lipid transporter (27), Sulfate transporter (28) and Aquaporin (glycerol transporter) (30). These different metabolic pathways involved in transport are essential to understand the symbiosis between Symbiodiniaceae and its coral host.

### I.5.6 Symbiodiniaceae adaptation in different location

There is also evidence of micro-algae adaptation in the same geographical areas. Indeed, micro-algae from clade D are thermal tolerant compared to the other clades in the Pacific Ocean. Genes like fatty acid desaturase, molecular chaperones and proteins involved in photosynthesis are involved in the thermal tolerance of clade D compared to the other clades (**Ladner et al. 2012**).

However, corals from Arabian/Persian Gulf are able to survive extreme sea temperatures (> 35°C). The genetic structure of these populations suggests that both partners of the symbiosis may contribute to the adaptation of coral bleaching, because in these warmest regions coral reefs are adapted to their environment (**Smith et al. 2017**). Other studies, suggested a genetic adaptation of the coral *Platygyra daedalea* on these reefs localized in the Red Sea (**Smith et al. 2017**). In parallel, the study of corals associated with *Cladocopium C3* hypothesized a genetic adaptation to this warm environment. In fact, thermo-tolerant *Durusdinium* are extremely rare in these regions, whereas it is shown as an environmental adaptation in the Pacific Ocean.

The composition of the micro-algae in the host is shaped by the acclimation of the coral host to its environment. It appears that many corals contain multiple clades simultaneously, which may allow for rapid acclimation to change in different environmental conditions (**Ladner et al. 2012**). These changes in the composition of Symbiodiniaceae appear after successive events of heat stress (**Silverstein et al. 2014**).

## I.6 The *Tara* Pacific expedition

*Tara* is a sailing boat which has already traveled more than 400.000 kilometers, in more than 60 countries during 4 major expeditions which are: *Tara Arctic* (2006-2008), *Tara Oceans* (2009-2013), *Tara Mediterranean* (2014) and *Tara Pacific* (2016-2018).

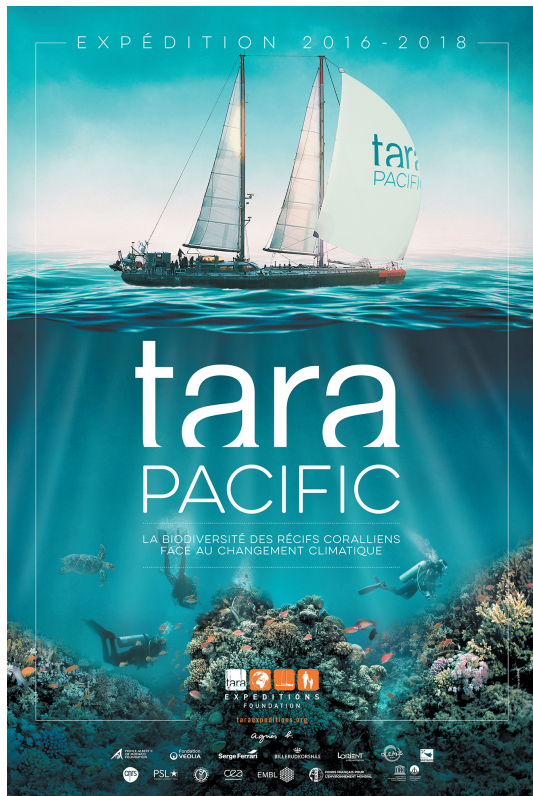


Figure I.37: The *Tara* Pacific expedition

In the context of bleaching events, the *Tara* Pacific expedition proposed to study coral reef ecosystems at the Pacific Ocean scale, which has never been performed before. The expedition in the Pacific Ocean began in May 2016 and finished in October 2018.

*Tara* traveled over 100.000 kilometers, during 2 years with different scientific teams on board. These teams are composed of researchers who work on different topics like biology, oceanography or bioinformatics.

To analyse the coral reef ecosystem, the *Tara* boat traveled in the Pacific Ocean from Panama to Japan during the first year and then from New Zealand to China during the second year.

The goals of this expedition were to study the biodiversity of coral reefs and their resilience to climate change and anthropogenic pressures. It was also to list and to quantify the micro-organisms living in and around corals, to analyse the population genetic structure of corals and micro-organisms living with them and to study the environment where corals live.

The expedition studied 40 archipelagos with identical protocols to collect, store and sequence samples, in order to understand the variations and the biodiversity of coral reefs, using comparative and interdisciplinary approaches.

In this expedition, researchers sampled two fishes species which are *Acanthurus tristegus* and *Zanclus cornutus* and three coral species which are *Millepora platyphylla*, *Pocillopora meandrina* and *Porites lobata* (**Figure I.38**). These corals and fishes were chosen because



their distribution areas cover all the Pacific Ocean, and because they are the subject of several studies at local scale.

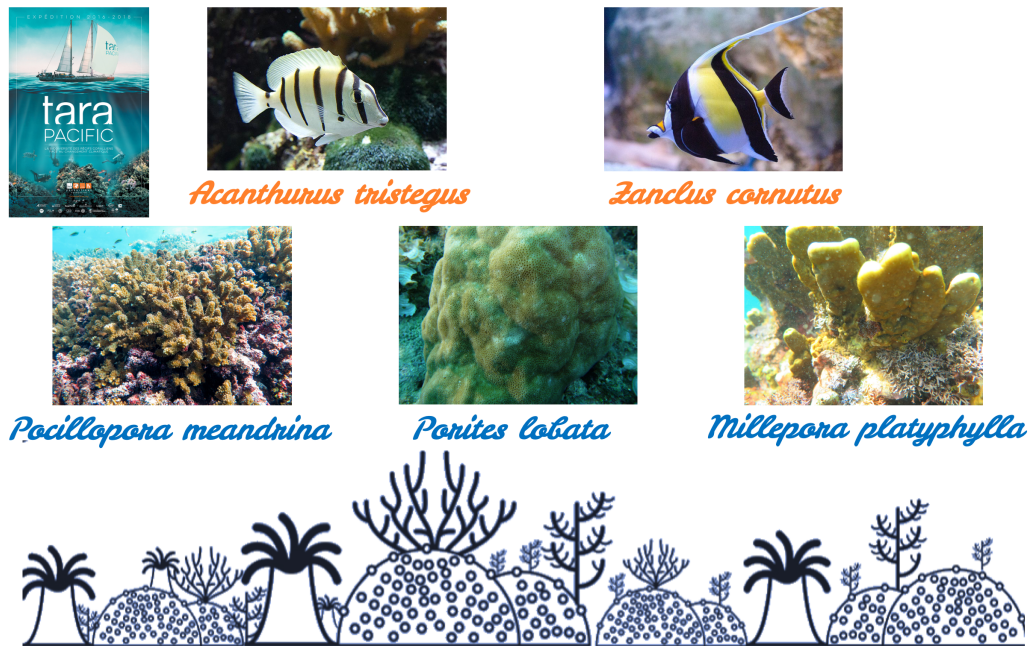


Figure I.38: Fishes and corals species sampled during the *Tara Pacific* expedition.

To collect samples, scientists used two protocols of sampling (**Figure I.39**) corresponding to:

- Continuous sampling: the boat contains instruments to measure physico-chemical parameters and nets of different mesh sizes to collect different samples of plankton with different size fractions.
- Fixed sampling: divers collect directly coral colonies, fishes and sea water around corals and measure environmental data (Temperature, pH, salinity, etc).

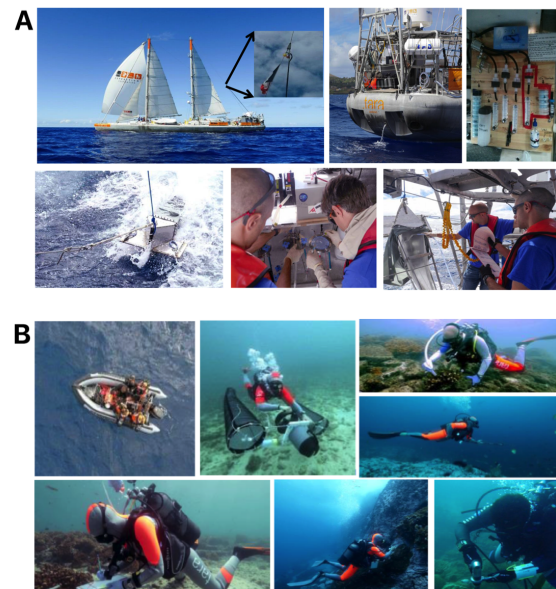


Figure I.39: Sampling processes during the *Tara Pacific* expedition.

**A.** Continuous sampling. **B.** Fixed sampling.

The sampling plan for coral species was set as follows: for each island, three sites were selected and in each site ten colonies of each coral species of interest were collected. Three of them were dedicated to transcriptomic analyses (**Figure I.40**).

Also, Coral Surrounding Water samples (CSW) were sequenced around *Pocillopora* sp. colonies. These samples were used to understand and to compare Symbiodiniaceae living inside and outside the host.

The Genoscope platform is in charge of sequencing of all the *Tara* Pacific samples. The sequencing concerns samples of corals with symbionts, fishes, plankton, sediments and environmental samples. The samples are sequenced at the Genoscope with metabarcoding, metatranscriptomic and metagenomic protocols.



Figure I.40: *Tara* Pacific sampling plan for each island with metatranscriptomic samples for each coral colonies.

## I.7 Aim of the thesis

Symbiodiniaceae are generally studied at the transcriptomic level in laboratories and/or at local scale. Thus, there is a lack of knowledge regarding the genetic diversity and gene expression of Symbiodiniaceae at the ocean scale.

The first question of the thesis was: what is the diversity of the symbiotic micro-algae of the coral in the Pacific Ocean, for three different coral species ?

I wondered if the diversity of Symbiodiniaceae across the Pacific Ocean is underestimated because most studies are focused on local diversity, at the scale of a single coral reef. In this way, I would be able to obtain a larger picture that would allow to better understand and to compare this diversity between different locations and between different scales.

To answer this question, I used genomic and transcriptomic references from different Symbiodiniaceae species, to determine the diversity present in the dual-transcriptomic samples of *Millepora sp.*, *Porites sp.* and *Pocillopora sp.*

For these three coral hosts, I determined the diversity in Symbiodiniaceae in the first 11 sequenced islands of the *Tara Pacific* expedition (**Figure II.4**), located in the South Pacific. This question is the core of the Chapter 2.



Figure I.41: *Tara Pacific* expedition map with islands sequenced in metatranscriptomic during the thesis indicated in pink.

In the context of coral bleaching events, literature clearly describes the role of Symbiodiniaceae. In consequence, understanding how Symbiodiniaceae adapted inside or outside of its host, to different environmental conditions and to different coral species would unravel the complex relationships between the symbiont and the host. Thus, the second main question of the thesis was: is there a specific adaptation of the symbiont to the host and to the environment ?

I answered this question by studying *Pocillopora sp.* coral and its symbiont *Cladocopium C1*, using different bioinformatics protocols and analyses. The results are presented in Chapter 3 and they are the subject of an article as co-first author included in this Chapter. Then, other analyses are presented in this Chapter linked to the other corals and also linked to other analyses.

Many studies have shown that Symbiodiniaceae can live outside its coral host. It is thus essential to understand how this micro-alga survives outside its host in the Pacific Ocean.

The surrounding water samples around the *Pocillopora sp.* colonies were sequenced in the first 11 islands from the *Tara* Pacific expedition. Using these samples, I asked the following questions: what is the diversity of the symbiotic micro-algae of the coral inside and outside of its coral host ? Is there a variation of gene expression between these two conditions ? If yes, what are the genes regulated for its survival outside its host ?

I answered these questions by studying *Pocillopora sp.* coral and its symbiont *Cladocopium C1* using the results obtained in Chapter 3, compared to the coral surrounding water samples. I used different bioinformatics protocols and analyses used before (Chapters 2 and 3) to compare these different conditions in terms of Symbiodiniaceae diversity and also to understand the genes which regulate the adaptation of *Cladocopium C1* in both conditions. The results are presented in Chapter 4.



## Chapter II

# Symbiodiniaceae diversity study in the *Tara Pacific* project

In this chapter, I tried to answer the question: what is the diversity of the symbiotic microalgae of the coral in the Pacific Ocean ?

To answer this question, I selected genomic and transcriptomic references from different populations of Symbiodiniaceae available from public databases. They were used to determine the diversity present in the dual-transcriptomic samples from 3 different corals (*Millepora sp.*, *Porites sp.* and *Pocillopora sp.*), in the first 11 islands sequenced from the *Tara Pacific* expedition.

To do this, an analysis of Symbiodiniaceae genomic and transcriptomic references gave rise to a first study to determine the potential contaminants in each reference.

Then, different bioinformatics methods were used on the different datasets to determine the Symbiodiniaceae diversity. A specific study of the *Cladocopium C1* diversity was made in the *Pocillopora sp.* samples.

## II.1 Materials and Methods

### II.1.1 Availability of Symbiodiniaceae transcriptomic and genomic references

With the advance of new sequencing technologies, there has been a considerable boom in genomic and transcriptomic references. Among these references, there are more than 30 transcriptomic and 8 genomic references of Symbiodiniaceae sequenced and assembled, and there is also one database available, which includes a multitude of Symbiodiniaceae ITS2 sequences.

These references were the subject of a first study made at the beginning of the thesis. The goal was to create a database of some various existing references of Symbiodiniaceae, to understand the genome and transcriptome differences and similarities between different clades and between different Symbiodiniaceae, and then use them to analyse the *Tara Pacific* data.

#### II.1.1.1 SymPortal: a database dedicated to Symbiodiniaceae ITS2 sequences

A database containing the ITS2<sup>1</sup> sequences of different Symbiodiniaceae was created by Benjamin Hume (**Hume *et al.* 2019**). This database gathers the following resources:

- **Published named sequences:** a fasta file containing sequences from the SymPortal database that have a name assigned to them. These sequences represent either sequences that the SymPortal database was seeded with when it was first created, or sequences that have been used to define an ITS2 type profile in a published study. Named sequences held in the SymPortal database that are only found in unpublished datasets are not included in this fasta file. It contains 1.314 published sequences from 3.203 samples from 19 published studies. 784 named but unpublished sequences were withheld from this release.
- **Published post-MED sequences:** a fasta file containing sequences from published samples that have been through the SymPortal quality control pipeline and Minimum Entropy Decomposition. It contains 8.367 sequences, from 3.203 samples reported in 19 published studies.

---

<sup>1</sup>Internal Transcribed Spacer 2

- **Published pre-MED sequences:** a fasta file containing sequences from published samples that have been through the SymPortal quality control pipeline but have not yet been through Minimum Entropy Decomposition. It contains 279.791 sequences from 3.203 samples reported in 19 published studies.

This database was first used to determine ITS2 profiles of Symbiodiniaceae present in the different dual-genomic datasets.

An ITS2 profile corresponds to a set of incorporated intragenomic sequences, leading to a core sequence, to understand specific combinations of intragenomic sequences (**Hume et al. 2019**). This study was done by Quentin Carradec and I used this analysis to compare Symbiodiniaceae diversity in the first *Tara Pacific* samples.

#### II.1.1.2 Symbiodiniaceae transcriptome references

28 transcriptomes of Symbiodiniaceae were selected in the bibliography (**Table II.1**). These 28 transcriptomes are available in public databases and were used to construct an integrative Symbiodiniaceae database. In each Symbiodiniaceae clade there are:

- *Symbiodinium* / *Clade A*: 6 references
- *Breviolum* / *Clade B*: 9 references
- *Cladocopium* / *Clade C*: 8 references
- *Durusdinium* / *Clade D*: 3 references
- *Effrenium* / *Clade E*: 1 reference
- *Fugacium* / *Clade F*: 1 reference

These transcriptomic references cover a large panel of the different clades of Symbiodiniaceae. They were used to determine Symbiodiniaceae diversity in the first coral dual-transcriptomic samples sequenced from the *Tara Pacific* expedition.



Table II.1: Symbiodiniaceae transcriptomes selected and used to analyse the Tara Pacific samples.

Clades / Species	Strains	Species	Hosts	Methods	Number of transcripts	Locations	Publications
A	CassKB8	<i>Symbiodinium natans</i>	<i>Cassiopeia</i>	Isolated and cultured	72.152	Kaneohe Bay Hawaii	Bayer <i>et al.</i> 2012
A1	CCMP2467	<i>Symbiodinium microadriaticum</i>	<i>Stylophora pistillata</i>	Isolated and cultured	198.297	Aqaba, Jordan	Baumgarten <i>et al.</i> 2013
A1	CCMP2467	<i>Symbiodinium microadriaticum</i>	<i>Stylophora pistillata</i>	Isolated and cultured	58.592	Aqaba, Jordan	Aranda <i>et al.</i> 2016
A2		<i>Symbiodinium corcolorum</i>	<i>Zoanthus sociatus</i>	Isolated and cultured	57.938	Caribbean	Rosic <i>et al.</i> 2015
A3	CCMP2430	<i>Symbiodinium tridacnorum</i>	<i>Tridacna maxima</i>	Isolated and cultured	51.448	Pacific Ocean	Lawrence <i>et al.</i> 2017
A3		<i>Symbiodinium sp.</i>	<i>Tridacna crocea</i>	Isolated and cultured	87.464	Okinawa Japan	Shoguchi <i>et al.</i> 2018
B	SSB01	<i>Breviolum muscatinei</i>		Cultured	59.669		Xiang <i>et al.</i> 2015
B		<i>Breviolum muscatinei</i>	<i>Anthopleura elegantissima</i>	Extracted	50.575	Orcas USA	Macrander <i>et al.</i> 2017
B	Mf1.05b	<i>Breviolum muscatinei</i>	<i>Montastraea faveolata</i>	Isolated and cultured	76.284	Florida	Bayer <i>et al.</i> 2012
B1	Mf1.05b	<i>Breviolum minutum</i>		Cultured	47.014		Shoguchi <i>et al.</i> 2013
B1	mac703/rt002/ Mf1.05b/rt351	<i>Breviolum minutum</i>		Cultured	51.199		Parkinson <i>et al.</i> 2016
B1	rt146 (CCMP3450)	<i>Breviolum pseudominutum</i>	<i>Oculina diffusa</i>	Isolated and cultured	47.411	North Atlantic	Parkinson <i>et al.</i> 2016
B19	mac04-487 (CCMP3448)	<i>Breviolum aenigmaticum</i>	<i>Porites</i>	Isolated and cultured	45.343	Florida	Parkinson <i>et al.</i> 2016
B19	HIAp/rt141/ Mf10.14b.02/ PurPFlex	<i>Breviolum psygmophylum</i>		Cultured	50.745		Parkinson <i>et al.</i> 2016
B2		<i>Breviolum psygmophylum</i>	<i>Oculina diffusa</i>	Isolated and cultured	87.171	Bermuda	Rosic <i>et al.</i> 2015
C1	CCMP829	<i>Cladocopium goreai</i>	<i>Acropora tenuis</i>	Cultured	52.421	Australia	MMETSP 2014
C1		<i>Cladocopium goreai</i>	<i>Acropora tenuis</i>	Isolated and cultured	116.479	MI, Australia	Levin <i>et al.</i> 2016
C1		<i>Cladocopium goreai</i>	<i>Acropora tenuis</i>	Isolated and cultured	131.066	SM, Australia	Levin <i>et al.</i> 2016
C1		<i>Cladocopium goreai</i>	<i>Discosoma sanctithomae</i>	Isolated and cultured	76.371	Jamaica	Rosic <i>et al.</i> 2016
C15		<i>Cladocopium sp.</i>	<i>Porites australiensis</i>	Extracted	26.627	Japan	Shinzato <i>et al.</i> 2014
C15		<i>Cladocopium sp.</i>	<i>Porites compressa</i>	Extracted	51.721	Pacific Ocean	MMETSP 2014
C3	Mp	<i>Cladocopium thermophilum</i>	<i>Mastigias papua</i>	Isolated and cultured	52.644	Pacific Ocean	Lawrence <i>et al.</i> 2017
C3k		<i>Cladocopium thermophilum</i>	<i>Acropora hyacinthus</i>	Isolated and cultured	26.986	Samoa	Ladner <i>et al.</i> 2012
D1		<i>Durusdinium glynni</i>	<i>Porites annae</i>	Isolated and cultured	68.957	Ist Pacific	Rosic <i>et al.</i> 2015
D1a		<i>Durusdinium glynni</i>	<i>Montastraea faveolata</i>	Cultured	56.916	Tennessee reef Atlantic	MMETSP 2014
D2		<i>Durusdinium sp.</i>	<i>Acropora hyacinthus</i>	Isolated and cultured	23.777	Samoa	Ladner <i>et al.</i> 2012
E	CCMP421	<i>Effrenium voratum</i>			77.821	Pacific Ocean	Lawrence <i>et al.</i> 2017
F	CCMP2468	<i>Fugacium kawagutii</i>	<i>Montipora verrucosa</i>	Cultured	18.489	Hawaii	MMETSP 2014

### II.1.1.3 Symbiodiniaceae genome references

8 Symbiodiniaceae reference genomes are available in the bibliography. Genomic coding sequences, abbreviated as CDS (CoDing Sequences), associated with 6 genomes were analysed (**Table II.2**). They were used to determine Symbiodiniaceae diversity in the first coral dual-transcriptomic and dual-genomic samples sequenced, from the *Tara Pacific* expedition.

Table II.2: Symbiodiniaceae genomes selected and used to analyse the *Tara Pacific* samples.

Clades	Strains	Species	Hosts	Methods	Number of CDS	Locations	Publications
A1	CCMP2467	<i>Symbiodinium microadriaticum</i>	<i>Stylophora pistillata</i>	Isolated and cultured	49.109	Aqaba, Jordan	Aranda <i>et al.</i> 2016
A3		<i>Symbiodinium sp.</i>	<i>Tridacna crocea</i>	Isolated and cultured	90.078	Okinawa Japan	Shoguchi <i>et al.</i> 2018
B1	Mf1.05b	<i>Breviolum minutum</i>		Cultured	47.014		Shoguchi <i>et al.</i> 2013
C1		<i>Cladocopium goreau</i>	<i>Acropora tenuis</i>	Isolated and cultured	35.913	MI, Australia	Liu <i>et al.</i> 2017
C15		<i>Cladocopium sp.</i>	<i>Porites lutea</i>	Extracted	22.508	Orpheus Island (GBR)	Robinson <i>et al.</i> 2019
F		<i>Fugacium</i>	<i>kawagutii</i>	Cultured	36.850	Hawaii	Lin <i>et al.</i> 2015

### II.1.2 GC content estimation and cleaning of Symbiodiniaceae references

To analyse the different Symbiodiniaceae genomic and transcriptomic references, I calculated the occurrence of each nucleotide A, T, G and C and also the number of undetermined bases (N or IUPAC Code<sup>2</sup>), to deduce the GC content. To do this, I programmed Python scripts and applied them, on each transcript and CDS, respectively from each reference transcriptome and genome.

After, I replaced letters other than A, C, T, G and N by N letters in each transcript or CDS, to have only letters that correspond to nucleotide letters (A, C, T, G and N). Transcripts or CDS with more than 90% of N content were removed to keep only transcripts or CDS with significant nucleotide information.

### II.1.3 Detection of potential contamination in Symbiodiniaceae references

To estimate the level of contamination in each transcriptome and genome, I performed a BLASTx of Symbiodiniaceae transcripts and CDS, against large protein databases by carrying

<sup>2</sup>International Union of Pure and Applied Chemistry (Stothard *et al.* 2000)

out a taxonomic analysis of the transcripts and the CDS.

I first performed a BLASTx only on the reference transcripts using the Uniref90 database (version 09/2018). Then, I performed a BLASTx on reference CDS and transcripts using a compilation of different marine databases created by Quentin Carradec and Benjamin Noel at the Genoscope (version 09/2018).

**BLAST (Basic Local Alignment Search Tool)** is a method used in bioinformatics, to find similar regions between at least two or many sequences of nucleotides or amino acids, by doing a multiple alignment of these sequences. It uses a query sequence and searches in databases for sequences with high similarity and classifies results using statistics. BLASTx option takes nucleotide sequences as input, translates these sequences in the 6 frames and aligns them against a protein database.

First, I used the Uniref90 database containing 557.491 proteins from the Swiss-Prot database and 111.678.811 proteins from the EMBL database clustered at 90% sequence identity and 80% overlap of identity with the longest sequence of the given cluster. This database contains only one reference for a Symbiodiniaceae proteome, that is *Symbiodinium microadriaticum* (clade A).

Then, I used a database created at the Genoscope, in September 2018 containing data from:

- MMETSP<sup>3</sup>: a collection of marine transcriptomes actualized by **Johnson *et al.* 2016** - [https://figshare.com/articles/dataset/Marine\\_Microbial\\_Eukaryotic\\_Transcriptome\\_Sequencing\\_Project\\_re-assemblies/3840153/3](https://figshare.com/articles/dataset/Marine_Microbial_Eukaryotic_Transcriptome_Sequencing_Project_re-assemblies/3840153/3)
- Marine Genomic database: from Okinawa Institute of Science and Technology Graduate University - <http://marinegenomics.oist.jp>
- The Echinobase: from **Cary GA *et al.* in 2018** - <https://www.echinobase.org/entry/>
- Reef Genomics: <http://reefgenomics.org/>
- Refugee 2020: from **Voolstra C. *et al.* 2015** - <http://refuge2020.reefgenomics.org/>
- NR: Database from NCBI<sup>4</sup>

<sup>3</sup>The Microbial Eukaryotic Transcriptome Sequencing Project (**Keeling *et al.* 2014**)

<sup>4</sup>National Center for Biotechnology Information

```
ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/accession2taxid/prot.accession2taxid.  
gz
```

The best **BLASTx** alignment was selected for each Symbiodiniaceae transcript or CDS, and then a taxonomic affiliation was given.

To simplify the analyses, the taxonomic affiliation levels were summarized using some taxonomic levels chosen (**Figure II.1**). When the transcripts did not match any Uniref90 proteins or any other protein databases, the taxonomic name given is noted "ID-NOT-FOUND".

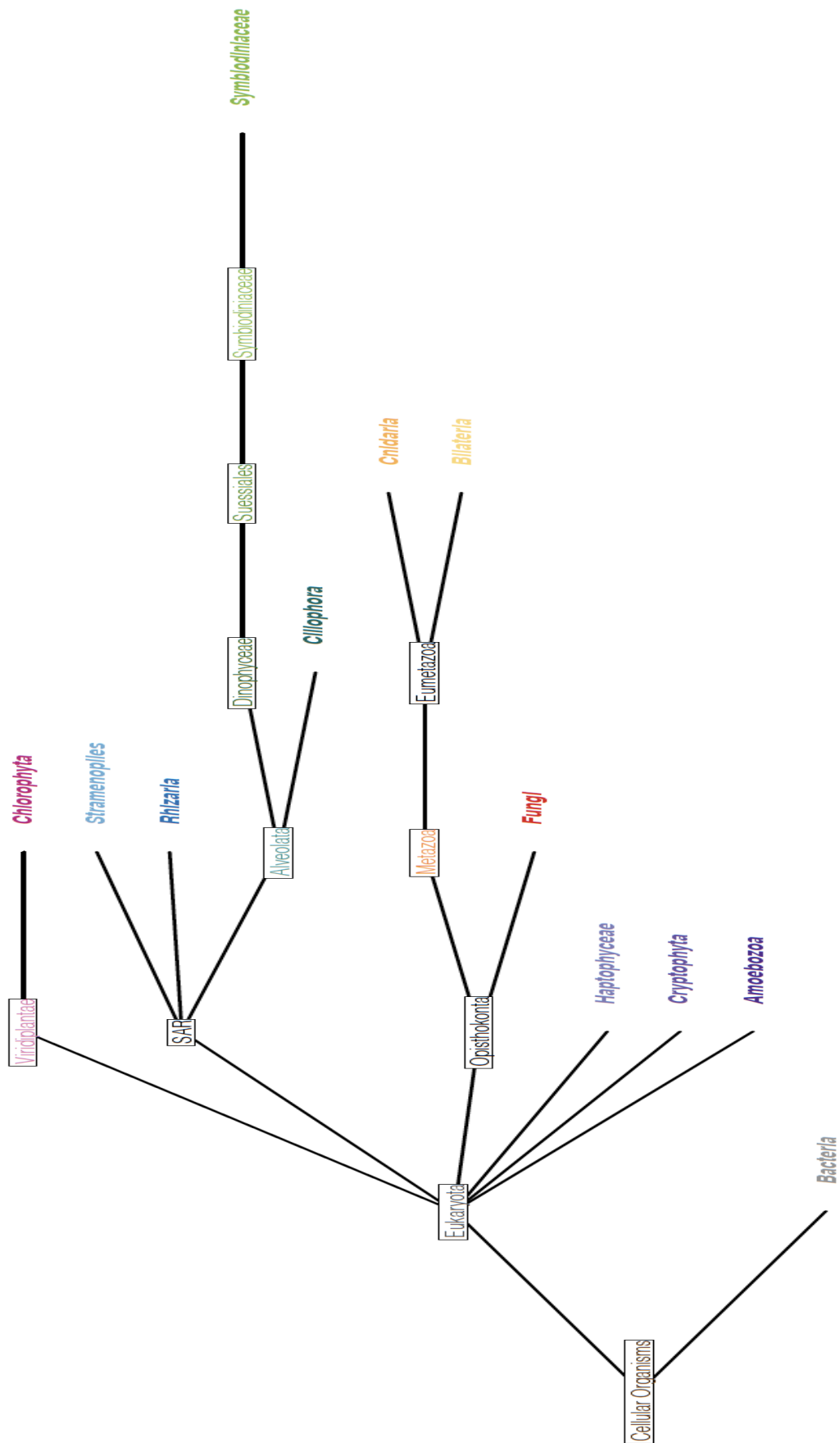


Figure II.1: Phylogenetic tree of the taxonomic lineages selected, to summarize the taxonomic affiliation results.

## II.1.4 Coral sampling for RNA sequencing

During the *Tara Pacific* expedition, researchers sampled three coral species (*Millepora sp.*, *Porites sp.* and *Pocillopora sp.*), and among these species, I studied the symbiotic micro-algae living in each sample of these different corals.

### II.1.4.1 Coral sampling

In the *Tara Pacific* expedition, three coral species, *Millepora platyphylla*, *Pocillopora meandrina* and *Porites lobata* were chosen for the studies (**Figure II.2**). These corals were chosen for sampling because they are present and abundant across the Pacific Ocean, and therefore represent important players in the study of the coral reef ecosystem.



Figure II.2: Corals species sampled during the *Tara Pacific* expedition.

All the coral samples were sent to the Genoscope for storage and sequencing. During this thesis, I was interested in "dual-transcriptomic" data (**Figure II.3**). The sampling plan for these samples was as follows: for each island, 3 sites were selected and in each site 10 colonies of each coral of interest were collected.

For each coral, researchers in charge of the sampling had difficulties to identify the corals initially chosen in the frame of the *Tara Pacific* expedition, that is why, I will talk about the genera and not the species of the different coral, in the following parts of the manuscript.



Figure II.3: The *Tara Pacific* sampling plan for each island and dual-transcriptomic samples using the three different corals.

During the thesis and with the *Tara* consortium decisions, sequencing of metatranscriptomic data from 11 islands were done (Figure II.4), for *Pocillopora sp.* and *Millepora sp.* For *Porites sp.*, only the samples from Coïba, Gambier, Samoa and Guam islands were sequenced because, there are rRNA<sup>5</sup> contaminations for this coral which require development of new protocols by the production team at the Genoscope.



Figure II.4: The *Tara Pacific* expedition map with islands sequenced in metatranscriptomic during the thesis (indicated in pink).

<sup>5</sup>ribosomal RNA

For each sample, around 10g of coral were sampled and placed in 15mL Lysing Matrix A tube (reference: 116930050 by MP Biomedicals). Then DNA/RNA shield was added (reference: R1100-250 by Zymo Research) and the samples were conserved at  $-20^{\circ}\text{C}$ .

#### II.1.4.2 DNA and RNA extractions

Coral samples were subjected to mechanical grinding using a FastPrep instrument (reference: 116005500 by MP Biomedicals) and then separated in 10 aliquots of 500 micro-liters for the different sequencing experiments. In fact, using the sampling plan of the 10 colonies for each coral sampled by site and by island, the 10 colonies were used for barcoding experiments while transcriptomic and genomic sequencing were done on 3 of them.

In each aliquot proteinase K was added to digest proteins and to remove contaminants. Then DNA and RNA were co-extracted and co-purified using ZR-Duet DNA/RNA MiniPrep Plus (reference: D7003, Zymo Research) for one aliquot.

All DNA and RNA extractions of coral samples were performed at the Genoscope by the Production team.

#### II.1.4.3 mRNA selection and libraries preparation

The TruSeq - Stranded Poly A libraries were used to prepare and to index libraries using RNA samples. This protocol was set up and applied at the Genoscope on the islands of Coïba, Gambier, Guam and Samoa for the three corals *Millepora sp.*, *Porites sp.* and *Pocillopora sp.*

The RNA NEB U2 - Stranded PolyA libraries were used to prepare and to index libraries using RNA samples and were applied on all the other samples.

All mRNA<sup>6</sup> selection and library preparations were performed at the Genoscope by the Production team.

#### II.1.4.4 Illumina sequencing

188 cDNA libraries were sequenced in paired-end with 151 nucleotides using Illumina sequencer Hi-Seq 4000 or NovaSeq 6000 (**ANNEXE II - Figure 1**).

---

<sup>6</sup>messenger RNA



Sequenced reads were filtered with a workflow created by the LBGB team (Laboratory of bioinformatics for Genomic and Biodiversity) at the Genoscope (**Figure II.5**).

A short name was given to each sample to summarize the origin of the sample, like for example: sample **I02S01C010MIL** indicating that the sample is from colony 10 (C010) from the first site (S01) of Coïba (I02), and the species is *Millepora sp.* (MIL).

All the samples sequenced and used in metatranscriptomic, during the thesis are presented in the **Table II.3**. The number of available samples differs slightly from the initial sequencing plan. Indeed, Rapa Nui island has 4 sampling sites and Malpelo island has only one sampling site. For the *Millepora* coral, the targeted species is extinct in all the East Pacific Ocean. The two samples from Coïba island are samples of *Millepora intricata* instead of *Millepora platyphylla*.



### II.1.5 Bioinformatics methods to align dual-transcriptomic reads on Symbiodiniaceae references

After the construction of a database with several genomic and transcriptomic references of Symbiodiniaceae and the sequencing of the dual-transcriptomic samples, I wondered what is the diversity of Symbiodiniaceae in the *Tara Pacific* samples? To answer to this question I used bioinformatics methods and I created a pipeline to align samples against the Symbiodiniaceae references.

Alignment is a method used in bioinformatics analyses to compare two data models using an algorithm. First, I aligned paired-end reads of the 188 dual-transcriptomic samples of corals, against the 28 reference transcriptomes using different protocols such as:

1. Burrows-Wheeler Aligner algorithm (BWA - version 0.7.15) to align paired-end reads of the 188 coral samples against the 28 reference transcriptomes. This tool maps short sequences against a reference genome or transcriptome. This algorithm was published in **1994 by Michael Burrows and David Wheeler** and reviewed by **Li *et al.* in 2009**.
2. Aligned reads were sorted with Samtools version 1.6 (samtools -sort).
3. Then, I selected aligned reads with different parameters such as:
  - A percentage of identity over 70% and a coverage over 30 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept.
  - A percentage of identity over 95% and a coverage over 50 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept.
  - A percentage of identity over 98% and a coverage over 80 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept.
4. After, I wrote a script using Python, to quantify the number of reads aligned on each transcript and to give: **The name of the transcript reference; the average of the read coverage; the mean identity percent; the number of reads aligned; the length of the transcript and the origin of the sample.**
5. Finally, this protocol was applied for some of the different conditions such as:

- All reference transcriptomes (**Table II.1**) against all dual-transcriptomic samples (**Table II.3**).
- Reference CDS of *Symbiodinium A3* (**Table II.2**) against dual-transcriptomic samples of *Millepora sp.* (**Table II.3**).
- Reference CDS of *Cladocopium C1* (**Table II.2**) against dual-transcriptomic samples of *Pocillopora sp.* (**Table II.3**).
- Reference transcriptome of *Durusdinium D2* (**Table II.1**) against dual-transcriptomic samples of *Pocillopora sp.* (**Table II.3**).
- Reference CDS of *Cladocopium C1* (**Table II.2**) and reference transcriptome of *Durusdinium D2* (**Table II.1**) against dual-transcriptomic samples of *Pocillopora sp.* (**Table II.3**).
- Reference CDS of *Cladocopium C15* (**Table II.2**) against dual-transcriptomic samples of *Porites sp.* (**Table II.3**).

### II.1.6 Population structure of *Cladocopium C1*

Among the *Pocillopora* samples, I performed a population study of *Cladocopium C1*. These samples were chosen during the thesis because *Pocillopora* coral was the first genome assembled from the *Tara Pacific* expedition.

Moreover, dual-transcriptomic samples associated were the easiest samples to sequence compared to *Porites sp.* samples (See previously). Also, the fact that I had larger datasets for this species had also guided the research on this coral compared to *Millepora sp.* samples where some islands were missed (See previously).

To carry out this study, I first used the work of Benjamin Hume on metabarcoding samples, which concerns the study of Symbiodiniaceae populations in the *Tara Pacific* samples, using a database that he created: SymPortal (**Hume et al. 2019 and See previously**).

Then, to carry out a population study on this species, I first used the metagenomic samples, but these samples did not prove to be conclusive during the thesis.

Thus, to solve this problem, metatranscriptomic samples were used to perform the population structure of *Cladocopium C1*, that has still never been done at the ocean scale and for this kind of study.

### II.1.6.1 Metagenomic samples used to understand *Cladocopium C1* diversity

To do the specific study of *Cladocopium C1* populations, I looked at the metagenomic samples from the *Tara Pacific* expedition (**Table II.4**).

Table II.4: Number of metagenomic samples per island and site from *Pocillopora sp.* samples sequenced at the Genoscope.

	I01 - Las perlas			I02 - Coiba			I03 - Malpelo	I04 - Rapa Nui				Total	
	S01	S02	S03	S01	S02	S03	S01	S01	S02	S03	S04	Total	
<i>Pocillopora sp.</i>	3	5	3	4	3	3	10	3	3	3	3	43	
	I05 - Ducie			I06 - Gambier			I07 - Moorea			I08 - Cook			Total
	S01	S02	S03	S01	S02	S03	S01	S02	S03	S01	S02	S03	Total
<i>Pocillopora sp.</i>	3	3	3	3	3	3	3	3	3	3	3	3	36
	I09 - Niue			I10 - Samoa			I15 - Guam			Total	Total		
	S01	S02	S03	S01	S02	S03	S01	S02	S03	Total	Total		
<i>Pocillopora sp.</i>	3	3	3	3	3	3	3	3	3	27	106		

Using these metagenomic samples I performed an alignment as presented previously, according to the following protocol: BWA algorithm was used to map short sequences against reference genome of *Cladocopium C1* (**Table II.2**) and reference transcriptome of *Durusdinium D2* (**Table II.1**). Then the alignments were sorted using Samtools and I selected aligned reads with different parameters as a percentage of identity over 98% and a coverage over 80 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept. After, the Python script used on the previous alignment to calculate on each read aligned and for each contig some statistics was applied.

The use of two different references from two different genera was to avoid non-specific alignment on samples with both Symbiodiniaceae genera. Samples with a majority of *Durusdinium* were removed from the analysis and *Durusdinium* transcripts were also removed to avoid biasing on the results to do population structure of *Cladocopium C1*. Then, to carry out the population structure of *Cladocopium C1*, I first looked at the average coverage per contig, to determine if the coverages were sufficient to obtain allele frequencies and therefore SNPs.

### II.1.6.2 Metatranscriptomic samples: a new kind of data to do population structure

Following the results obtained with the metagenomic data, we decided to use metatranscriptomic samples to perform the population structure of *Cladocopium C1*, using the following

protocol:

1. I used the reads aligned on the *Cladocopium C1* CDS and *Durusdinium D2* transcripts against dual-transcriptomic samples. Then, I selected aligned reads with different parameters as a percentage of identity over 98% and a coverage over 80 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept.
2. I used the Genome Analysis ToolKit program (GATK, version 3.7.0 developed by **Van der Auwera and O'Connor in 2020**) following the best practices guide for variant discovery<sup>7</sup> as described in the article **Armstrong, Lê-Hoang et al.**

%

```
%Indexation of Cladocopium C1 CDS reference and Durusdinium D2 transcripts  
(picardtools v2.6.0, CreateSequenceDictionary). % Identification of realignment targets  
(GATK RealignerTargetCreator). % Realignment around detected indels (GATK  
IndelRealigner). % Variants calling for each colony individually (GATK Haplotype-  
Caller) using a ploidy equal to 1 to generate the variant call files (vcf). % Elimination  
of Durusdinium D2 transcripts. % The results were merged into island-specific cohorts  
(GATK CombineGVCFs) and after were merged using a perform joint genotyping  
across all 11 islands (GATK GenotypeGVCFs). %
```

3. Then, to avoid cross-contamination between different Symbiodiniaceae, samples with a majority of the *Durusdinium* genus or with 2 different *Cladocopium* based on the ITS2 analysis were removed.
4. SNPs were filtered using VCFtools (version 0.1.12 from **Danecek et al. in 2011**) to include only biallelic sites with a quality score  $\geq 30$ .
5. I made a python script to filter the SNPs obtained and kept only the biallelic SNPs with a coverage greater than or equal to 4 and then calculated the allelic frequencies.
6. A Hierarchical clustering (hclust function in R) was used with the complete linkage option to obtain a classification of *Cladocopium C1* samples according to their SNP frequencies.

Moreover, the average coverage per CDS was calculated and compared to the results obtained with metagenomic samples.

<sup>7</sup> <https://gatk.broadinstitute.org/hc/en-us/articles/360035890431-The-logic-of-joint-calling-for-germline-short-variants>

To finish, these results were then compared with the results obtained with the ITS2 profiles (carried out by Benjamin Hume and Quentin Carradec) and with the population structure determined in the host organism *Pocillopora sp.* (carried out by Eric Armstrong and Didier Forcioli).

## II.2 Results

### II.2.1 Detection of the contaminations in Symbiodiniaceae references

In a first analysis, Symbiodiniaceae references were studied to detect potential contaminants and to keep only the cleanest references using GC content.

As the mean GC content is homogenous within a species and variable between species, it was used to determine if datasets from genomes or transcriptomes were contaminated. In fact, if the repartition of the GC content for one reference diverged, it often means that there is a contamination from another organism in the reference given.

The results obtained using this method were not sufficient to determine the potential contaminations of the Symbiodiniaceae references (**APPENDIX III - Figures 2, 3 and 4**). These results were compared with the article from **Gonzales *et al.* in 2017** providing 24 Symbiodiniaceae reference transcriptomes. I concluded that in these different cases, GC content is not a good strategy to detect contaminants, that is why, we decided to use another method on Symbiodiniaceae references to detect potential contaminations.

To detect the potential contaminants in each Symbiodiniaceae reference, I performed a first time, a Blastx of all Symbiodiniaceae transcriptomes against the protein database Uniref90 and kept the best match for each transcript.

A taxonomic identification was obtained for 1.125.826 Symbiodiniaceae transcripts and these transcripts were affiliated to taxonomic levels (**Figure II.1**). 520.239 transcripts did not match significantly any Uniref90 proteins and were labelled "ID-NOT-FOUND" (**Figure II.6**).

In all transcriptomes, between 40% and 90% of transcripts were, as expected, affiliated to Symbiodiniaceae. From 10% to 50% of the transcripts were "ID-NOT-FOUND" meaning they are unknown in the Uniref90 database. This result was expected because the Uniref90 database contains a partial proteome of *Symbiodinium microadriaticum* from clade A.

10% and 25% of the transcripts of *Cladocopium-C15-Pau* and *Cladocopium-C15-Pco* respectively were affiliated to *Cnidaria* phylum, suggesting a strong coral contamination for these 2 transcriptomes. Indeed, these transcriptomes come from environmental samples and RNA from Symbiodiniaceae and its coral host were extracted, sequenced and assembled together though dual-transcriptomic. Host and symbiont transcripts were then separated by bioinformat-



ics methods that seem poorly efficient in this case. For the other reference transcriptomes, the number of contaminated transcripts represents a small proportion of the transcriptomes.

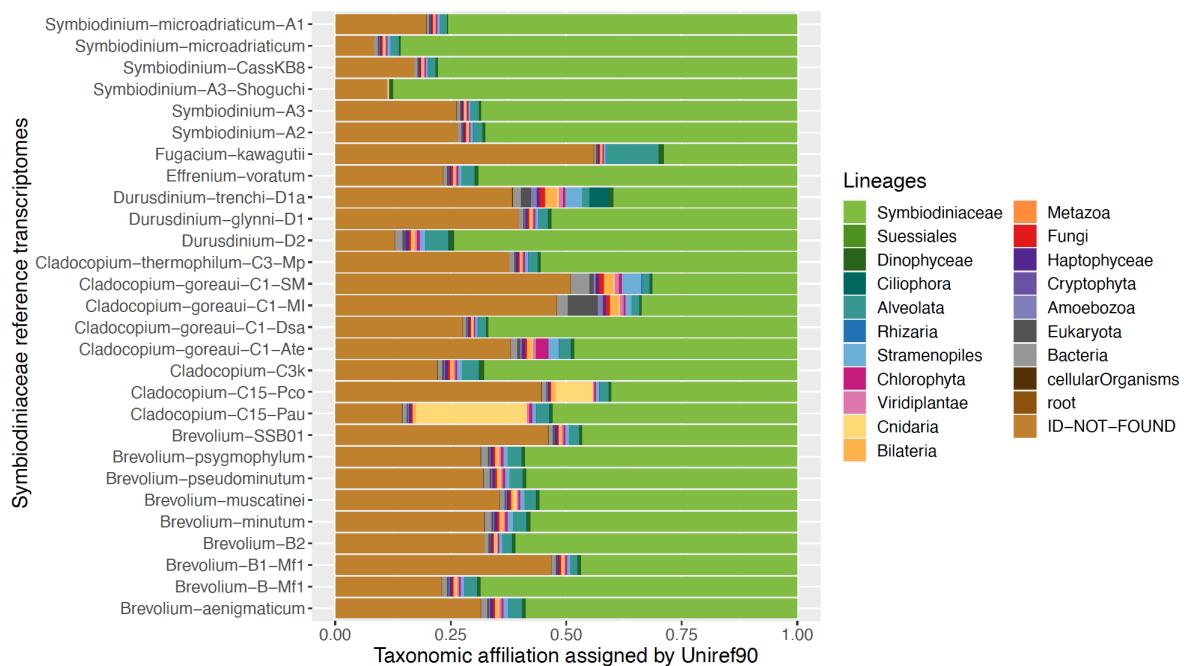


Figure II.6: Taxonomic affiliation of transcripts from the 28 Symbiodiniaceae reference transcriptomes selected, using Uniref90 database.

Each color represents the fraction of the transcripts assigned to a taxonomic level selected (**Figure II.1**). The legend for each color is indicated on the right. "ID-NOT-FOUND" means that the transcript did not match any protein.

The same analyses were done using another database created at the Genoscope (**Figure II.7**) and a significant match was obtained for 1.638.818 transcripts that were affiliated to taxonomic levels (**Figure II.1**). The 302.944 remaining transcripts did not significantly match to this protein database and were labelled "ID-NOT-FOUND".

In this analysis there are between 50% and 90% of transcripts which were as expected affiliated to Symbiodiniaceae. From 5% to 25% of the transcripts were "ID-NOT-FOUND" meaning they are not found in this database.

Compared to the previous analyses, there were between 5% and 25% of the transcripts, which were assigned to *Cnidaria* phylum, once again suggesting a coral contamination. In these cases, co-extraction of RNA with the host and the symbiont, must have contaminated these transcriptomes. I can also assume that the databases used and selected also present certain contaminations. Indeed, these databases contain exclusively marine organisms and

incorrect assignments in some transcripts can lead to these results. I also observed fewer contaminants for these analyses compared to the previous ones. There were also fewer sequences that were not assigned compared to Uniref90 taxonomic analyses.

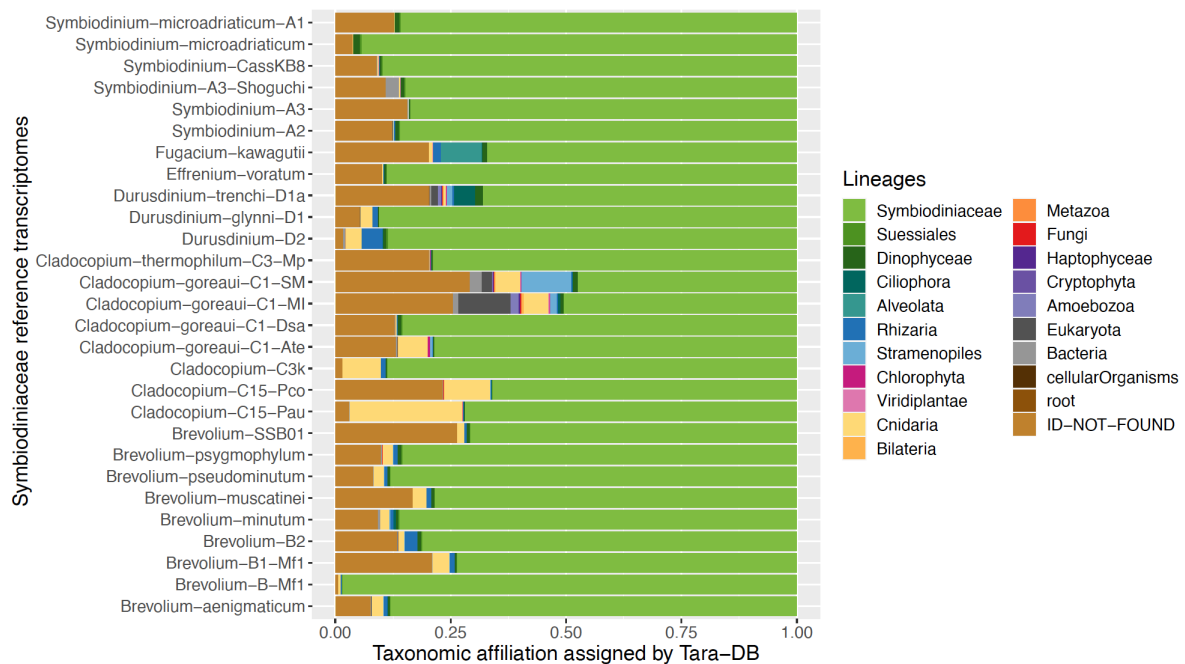


Figure II.7: Taxonomic affiliation of transcripts from the 28 Symbiodiniaceae reference transcriptomes, using a database created at the Genoscope.

Each color represents the fraction of the transcripts assigned to a taxonomic group selected (**Figure II.1**). The legend for each color is indicated on the right. "ID-NOT-FOUND" means that the transcript did not match any protein.

Then, I did the same analysis for the Symbiodiniaceae CDS from genome references, using the database created at the Genoscope. A significant match was obtained for 314.371 CDS that were affiliated to taxonomic levels (**Figure II.1**). 25.452 remaining transcripts did not match significantly with the protein database and were labelled "ID-NOT-FOUND".

I observed a higher assignment to Symbiodiniaceae for the 6 genomes compared to transcriptomes, with between 90% to 99% of CDS assigned to this taxonomic level. There were also between 1% and 10% of CDS assigned to "ID-NOT-FOUND".

To compare to transcriptome results, there is a higher assignment to Symbiodiniaceae family, when I used CDS sequences. From these results, we decided to flag CDS and transcripts to be able to recognize them in further analyses.

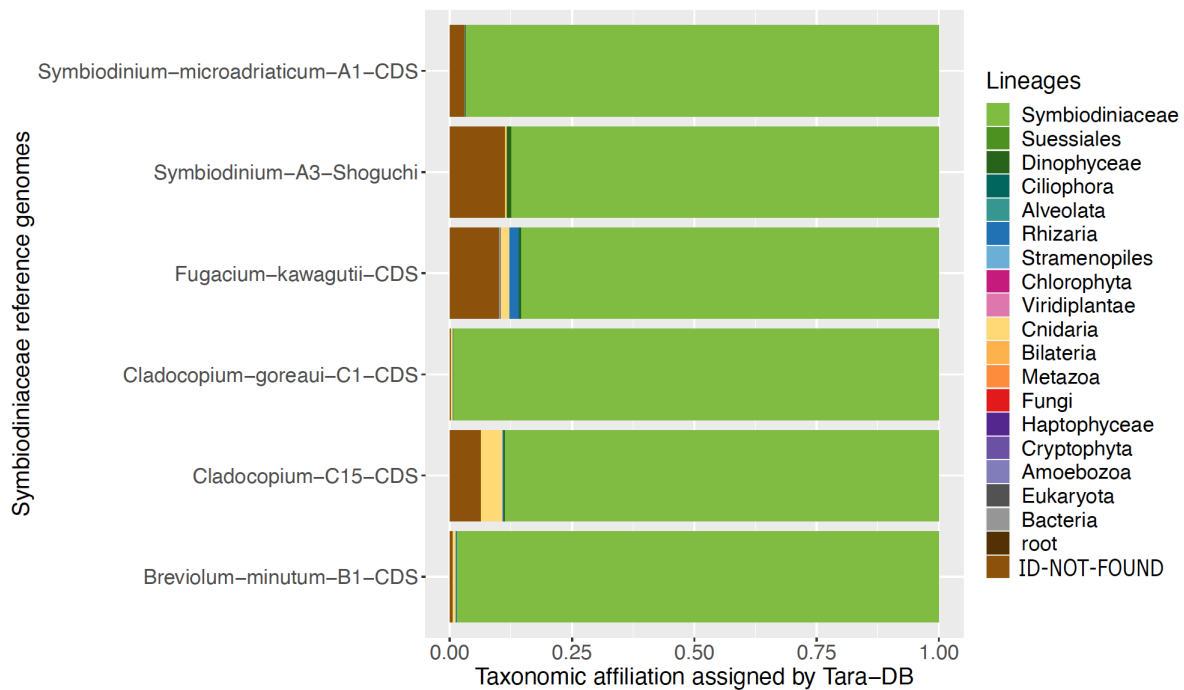


Figure II.8: Taxonomic affiliation of CDS from 6 Symbiodiniaceae reference genomes, using a database created at the Genoscope.

Each color represents the fraction of the CDS assigned to a taxonomic level selected (**Figure II.1**). The legend for each color is indicated on the right. "ID-NOT-FOUND" means that the transcript did not match any protein.

## II.2.2 Estimating Symbiodiniaceae diversity in the *Tara Pacific* coral samples using all Symbiodiniaceae reference transcriptomes

In order to identify Symbiodiniaceae species present in the first 11 islands from the *Tara Pacific* expedition, dual-transcriptomic samples reads were aligned simultaneously on all Symbiodiniaceae transcriptome references studied previously.

A first analysis with low-filtered parameters was done (**Appendix IV - Figure 7**). Then, the same alignments were done, using a percentage of identity over 95% and a coverage over 50%, with all Symbiodiniaceae reference transcriptomes against all dual-transcriptomic samples (**Figures II.9, II.10 and II.11**).

For *Pocillopora sp.* samples in the East Pacific (Las Perlas, Coïba and Malpelo islands and one sample in Guam island), reads were aligned on two Symbiodiniaceae genera (*Cladocopium* and *Durusdinium*). These results indicate that at least two species are present in these samples. According to the coral colony, either only *Durusdinium*, only *Cladocopium* or both

genera are detected. In the Central and West Pacific, the reads were only aligned on *Cladocopium*.

However the *Breviolum* clade is unknown for this coral host in this location. Looking at the associated transcriptomic reference corresponding to *Breviolum-Odi-D2*, it appeared that this reference was contaminated by another transcriptome from *Durusdinium* clade. Indeed, in the protocol used to carry out this reference, there is a mixture of these two species. I therefore assumed that this reference was contaminated and could not be used for the rest of the analyses.

For *Porites sp.* samples scattered on four different islands across the Pacific Ocean (Coïba, Gambier, Samoa and Guam islands), reads were aligned on two transcriptomes of *Cladocopium* genus. These two transcriptomes are assigned to one species, *Cladocopium-C15* which is distinct from the other reference of *Cladocopium* used. These results are coherent with the bibliography.

Finally, the results of the alignment of the *Millepora sp.* transcriptomic samples on Symbiodiniaceae reference transcriptomes were analysed. In these samples, I observed for the islands of Guam and Samoa a correspondence with some transcriptomes of the *Cladocopium* genus, more precisely affiliated to *Cladocopium C1* transcriptomes. On the other hand, Coïba, Gambier, Moorea, Cook and Niue islands, reads were aligned to *Symbiodinium* and *Durusdinium* from clade A and clade D genera respectively.

Looking further, there is an assignment for the two reference transcriptomes corresponding to the species *Symbiodinium-A3* and *Durusdinium-trenchi-D1a-Mfa*. Looking at this last reference in more detail, it turned out that the protocol applied to this reference could have resulted in contamination. Indeed, obtaining a transcriptome of the *Durusdinium* genus in the *Millepora sp.* coral host in this location, had never been proven in the bibliography. Moreover, there are known contaminants in the MMETSP<sup>8</sup> database where this reference was provided.

After the obtaining of these results, I focused more on the precise contaminations by the transcriptomes used, to obtain if possible a reference of Symbiodiniaceae for each dual-transcriptomic sample. To do this, I summarized contaminations of transcriptomes in different tables (**APPENDIX V - Tables 1, 2 and 3**).

---

<sup>8</sup>The Marine Microbial Eukaryote Transcriptome Sequencing Project (Keeling *et al.* 2014)

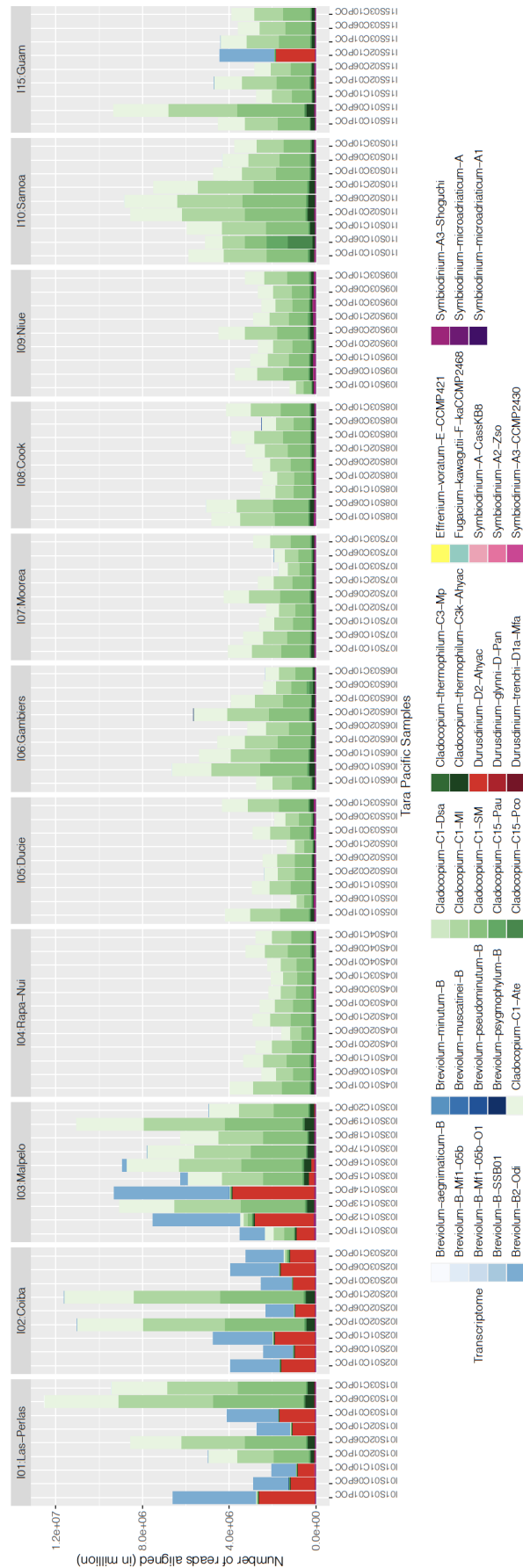


Figure II.9: Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes.

Each color corresponds to a specific transcriptome. Similar colors are transcriptomes of the same Symbiodiniaceae genus in *Pocillopora sp.* samples.

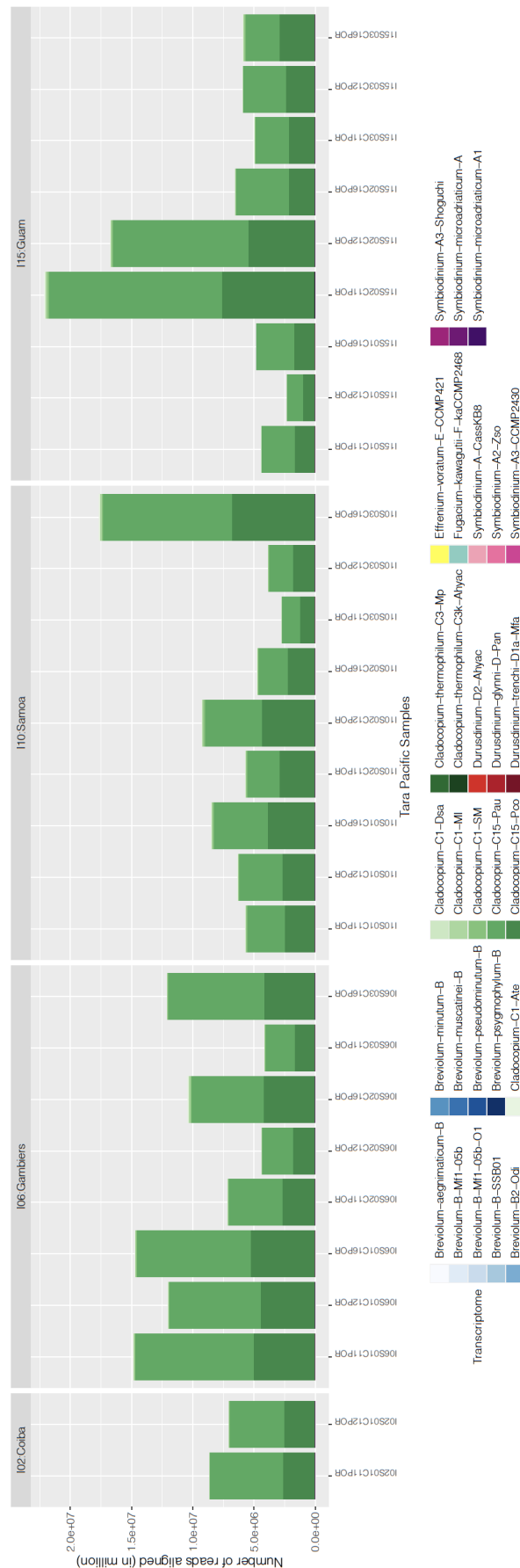


Figure II.10: Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes.

Each color corresponds to a specific transcriptome. Similar colors are transcriptomes of the same Symbiodiniaceae genus in *Porites sp.* samples .

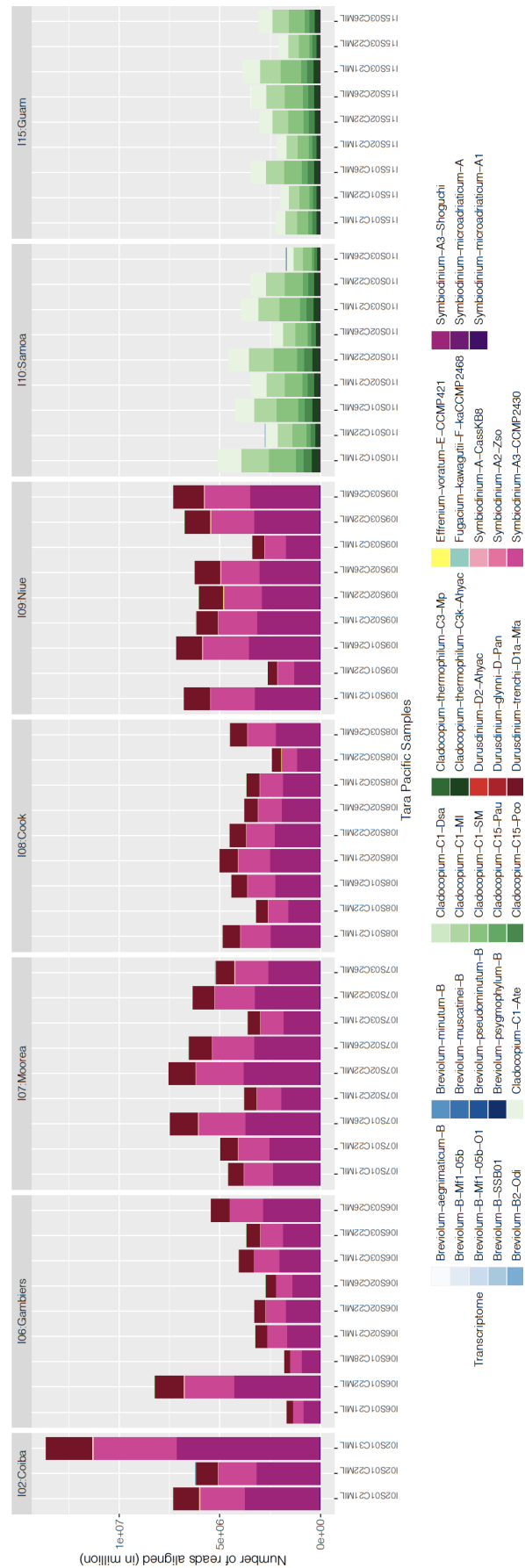


Figure II.11: Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes.

Each color corresponds to a specific transcriptome. Similar colors are transcriptomes of the same Symbiodiniaceae genus in *Millepora sp.* samples.

For the *Millepora sp.* coral (**APPENDIX V - Table 1**), a maximum number of reads were aligned to the reference *Symbiodinium-A3-Shoguchi*, followed by *Symbiodinium-A3-CCMP2430*. Similar results were obtained with the reference *Durusdinium-trenchi-D1a-Mfa*, which confirmed the hypothesis of a contamination of this transcriptome. For each of these references, the average percentage identity per transcript was very high (greater than 99%) as well as the average coverage per transcript was also high (greater than 50%). In order to study symbiont contained in *Millepora sp.* for all islands except Samoa and Guam islands, a corresponding transcriptomic reference to *Symbiodinium-A3-Shoguchi* is available.

The average identity percentages of reads aligned on *Cladocopium C1* transcriptomes were around 97%. These results suggest that the *Cladocopium* present in *Millepora sp.* samples is different from all Symbiodiniaceae reference transcriptomes selected and should be another *Cladocopium* species.

For the *Porites sp.* samples (**APPENDIX V - Table 2**), the majority of reads aligned with two references belonging to the same species *Cladocopium-C15-Pco* and *Cladocopium-C15-Pau*. These references showed an average percentage identity per transcript of 99% and an average coverage per transcript between 58-75%. These results showed that for the study of these samples, there are two reference transcriptomes available.

Finally, for the *Pocillopora sp.* samples (**APPENDIX V - Table 3**), the majority of reads aligned to the following transcriptomic references *Cladocopium-C1-SM*, *Cladocopium-C1-MI* and *Cladocopium-C1-Ate*. These three references belong to the same Symbiodiniaceae species, named *Cladocopium C1*. These results showed that references are available to study these samples. In addition, the percentages of average identity and average coverage per transcript were high, around 99% for identity and between 59 and 68% for coverage.

Looking at the *Breviolum-B2-Odi* reference transcriptome, a 99% average percentage identity per transcript and a 83% average coverage per transcript were observed. These results confirmed the contamination of this transcriptomic reference by *Durusdinium* species.

For samples containing *Durusdinium*, the reference with a maximum of aligned reads corresponds to the reference *Durusdinium-D2-Ahyac* with 99% average identity per transcript and 78% average coverage per transcript.

These results for the *Pocillopora sp.* samples, revealed that 3 different references belonging to the species *Cladocopium-C1* are available to do the transcriptomic analyses. There is also



the transcriptomic reference of *Durusdinium-D2-Ahyac* for samples containing both species or only this species in Las Perlas, Coïba, Malpelo and Guam islands.

Thus, for the rest of the analyses these results were used to represent the corresponding species in each of the dual-transcriptomic samples for each different coral.

After having determined the species present in each of the dual-transcriptomic samples and used Symbiodiniaceae genomic and transcriptomic references available, I did specific alignments of the dual-transcriptomic reads, using the closest genomes or transcriptomes available (**Figures II.12, II.13 and II.14**).

For *Pocillopora sp.* samples, I aligned reads simultaneously on *Cladocopium C1* CDS from the reference genome and *Durusdinium-D2* reference transcriptome, to avoid non-specific reads aligned. Indeed, in the case where the samples include both species, I wanted to have a more specific distribution of the reads for closely homologous genes in Symbiodiniaceae. With this method, it is also easier to distinguish the proportionality of the two species in a given sample. Indeed, 5 samples from Las Perlas, 7 samples from Coïba, 6 samples from Malpelo and 1 sample from Guam islands contain the two species. In all of these samples except two from Malpelo island, the Symbiodiniaceae population is dominated by *Durusdinium-D2*.

The *Porites sp.* samples had been aligned on the CDS from the reference genome of *Cladocopium C15*.

For the *Millepora sp.* samples, transcriptomic reads were aligned with the CDS of the *Symbiodinium A3* reference genome in the first five islands (Coïba, Gambier, Moorea, Cook and Niue islands).

However, for the last two islands, the samples aligned against *Cladocopium thermophilum C3k* reference transcriptome showed very few aligned reads. These were expected results, because I previously observed that I did not have genomic and transcriptomic Symbiodiniaceae references to study this species. I also noticed that *Cladocopium* density is lower than *Symbiodinium* density in *Millepora sp.* coral samples.

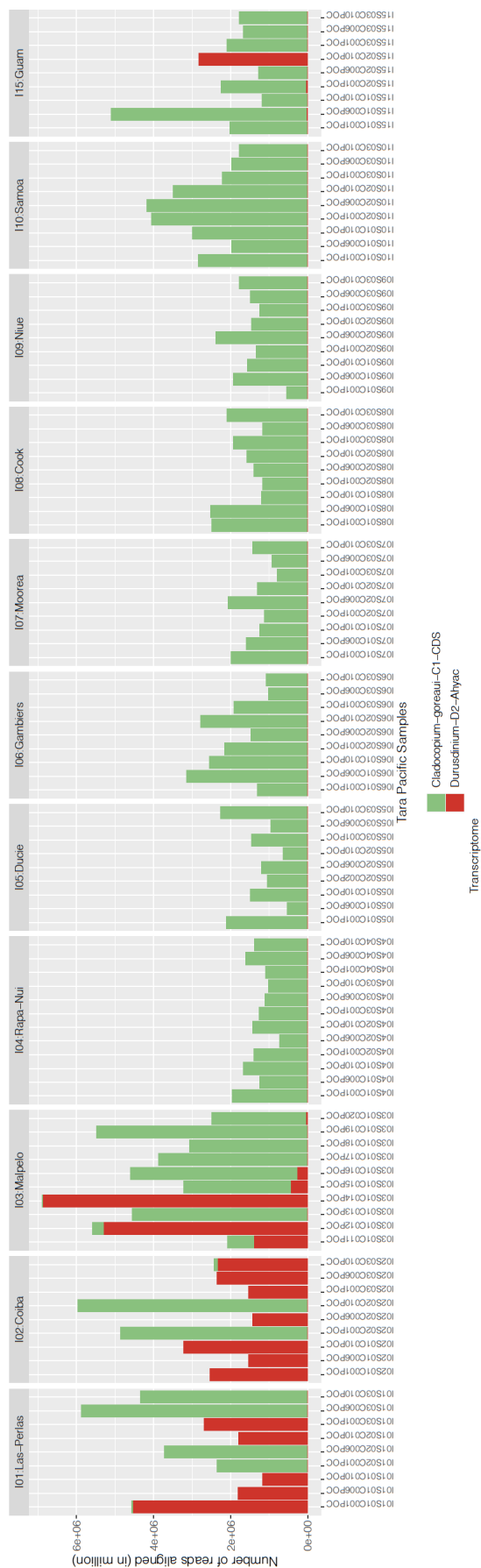


Figure II.12: Barplots of the number of reads aligned on specific Symbiodiniaceae transcriptome and genome references.

Each color corresponds to a specific transcriptome or genome in *Pocillopora sp.* samples.

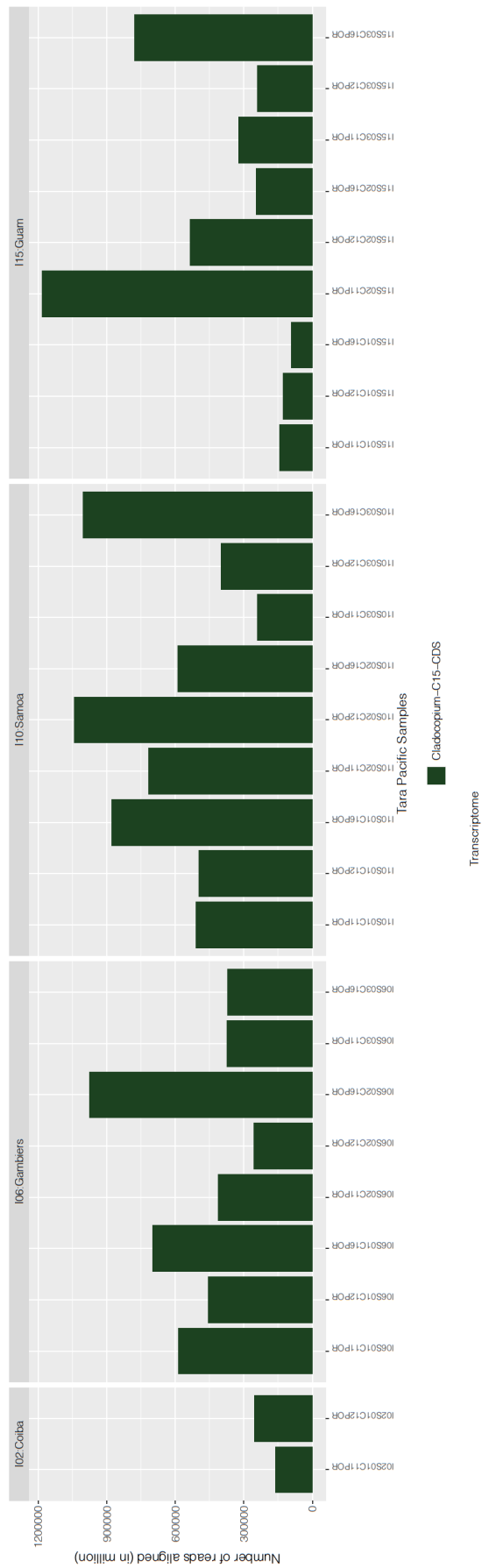


Figure II.13: Barplots of the number of reads aligned on specific Symbiodiniaceae transcriptome and genome references.

Each color corresponds to a specific transcriptome or genome in *Porites sp.* samples.

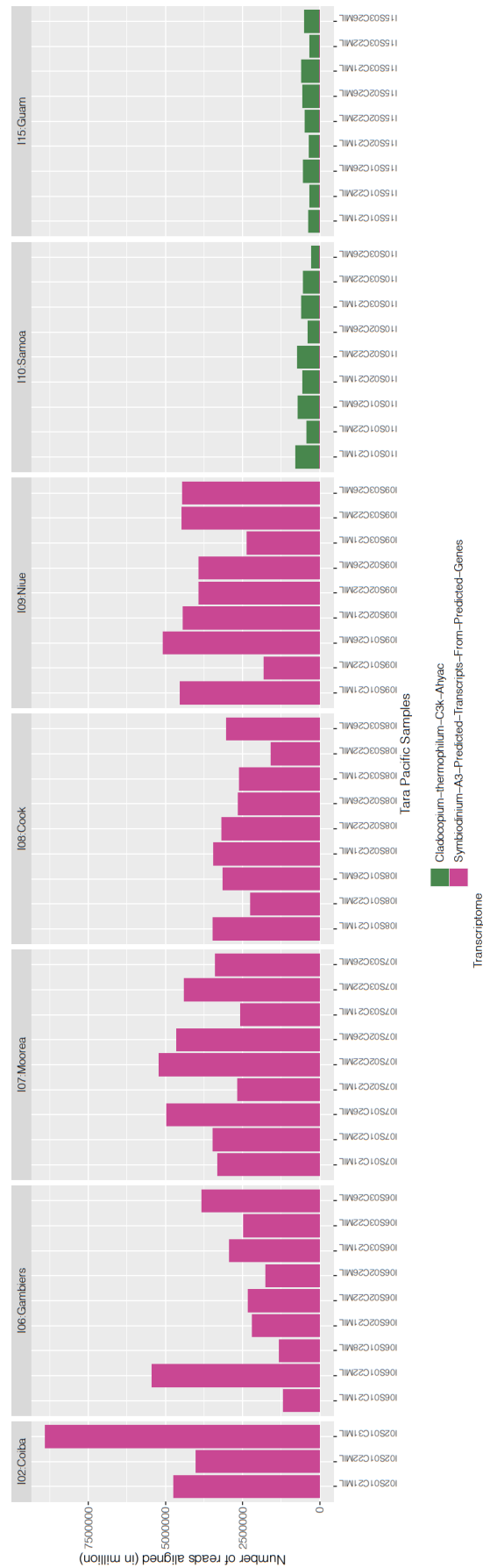


Figure II.14: Barplots of the number of reads aligned on specific Symbiodiniaceae transcriptome and genome references.

Each color corresponds to a specific transcriptome or genome in *Millepora sp.* samples.

The number of reads aligned, the average percentage of identity per transcript or CDS, the average percentage of coverage per transcript or CDS were calculated for each island, by selected references and by coral studied (**APPENDIX VI - Table 4**).

These results confirmed that thanks to these Symbiodiniaceae references, a transcriptomic study can be done for all the samples except the *Millepora sp.* samples from Samoa and Guam islands.

The Symbiodiniaceae algae present in each coral colony were identified using the transcriptomic reads aligned on various available transcriptomes. However, the number of available transcriptomes remains limited and the identity percentages of aligned reads suggested that different populations of Symbiodiniaceae were aligned on the same reference.

To resolve the populations present in each sample, a population structure of *Cladocopium C1* was performed in *Pocillopora sp.* samples using the SNP detected on the *Cladocopium C1* reference genome.

## II.2.3 Population structure of *Cladocopium C1* in *Pocillopora sp.*

### II.2.3.1 Analyses of ITS2 sequences using metabarcoding data

The ITS2<sup>9</sup> sequences are small DNA sequences located between the 5.8S and 28S ribosomal RNA. These small sequences of around 300 base pairs are highly variable among Symbiodiniaceae species and have been proposed as taxonomic markers.

Initially, an analysis of the ITS2 sequences was carried out by Benjamin Hume from the University of Konstanz, in Germany. This analysis used metabarcoding data from *Pocillopora sp.* samples to determine the diversity of Symbiodiniaceae.

For this, the SymPortal database was used to identify ITS2 sequences present in each sample (**Hume et al. 2019**) and was applied on each *Pocillopora sp.* samples (**Figure II.15**).

The analysis of the ITS2 profiles showed different profiles containing several D1 sequences belonging to the *Durusdinium* genus in the islands of the East Pacific (Las Perlas, Coïba, Malpelo and one sample in Guam islands). It also showed several C42 and C1 sequences belonging to the *Cladocopium* genus. These results revealed the presence of two different Symbiodiniaceae genera in the same coral colony.

On Rapa Nui island a single ITS2 profile containing C1cf sequence was detected. For the islands of the Central Pacific (Ducie, Gambier and Moorea islands), more complex ITS2 profiles were obtained. On Ducie and Gambier islands and one sample from Moorea, the C42 ITS2 sequences were obtained. There was one sample of Ducie with a mixture of C1 and C42 sequences suggesting the presence of two species of the same genus in the same colony.

The island of Moorea, in the Central Pacific Ocean also presented complex profiles. Indeed, several ITS2 profiles depending on the sample were observed, like: C1 sequence, C1 and C42 sequences, C42 and D1 sequences.

Then different ITS2 profiles were obtained in the West Pacific (Guam, Samoa, Niue and Cook islands). As in the Central and the East Pacific, there were different ITS2 profiles which combined C42 sequences and C1 sequences with variability between sequences and between profiles.

These results obtained from the metabarcoding data revealed a large diversity of Symbiodiniaceae in *Pocillopora sp.* samples, belonging to the genera *Cladocopium* and *Durusdinium*.

---

<sup>9</sup>Internal Transcribed Spacer 2

They also showed the important diversity in *Cladocopium* ITS2 sequences and the results revealed that *Cladocopium* C1 diversity is more complex than what it was observed in the previous results. Indeed, it appeared that there are different sub-species in *Cladocopium* C1 species, such as *Cladocopium* C42, *Cladocopium* C1 and *Cladocopium* C1cf, which are themselves different depending on the geographic location (East, Central and West Pacific).

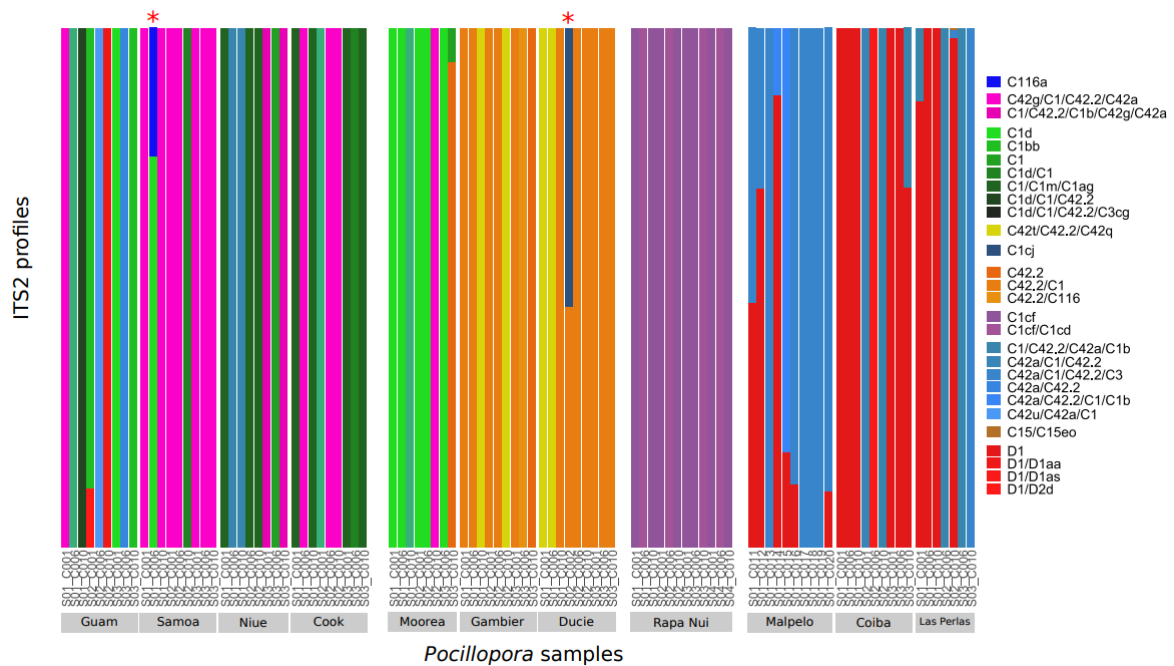


Figure II.15: Relative abundance of Symbiodiniaceae ITS2 profiles in *Pocillopora* sp. samples. The abundance of each ITS2 profile was obtained from ITS2 reads analysed with the SymPortal method. Each color corresponds to a different profile named by its most abundant ITS2 sequence. *Pocillopora* sp. colonies indicated by a star contain 2 different *Cladocopium* ITS2 profiles and have been removed for the population structure analysis.

### II.2.3.2 Revealing sub-variants of the ITS2 sequences in *Cladocopium* C1

To determine the phylogenetic differences between the different *Cladocopium* C1 species, I carried out a study of the ITS2 sequences obtained previously to understand the diversity of the different ITS2 profiles.

First, I looked at the different types of mutations between the sequences obtained previously (**APPENDIX VII - Table 5**). These results showed slight variations between sequences and more precisely slight variations between the C1 and C42 sequences. These variations were greater when I compared the C116, C15 or C3 sequences with the C1 and C42 sequences.

Looking more closely into the ITS2 sequences, I compared the nucleotides to each other, for some sequences (**Figure II.16**). These results showed that certain sequences such as the C1cf sequence is very distant from the other sequences with more than 11 substitutions with the C42.2, C1d, C42.a and C42g sequences. This sequence was found in Rapa Nui island. Comparisons of the ITS2 sequences of C42 and C1d, gave some substitutions between the sequences, with between 2 and 3 substitutions for the sequences C42.2, C1d, C42.a and C42.g.

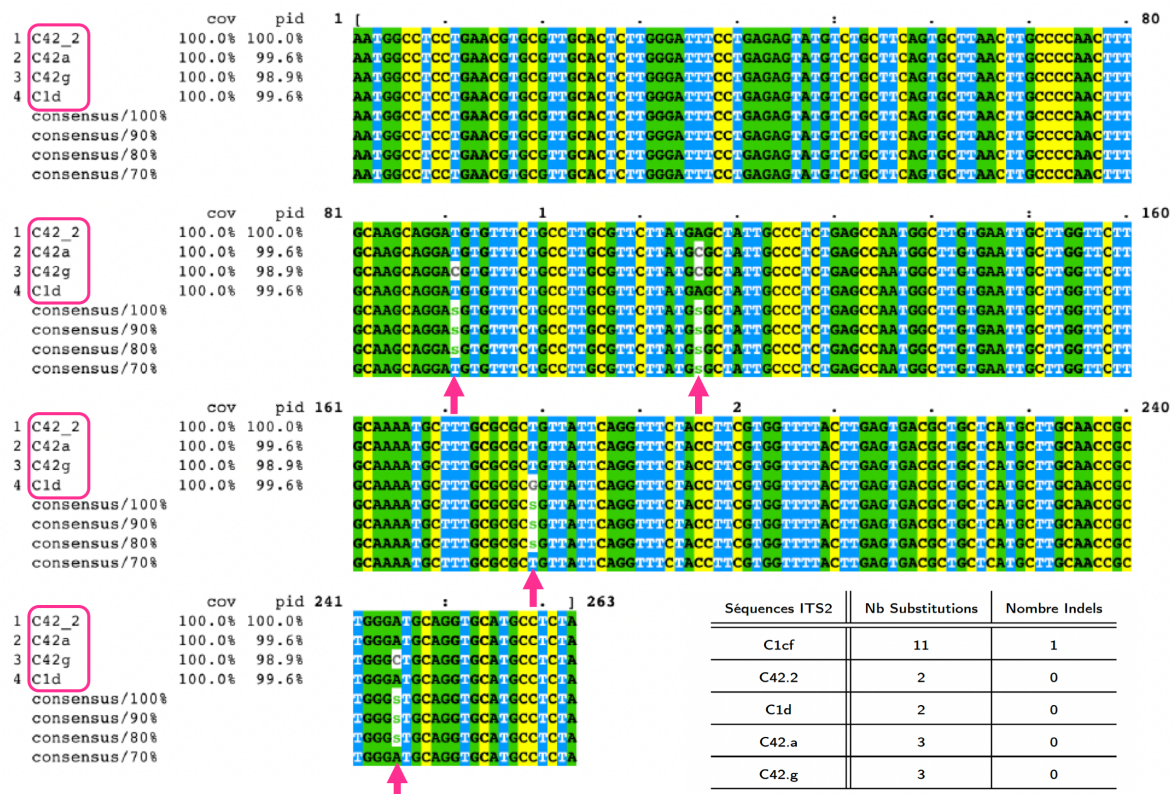


Figure II.16: Comparisons of some ITS2 nucleotide sequences of different *Cladocopium C1* obtained in ITS2 profiles analyses.

These results were obtained using MVIEW<sup>10</sup> tool and showed that the use of ITS2 sequences help to define Symbiodiniaceae populations in the coral samples.

Indeed, there is a sub-species of *Cladocopium C1* for Rapa Nui island, distinct from the other sub-species of *Cladocopium C1* in the other samples. However, when I compared *Cladocopium C1* and *Cladocopium C42* species the distinction appeared to be less precise. Thus, these results allowed to distinguish *Cladocopium C1* from Rapa Nui island from other samples,

<sup>10</sup>Multiple alignment VIEWer developed by EMBL and actualised in 2021



but the distinction between all other samples is still uncertain.

Taken with previous results, the ITS2 sequences are highly variable between samples but also within a coral colony. This diversity could come from:

- The presence of different Symbiodiniaceae species within the same coral colony.
- The polymorphism of these sequences in one population.
- The intragenomic diversity from rRNA sequences, because they are in multiple copies in the micro-algae.

Therefore, I could not use this information to determine how many Symbiodiniaceae lineages are present in *Pocillopora sp.* samples, that is why, we decided to use the entire genome to detect variants between Symbiodiniaceae from coral colonies and then figure out the population structure of *Cladocopium C1* within *Pocillopora sp.*

### II.2.3.3 Genomic vertical coverage to identify the population structure of *Cladocopium C1*

After alignments using dual-genomic samples against *Cladocopium C1* reference genome, the mean of the identity percent by gene was calculated and seemed to be consistent with the alignments with metatranscriptomic samples (**APPENDIX VIII - Figure 8**).

However, looking at the average coverage per contig and comparing these results with the metatranscriptomic samples (**Figure II.17**) it appeared that in the metagenomic samples, the average coverage per contig is less than 2. This means that there is not sufficient coverage of the *Cladocopium C1* genes to allow examination of population structure.

By comparing these results with metatranscriptomic samples, an average transcript coverage greater than 4 was observed. Then, it was decided to not use metagenomic samples to study the population structure. The use of metatranscriptomic samples seemed to be a better way to obtain population structure of *Cladocopium C1*.

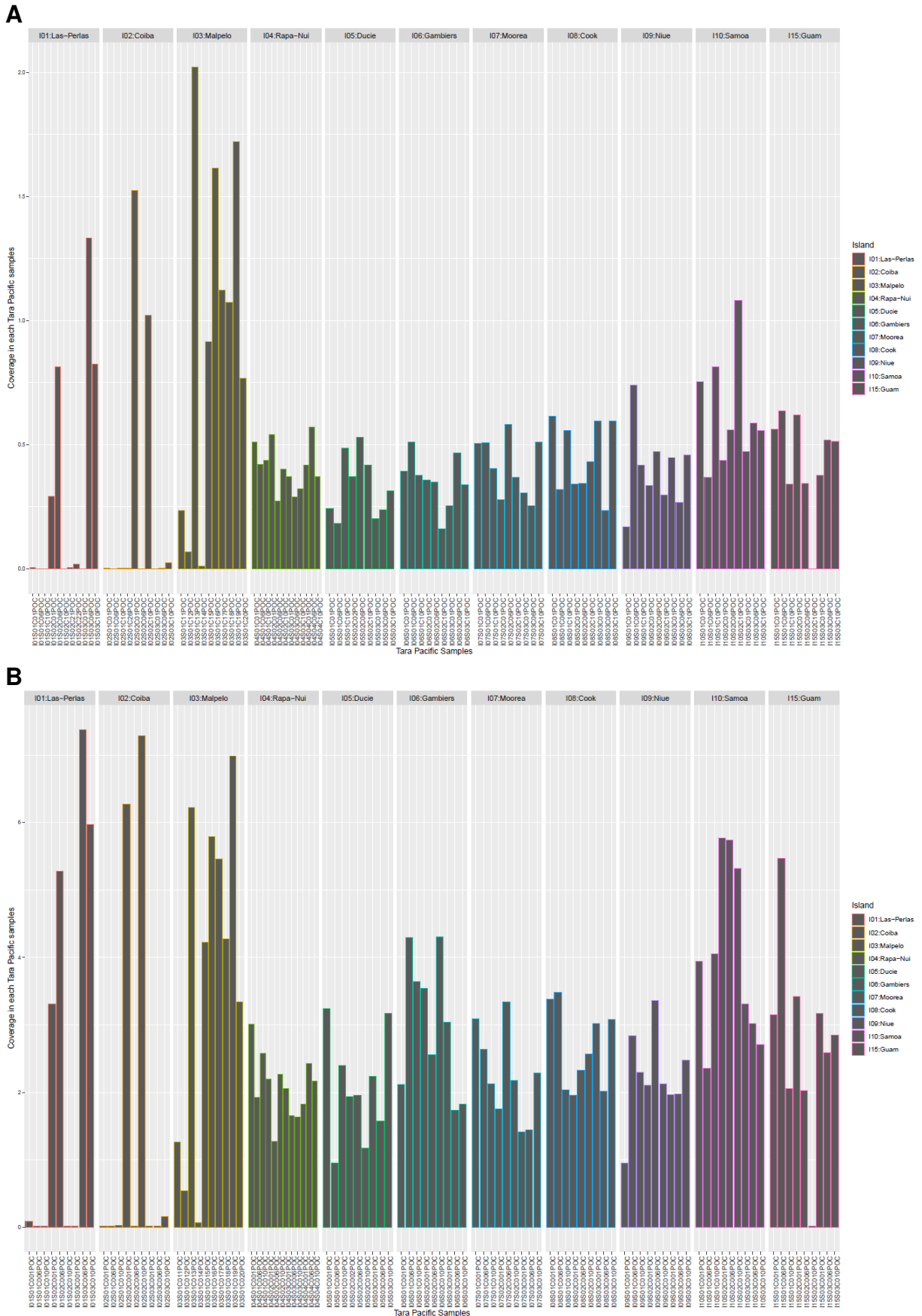


Figure II.17: Barplots of the mean coverage by contig (A) and CDS (B) from *Cladocopium C1* reference genome and CDS aligned on dual-genomic (A) and dual-transcriptomic (B) samples of *Pocillopora* sp.

### II.2.3.4 Population structure of *Cladocopium-C1* using metatranscriptomic samples

SNPs<sup>11</sup> were identified within *Cladocopium* transcripts using the GATK pipeline (**Materials and Methods**) and then a number of genes with a number of SNPs were obtained at each step (**Table II.5**).

Table II.5: Number of remaining SNPs and number of genes covered by these SNPs in the different filtering steps of the GATK pipeline.

Tools	Programs/scripts	Number of genes	Number of SNPs
GATK - T GenotypeGVFs	Perfomed joint genotyping	26.202	302.210
Grep	Removed <i>Durusdinium D2</i> transcripts and samples	23.054	296.798
vcftools -min alleles 2 -max alleles 2	Kept only biallelic variance	23.033	272.290
bcftools view -i % qual ≥ 30	Filtered SNPs with quality score ≥ 30	22.783	259.725
Python script	Kept Biallelic SNPs SNP Coverage ≥ 4 in all samples and frequency calculs	22.521	221.400

From these results I kept 3.712 SNPs for 1.354 genes which allowed to build a dendrogram (**Figure II.18**), after using the Hclust function and the "Complete linkage" option.

The results of this analysis showed the distinction of 5 different lineages of *Cladocopium C1* in the different *Pocillopora sp.* samples.

The lineage 1 brings together all the samples belonging to Rapa Nui island. I previously observed that this island had a different symbiont profile from other islands (**ITS2 analyses**).

Lineages 3 and 5 are a mix of samples belonging to the islands of the East, Central and West Pacific. Indeed, lineage 3 includes samples from the islands of Las Perlas, Coïba, Malpelo, Moorea, Cook, Niue, Samoa and Guam islands. Lineage 5 includes samples from the islands of Ducie, Moorea, Cook, Niue, Samoa and Guam islands. I thus observed a certain complexity within these *Cladocopium C1* lineages, which are present on several islands of the Pacific Ocean.

Finally, lineages 2 and 4 are present in only a few samples, 11 and 5 respectively, as compared to the other lineages. These samples relate to islands in the Central Pacific, which are Ducie, Gambier and Moorea islands. The results indicated that these two lineages have a

<sup>11</sup>Single Nucleotide Polymorphism

more restricted geographic distribution.

These results were compared to the ITS2 profiles (**APPENDIX IX - Table 6**). There was a match between ITS2 profiles, coral host and dual-transcriptomic population structure of *Cladocopium C1* for each sample. These results contributed to the study of the genetic and environmental adaptation of *Cladocopium C1* in the Pacific Ocean and are presented in the article of the **Chapter 3 - Armstrong, Lê-Hoang et al.**

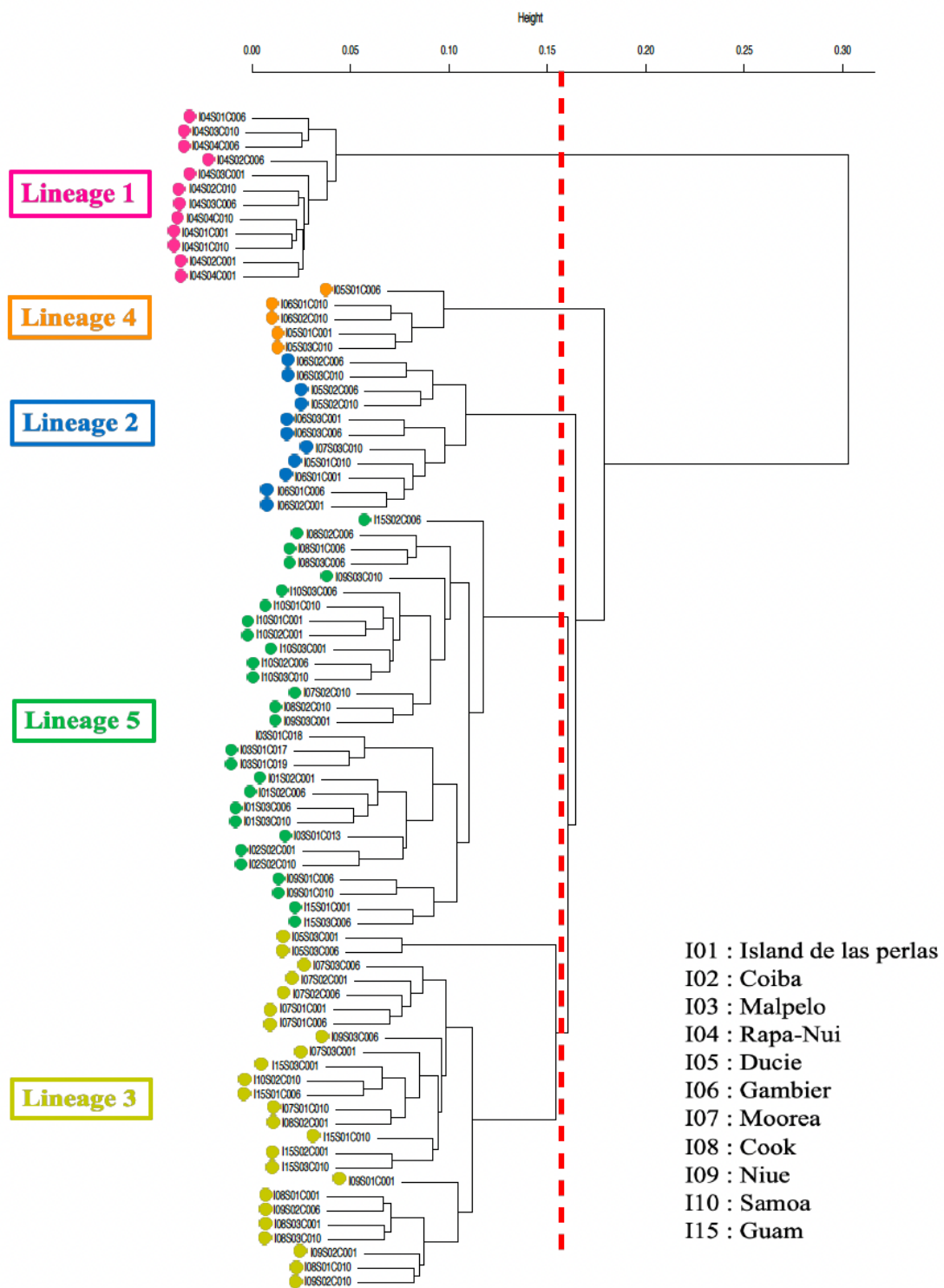


Figure II.18: Hierarchical clustering of *Cladocopium C1* populations in metatranscriptomic samples of *Pocillopora sp.*

Dendrogram representation of the hierarchical clustering of 3.712 SNPs detected by the alignments of metatranscriptomic reads on 1.354 genes of the *Cladocopium C1* reference genome. 82 *Cladocopium C1* samples were clustered into 5 genetic lineages based on a tree height cutoff threshold (red line).

## II.3 Discussion and Conclusions

### II.3.1 A multitude of Symbiodiniaceae genomes and transcriptomes

Symbiodiniaceae reference transcriptomes and genomes were used to construct a Symbiodiniaceae database. The first analyses were focused on the potential contaminants of these different references.

At first sight, GC content results displayed potential contamination of some reference transcriptomes and genomes by the observation of different averages. Because mitochondrial, chloroplastic and nuclear genomes have different GC contents, I could not use the GC content to determine the presence of contaminants in a given reference.

The taxonomic affiliation of each Symbiodiniaceae transcript provided interesting information about the potential contaminants. Moreover, this method appeared sufficient to analyse the Symbiodiniaceae references used, giving a simple view of the potential contamination at different taxonomic levels.

Comparisons between the Uniref90 database with only one proteome from *Symbiodinium microadriaticum* A1 (clade A) and the database created at the Genoscope gave different taxonomic results. In fact, using Uniref90 showed a missed of Symbiodiniaceae references and gave a more important datasets of unknown sequences. Even if the database contains very important datasets, some marine organisms are still not referenced in Uniref90, that is why the choice of the reference to study is essential to do specific analyses.

The construction of a database specialized in marine organisms gave higher assignation to Symbiodiniaceae transcriptomes and genomes than only UniRef90. These results indicated that it is still difficult to find a global database with all marine genome and transcriptome references. Some contaminations were also observed in each transcriptome and especially in two reference transcriptomes (*Cladocopium* C15). These contaminations are probably present because Symbiodiniaceae RNA were extracted and sequenced with the coral hosts then separated by bioinformatic methods.

In biology and also in bioinformatic it is difficult to remove contaminants from datasets as it cannot be entirely removed from the biological samples. Moreover, I worked on dual-transcriptomic samples, which means that I worked on a mixture of RNA from different organisms. Even if the protocols of decontamination are efficient, it is still impossible to avoid all

contaminations. In the case of dual-transcriptomic samples it is still impossible to avoid coral contaminations.

Furthermore, the analyses are based on databases that contain contaminants in the sequences or sequences not assigned to the correct organism. It appeared that these methods are not enough confident to assign some sequences without doubt.

### II.3.2 Diversity of Symbiodiniaceae in the three coral species studied

The objectives of the *Tara Pacific* expedition were to collect samples from three different coral genera (*Millepora sp.*, *Pocillopora sp.* and *Porites sp.*). Then, each sample was extracted and sequenced using the same protocol to be able to compare them.

The alignment of the 188 *Tara Pacific* samples against the reference transcriptomes and genomes gave different results for each coral species that can be separated into three different cases. In the three studied species, I had the opportunity to study the coral *Porites sp.* with a stable symbiosis across the Pacific Ocean, the hydrozoan *Millepora sp.* with two symbionts according to the sampled island and finally the coral *Pocillopora sp.* carrying various Symbiodiniaceae symbionts depending on environmental conditions.

#### II.3.2.1 A limited Symbiodiniaceae diversity in *Porites sp.* colonies

In *Porites sp.*, all colonies were assigned to *Cladocopium C15* in each *Tara Pacific* sample. This result indicates that there is a stable symbiosis between *Porites sp.* and *Cladocopium C15* across the Pacific Ocean (Coïba, Gambier, Samoa and Guam islands).

These results are in accordance with previous results (**LaJeunesse et al. 2004**). However, this broad scale analysis of *Cladocopium C15* may mask a higher genetic diversity not detectable with a read mapping on a limited number of available transcriptomes.

A recent study has shown a correspondence between two cryptic genetic structures of *Porites sp.* and their respective *Cladocopium C15* symbionts suggesting a co-evolution between these two partners (**Forsman, Z.H. et al. 2020**). *Cladocopium C15* analysis at the variant level may resolve these potential sub-populations present in *Porites sp.*

However, the lack of samples for this coral due to the difficulty to sequence their mRNA, did not allow for a global study of these samples. Indeed, for an unknown reason, the polyA

enrichment of mRNA was not efficient resulting in a high proportion of ribosomal RNA in the sequenced reads. New biological protocols are being tested at the Genoscope to resolve this problem and therefore subsequently to give the opportunity to analyse this multitude of *Porites sp.* samples and their symbiont across the Pacific Ocean.

### II.3.2.2 Two Symbiodiniaceae genera in *Millepora sp.* colonies

In *Millepora sp.* colonies, two distinct genera of Symbiodiniaceae were identified depending on the sampled island. *Symbiodinium A3* (clade A) in the East and the Central Pacific Ocean (Gambier, Moorea, Cook and Niue islands) and *Cladocopium* (clade C) in the West Pacific Ocean (Samoa and Guam islands).

It was already known that the *Millepora sp.* coral contains Symbiodiniaceae from clade A and clade C in the Pacific Ocean (**LaJeunesse et al. 2009 and Riddle et al. 2016**). Generally, *Millepora sp.* symbionts seem determined by the geographic location (**Rodríguez L. et al. 2019**), suggesting an important horizontal transmission of the symbiont and a weak host/symbiont specificity. Pour en savoir plus sur la diversité des *Millepora* et leur symbiont, tu devrais lire ça: <https://www.intechopen.com/chapters/69545>

Moreover, *Symbiodinium A3* in the East and the Central Pacific are present in two different *Millepora* species (*Millepora intricata* in Coïba and *Millepora sp.* in the four other islands). These results confirm that *Millepora* symbionts are not specific to the coral host.

To conclude, *Symbiodinium A3* is able to adapt in different corals and especially in two different *Millepora sp.* It also confirms an adaptation of Symbiodiniaceae to the geographical location with *Cladocopium* and *Symbiodinium*, which are two different genera in different islands.

Because, there is no available transcriptomes for this *Cladocopium* species, it was not possible to directly analyse its population structure. A *de novo* assembly of metatranscriptomic reads from these environmental samples is a mandatory preliminary step to study precisely this symbiont.

During the thesis, I started doing this *de novo* assembly transcriptome reference. However, due to the fact that the *Millepora sp.* genome was not ready before the one of *Pocillopora sp.* genome, the analyses on the assembly of *Cladocopium* in these islands were abandoned.



These analyses leave the door open to building a new reference for these still unlisted *Cladocopium* species.

### II.3.2.3 High host/symbiont fidelity in *Pocillopora* sp. colonies.

The host-symbiont relationship in *Pocillopora* sp. samples are discussed in the article **Armstrong, Lê-Hoang et al.**

In *Pocillopora* sp. samples, two Symbiodiniaceae were detected: *Cladocopium goreau* and *Durusdinium D2* assigned respectively to the clades C and D. Moreover, the results showed three possibilities in Symbiodiniaceae diversity which are only clade C, only clade D or both clades in these samples. Once again, these results are in accordance with what it has been reported in the literature (**LaJeunesse et al. 2004 and Riddle et al. 2016**).

From these results I concluded that the Symbiodiniaceae diversity depends on the environmental conditions in *Pocillopora* sp. (**Ladner et al. 2012**). In fact, clade D is more thermal tolerant than clade C. It means that samples which contain both clades are in stressful conditions and are changing their Symbiodiniaceae diversity to adapt in the different environmental conditions. It appeared that samples with only clade D had already changed their Symbiodiniaceae diversity, because they were already exposed to very stressful conditions.

### II.3.3 Classical bioinformatics methods to construct *Cladocopium* C1 population structure

It has been shown that the use of the ITS2 sequences is sufficient to identify the different populations of Symbiodiniaceae (**Hume et al. 2019**).

However, the results of the analyses of the *Cladocopium* C1 ITS2 sequences in the samples of *Pocillopora* sp., revealed a significant complexity of ITS2 profiles. Indeed, the analysis of these ITS2 profiles revealed a multitude of intra-genomic variations of the ITS2 sequences (*C116, C42, C1, C3, C1cf, C15 and D1*).

These results revealed an important complexity of the analysis of the *Cladocopium* C1 populations in *Pocillopora* sp. samples, in contrast to their analyses from transcriptomic references as shown previously.

A complete analysis of these sequences revealed the nucleotide differences of these se-

quences. However, because the nucleotide mutations are variables between sequences related to the same species, it turned out to be complex to determine an ITS2 profile on *Cladocopium C1* populations. Indeed, certain ITS2 sequences for the species C42 for example, have 2 to 3 substitutions between them. These variations may be due to the presence of a different population of *Cladocopium C1* in the sample examined, but also due to the polymorphism of the sequences or even to the intragenomic diversity of the ITS2 sequences.

Therefore, the results of these analyses made it possible to define an ITS2 profile for each of the *Pocillopora sp.* samples. However, the complexity of the ITS2 sequences is too precise, that is why I was interested in the study of a population structure of *Cladocopium C1*.

These analyses showed that the ITS2 is not a good maker sequence to identify *Cladocopium C1* lineages because there is the presence of ITS2 variants between intragenomic copies and also the ITS2 sequences are variables in Symbiodiniaceae within a population given. There is no standard way to cluster ITS2 profiles and then to obtain robust Symbiodiniaceae lineages, that is why longer genomic regions or complete transcriptomes and/or genomes are required to separate Symbiodiniaceae lineages.

The use of a classical method of population structure in *Cladocopium C1* from metagenomic data was not conclusive. Indeed, the coverage of the reads was not sufficient to determine a population structure within these samples (coverage less than 2), because the sequencing depth on dual-genomic samples was insufficient.

To solve this problem and to better understand the diversity of Symbiodiniaceae in *Pocillopora sp.* samples, metatranscriptomic data were therefore necessary.

### **II.3.4 A new definition of Symbiodiniaceae diversity**

The use of different datasets such as genomic data and transcriptomic data allowed a new vision of the Symbiodiniaceae diversity.

Indeed, when they are available the use of transcriptomic references make it possible to identify with precision the different lineages in a dataset. However, there are still missing transcriptomic and genomic references which may hamper the study of the Symbiodiniaceae diversity.

The diversity of ITS2 sequences of Symbiodiniaceae made it possible to go further in the

definition of the Symbiodiniaceae populations. The diversity of these sequences allowed a study at the scale of Symbiodiniaceae populations. These results showed that these sequences are linked on the coral host and the geographical location (**Armstrong, Lê-Hoang *et al.***). They can also be used to precisely define the intra-genomic diversity of Symbiodiniaceae in a sample given.

With the advance of high throughput sequencing, it is now possible to combine a multitude of different kinds of samples combining genomic and transcriptomic data. These two kinds of data are essential for understanding the Symbiodiniaceae diversity.

The study of the *Cladocopium C1* populations using transcriptomic data allowed a new approach in the definition of the Symbiodiniaceae diversity. It made it possible to define lineages and therefore to better understand the extent of this diversity. These analyses represent a new approach for the definition of the Symbiodiniaceae populations.

## Chapter III

# Adaptation of Symbiodiniaceae in symbiosis across the Pacific Ocean

In this chapter I investigated the adaptation and the acclimation capacities of the Symbiodiniaceae from the *Tara* Pacific samples. In the context of mass coral bleaching events, Symbiodiniaceae populations can change between different corals to adapt to different environmental conditions. With these observations, this chapter tries to answer if there is a genetic and/or an environmental adaptation of the symbiotic micro-algae in the Pacific Ocean.

To answer these questions, I used results from the genomic and transcriptomic Symbiodiniaceae analyses, previously presented and then, bioinformatics methods were applied to the symbionts present in the 3 different corals (*Millepora sp.*, *Porites sp.* and *Pocillopora sp.*), in the first 11 islands sequenced from the *Tara* Pacific expedition.

The goal was to determine what are the genes involved in a genetic or an environmental adaptation of the symbiont. To study these questions, differential expression analyses were applied on the different datasets. Then, functional analyses of these genes were done using annotated genes from *Symbiodinium A3* in *Millepora sp.* samples and *Cladocopium C1* in *Pocillopora sp.* On the one hand, specific metabolic pathways involved in Symbiodiniaceae responses to the environment were studied. On the other hand, functional analyses without *a priori* assumptions regarding the genes functions were done on these two datasets. The *Pocillopora sp.* coral and its symbiont *Cladocopium C1* were more specifically studied in the article **Armstrong, Lê-Hoang et al.**

## III.1 Materials and Methods

### III.1.1 Differential expression analyses - Determination of biological replicates

#### III.1.1.1 Expression profiles of the *Tara* transcriptomic datasets

To examine differences between the gene expression profiles according to the environmental parameters at the site of sampling, I first carried out a PCA<sup>1</sup> of each dataset, to visualize data and to determine an expression profile.

Then, I used Pearson correlations on gene expression levels between each pair of coral samples, to determine if samples from each coral colony by site and by island can be used as biological replicates. This correlation coefficient returns:

- **If the coefficient of correlation is close to 1:** samples are correlated → samples involve in the same direction.
- **If the coefficient of correlation is closed to -1:** samples are anti-correlated → samples involve in opposite direction.
- **If the coefficient of correlation is close to 0:** samples are not connected → there is no relation between samples.

These two approaches were used initially to have a global view of the data by crossing samples of the same coral. It is also to give an overview of the samples from a transcriptomic point of view.

#### III.1.1.2 Determination of biological replicates - Environmental conditions

To understand the adaptation of the micro-algae in different environmental conditions, other analyses were applied to the datasets. Indeed, to understand which genes involved in the different environmental conditions, for each coral, two methods were applied to determine biological replicates (**Figure III.1**).

---

<sup>1</sup>Principal Composant Analyses

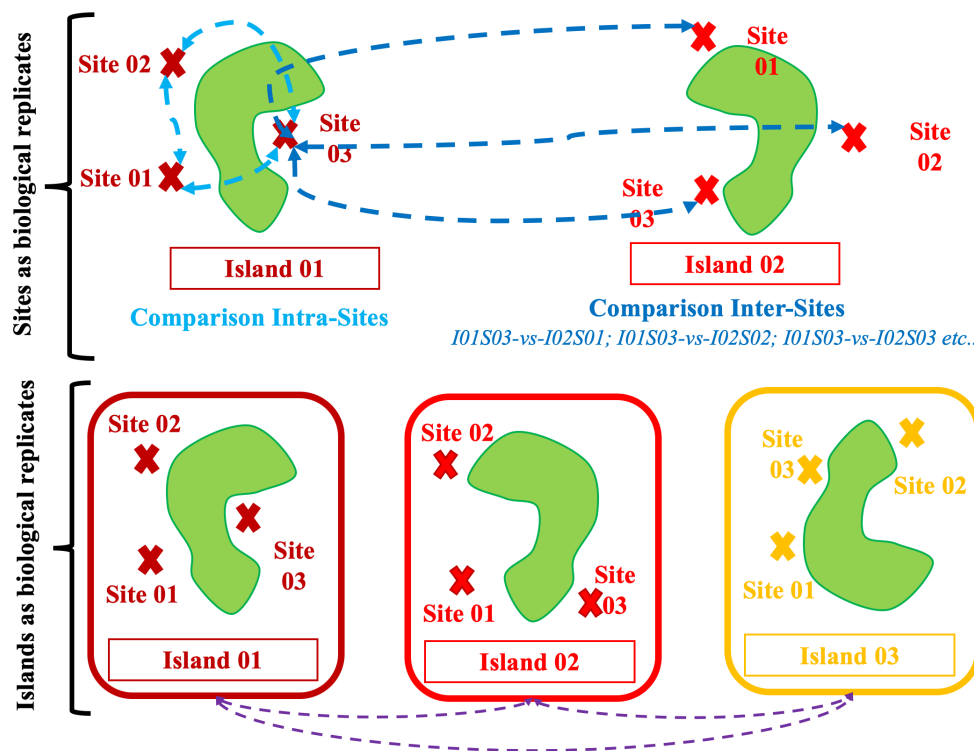


Figure III.1: Schema of the two methods applied on dual-transcriptomic samples from the *Tara* Pacific expedition, to define biological replicates in different environmental conditions.

These methods were used to define:

- **Sites as biological replicates:** Samples from the same site were used as biological replicates. They were compared site by site with those from the island and those from other islands. This analysis was performed for each Symbiodiniaceae expression dataset, independently of each coral.
- **Islands as biological replicates:** Samples from the same island were used as a biological replicates. They were compared island by island and this analysis was performed for each Symbiodiniaceae expression dataset, independently of each coral.

### III.1.1.3 Determination of biological replicates - Genetic conditions

In a second analysis, differential expression analyses were applied only to samples of *Pocillopora* sp. containing the micro-algae *Cladocopium* C1. In these analyses, the aim was to determine using the different populations of *Cladocopium* C1 obtained previously, the genes involved in these different lineages. Secondly, the aim was to obtain a list of genes independent

of the genetic of the symbiont and of the host, which make it possible to explain the adaptation of Symbiodiniaceae in different environmental conditions (**Armstrong, Lê-Hoang *et al.***).

To do this, I applied two different methods to define biological replicates to these datasets:

- **Cladocopium C1 lineages as biological replicates:** each *Cladocopium C1* lineage obtained from the population structure analyses, was used as biological replicates against other lineages, independently of the coral host and island conditions to have genes which vary between lineages.
- **Cladocopium C1, host and island as biological replicates:** is described in the article **Armstrong, Lê-Hoang *et al.***

### III.1.2 bioinformatics methods to perform differential expression analyses

After determining the different conditions to answer to the different questions on the genetic and environmental adaptation of Symbiodiniaceae, differential expression analyses were performed. To do this, the package **DESeq2** developed in R by **Anders S. and Huber W. in 2010** to detect Differentially Expressed Genes (DEG) was used. This tool calculates fold change of gene expression between two conditions taking in account biological replicates. Statistics methods (negative binomial generalized to linear models) were used to calculate a p-value for each differentially expressed gene. An adjust p-value for multiple-comparison was then calculated. Analyses were done according to DESeq2 protocol <sup>2</sup>.

For each analysis, the read counts were normalized using TPM (genes Per Kilobase Millions reads). These metrics were used to normalize read counts, by taking in account sequencing depth and gene length as the classical formula:

1. Divide the read counts by the length of each gene in kilobases.
2. Count same values in a sample and divide this number by 1.000.000.
3. Divide the values obtained in (1) by the values obtained in (2).

Then, a gene was consider as differentially expressed if the adjusted p-value was lower than 0.001. This step was done to have a higher confidence in the list of genes which are differentially expressed.

---

<sup>2</sup><http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html>

The Fold Change is a number which gives an idea of the differential expression between two conditions. For example in a first test between **Gambier-S01 vs Coïba-S03**, if Fold Change of one gene is equal to 3 it means that this gene is three times more expressed in Gambier-S01 than Coïba-S03. If Fold Change of one gene is equal to  $-4$  it means that this genes is four times less expressed in Gambier-S01 than Coïba-S03. This number was used to have an idea of power of differentially expressed genes between two conditions. Then, I developed script in R to select genes which were differentially expressed and shared between tests.

These DEG were then compared using environmental parameters measured during the *Tara Pacific* expedition. To do this, a Canonical Coordinates Analysis (CCA), to represent the results and also to visualize the link between environmental parameters and differential expressed genes was done.

### III.1.3 bioinformatics methods to explain the Symbiodiniaceae expression profiles with associated functions

After getting lists of differentially expressed genes in different conditions, I was interesting in the functions associated with these genes. The aim of this part was to understand the metabolic pathways involved in the genetic and environmental adaptation of Symbiodiniaceae across the Pacific Ocean.

To do this, bioinformatics methods were applied on the different datasets using several databases such as Gene Ontology (GO), Protein families (PFAM) or Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. The annotations of *Symbiodinium A3* and *Cladocopium C1* genomes were used. However, due to the lack of sufficient information for the *Symbiodinium A3* annotation, I performed a functional re-annotation.

During this thesis, other functional analyses were carried out. These analyses followed the results obtained in the article (**Armstrong, Lê-Hoang *et al.***). Among these analyses, the functional analyses of Symbiodiniaceae in the results of Variation Partitioning were done.

#### III.1.3.1 Gene Ontology or GO terms

Gene Ontology database was created in 2000 by the Gene Ontology consortium (**Ashburner M. *et al.* 2000**). The goal of this database was to develop a comprehensive, computational



model of biological systems, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life. Today there are more than 43.850 GO terms; 7.928.834 annotations; 1.568.828 gene products and 5.086 species annotated.

GO terms are defined in three categories which are **cellular component (CC)**, **molecular function (MF)** and **biological process (BP)**.

With this database and with the DEG obtained in the different analyses, several functional analyses using the GO terms of the different categories proposed were done. Using the following bioinformatics methods, enrichments in GO terms were carried out such as:

- **topGO**: topGO is a R package developed and actualized by **Alexa A. et al. in 2021**. This package was used to test GO terms using topology of the GO graph. Some statistic tests were applied on the different methods for eliminating local similarities and dependencies between GO terms.
- **REVIGO**: is a bioinformatics tool to summarize and to visualize long lists of GO terms. This method was developed by **Supek et al. in 2011** and it uses pvalue associated to GO terms to provide a visualisation based on the GO enrichment.
- **dcGOR**: dcGOR is a R package developed by **Fang et al. in 2014**. dcEnrichment function performs enrichment analysis from an input group of domains using a specified ontology. The enrichment analysis uses Fisher's exact test or Hypergeometric test, using the hierarchy of the ontology.

### III.1.3.2 Protein families or PFAM

**PFAM** (Protein families) is a database which contains data protein domains and which classifies these domains using alignment of multiple sequences and Hidden Markov Models (HMM). This database was created in 1997 and actualized often updated. It also gives informations about architecture of proteins domains and their distributions in living species and also their links with other databases. The HMM is an algorithm to identify recurrent sequences by machine learning and automatic recognition of domains.

From the PFAM domains annotated in the genes selected, I searched manually in each PFAM domain some characteristics about metabolic pathways in Symbiodiniaceae organisms

including **nutrient transports and photosynthesis** with *a priori*, and other functions without *a priori*.

I performed scripts which gave Symbiodiniaceae genes with GO terms based on the PFAM associated by genes from the annotation. To do this, corresponding file which contains the links between PFAM domains and GO terms (January 2018 version) was used.

Another bioinformatics analysis on the PFAM of different datasets was done, to have a significant enrichment between annotated genome and the results obtained under differentially expressed conditions. To do this, I used exact Fisher test on the enriched PFAM using a script that I wrote, to allow the data to do in the form of a table (**Table III.1**). Then I applied this Fisher test for each PFAM in each condition.

Table III.1: Table used to perform Fisher test applied on PFAM on the different datasets

	Differentially Expressed Genes	Not Differentially Expressed Genes	Total
PFAM IN datasets	x	m - x	m
PFAM NOT IN datasets	k - x	n - (k - x)	n
Total	k	(m + n - k)	m + n

The "two-side" method was applied in the Fisher tests and p-values were calculated, to allow an indication of the confidence of the results. After, each PFAM calculated with the Fisher exact tests, had a new adjusted p-values calculated using the "p.adjust" function in R and "BH" option from **Benjamini and Hochberg method in 1995**. Then, the PFAM with an adjusted p-value < 0.1 were kept for the analyses.

### III.1.3.3 Kyoto Encyclopedia of Genes and Genomes or KEGG

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a free and regularly updated database created in 2000 by **Minoru Kanehisa and Susumu Goto**. This database brings together systematic analyses of gene functions. Indeed, genes listed in this database have genomic and functional informations and it groups together these genes from sequenced genomes with a functional annotation.

With these informations, the database makes the link between the different metabolic pathways found. It thus makes possible to say with precision the functions involved for a given gene,

with the associated metabolic pathways.

This database also allows metabolic visualization of pathways at different levels and these informations can then be linked to associated chemical compounds, enzymatic molecules and enzymatic reactions.

As part of the functional analyses, the KEGG annotated on the genes of the genomes were used to make the link with the different analyses. To do this, I developed bioinformatics pipelines to automatize the annotations and the functions that can be provided thanks to the KEGG.

Then, to visualize these metabolic pathways I used the the **iPath3 (Interactive Pathways Explorer)** tool for visualizing KEGG terms in several targeted metabolic pathways (**Letunic et al. in 2008**).

#### III.1.3.4 *Symbiodinium A3* and *Cladocopium C1* annotations

The gene annotations of *Symbiodinium A3* from **Shoguchi et al. 2018** was given by the **PHYLO-METAREP** database<sup>3</sup>. To annotate this genome, they used their platform JCVI Phylo-Metagenomics reports which is an online resource to make large volumes of annotated marine genomes for exploratory data analyses<sup>4</sup>. These genes with a functional annotation contains PFAM domains, TIGER domains and GO terms annotated.

However, the annotation was incomplete for some genes, that is why I performed a re-annotation of the complete genome using the **InterProScan** tool developed by **Evgeni M. Zdobnov & Rolf Apweiler in 2001**. To do this, the classical parameters were applied on CDS from reference genome. Thanks to this, additional informations were added on *Symbiodinium A3* genome.

The functional annotation of *Cladocopium C1* given by **Liu et al. 2017** was used to perform functional analyses of this organism. It contains a rich and complete annotation for each gene with PFAM, KEGG and GO associated.

---

<sup>3</sup><http://www.jcvi.org/phylo-metarep/search/index/MMETSP1115>

<sup>4</sup><http://www.jcvi.org/phylo-metarep/>

## III.2 Results

This chapter presents the analyses of differentially expressed genes of the symbiont, carried out on the different datasets available from the three corals *Millepora sp.*, *Porites sp.* and *Pocillopora sp.* from the Tara Pacific expedition.

To do this, the study of the expression level of Symbiodiniaceae genes in dual-transcriptomic samples was the subject of the first analyses using visualization tools such as PCA and by calculating Pearson's correlations between samples.

A second study presents the differential expression analyses obtained on the different datasets, looking at the environment and then the genetic of the symbiont *Cladocopium C1*.

Finally, the functions associated with these differentially expressed genes were presented in *Symbiodinium A3* and in *Cladocopium C1*.

### III.2.1 Expression level of Symbiodiniaceae in dual-transcriptomic samples

#### III.2.1.1 Expression level of Symbiodiniaceae in dual-transcriptomic samples - PCA

First, the different expression profiles according to the sampling islands and sites were determined (**Figures III.2, III.3, III.4**).

The expression profile of *Symbiodinium A3* (**Figure III.2**) showed on PC2 axis a separation of two groups with Coïba, Gambier and some samples of Niue islands in a first group and the islands of Moorea, Aitutaki and Niue in another group. This axis reflects 10% of the explained variance of the datasets. It is interesting to note that despite two different host species in the datasets for Coïba and Gambier islands, the expression profiles seem to cluster together. In PC1 with 15% of the variance explained, the expression profile seems uncertain because I did not observe a specific pattern which came out on this axis.

For the samples of *Cladocopium C1* (**Figure III.3**), the first axis PC1 explains 13% of the variance and separates the samples containing *Cladocopium* from those containing *Durusdinium*. This result was expected because they are two very different genera. Looking at the PC2 axis explaining 6% of the variance, it appeared that the samples from Rapa Nui island are distinct from all the other samples. Once again, this island showed a different expression profile from other islands. This was an expected result, because the population of *Cladocopium C1*

associated with this island is very different from the other islands (**Armstrong, Lê-Hoang et al.**).

Finally, the *Cladocopium C15* samples (**Figure III.4**) showed on the second axis PC2, that 10% of the variance was explained by the environmental conditions. Indeed, there is a separation of three islands Gambier, Samoa and Guam which are grouped by island in this axis. The two samples from Coïba seem similar with those from Guam island. An interesting profile appears to determine the environmental adaptation of *Cladocopium C15* in *Porites sp.* samples.

To conclude on these results, it appeared that the determining factor which can explain the different datasets were the environment materialized by the different islands of the different samples and also the genetic materialized by the different Symbiodiniaceae genera in each coral sample.

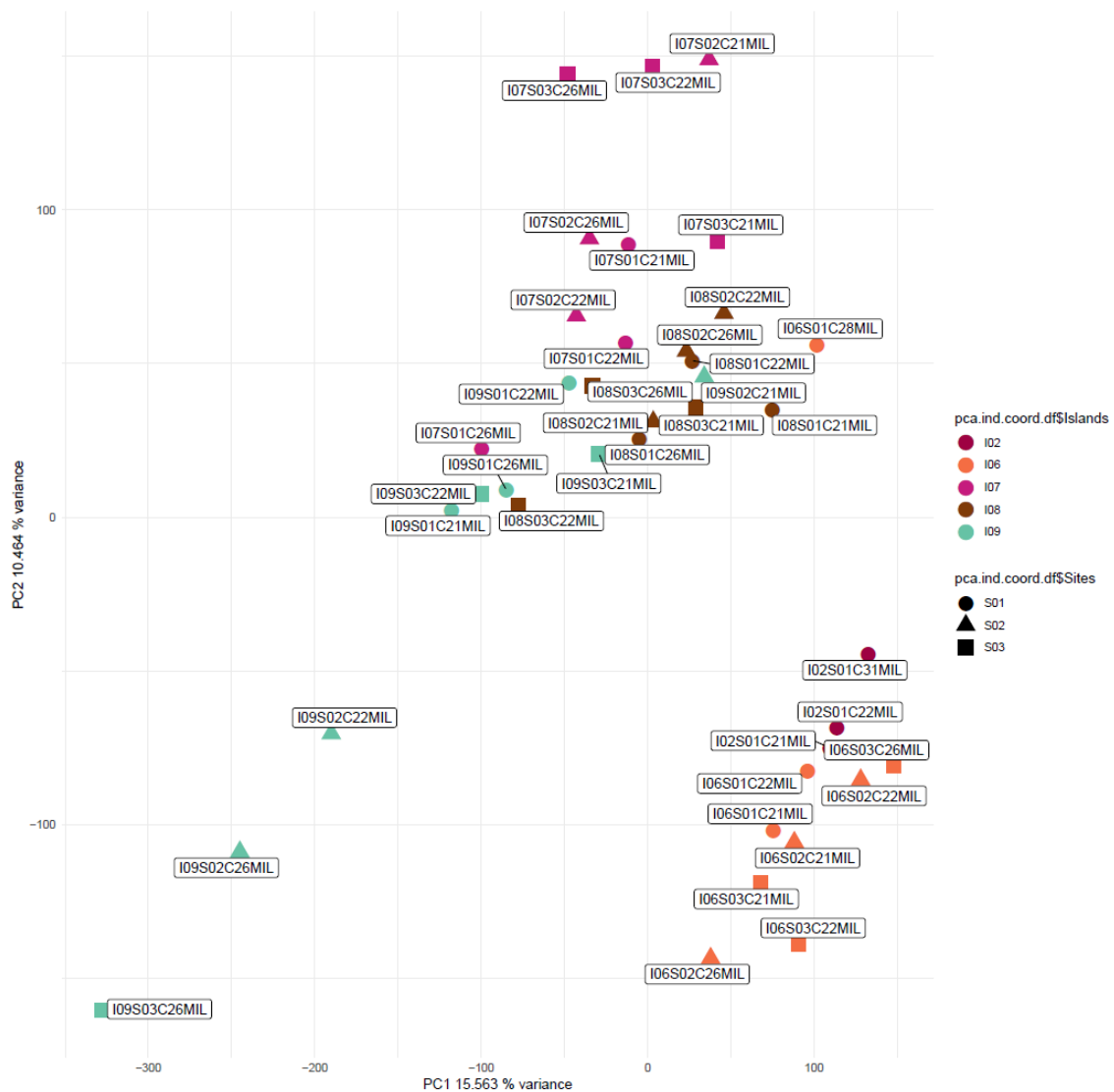


Figure III.2: PCA of gene expression normalized in TPM on each dual-transcriptomic samples. *Symbiodinium A3* expression in *Millepora sp.* Each color indicates an island; Forms indicate sites; I02: Coiba, I06: Gambier, I07: Moorea, I08: Cook and I09: Niue.



Figure III.3: PCA of gene expression normalized in TPM on each dual-transcriptomic samples. *Cladocopium C1* expression in *Pocillopora sp.* Each color indicates an island; Forms indicate *Cladocopium C1* populations; I01: Island de las perlas, I02: Coiba, I03: Malpelo, I04: Rapa Nui, I05: Ducie, I06: Gambier, I07: Moorea, I08: Cook, I09: Niue, I10: Samoa and I15: Guam.

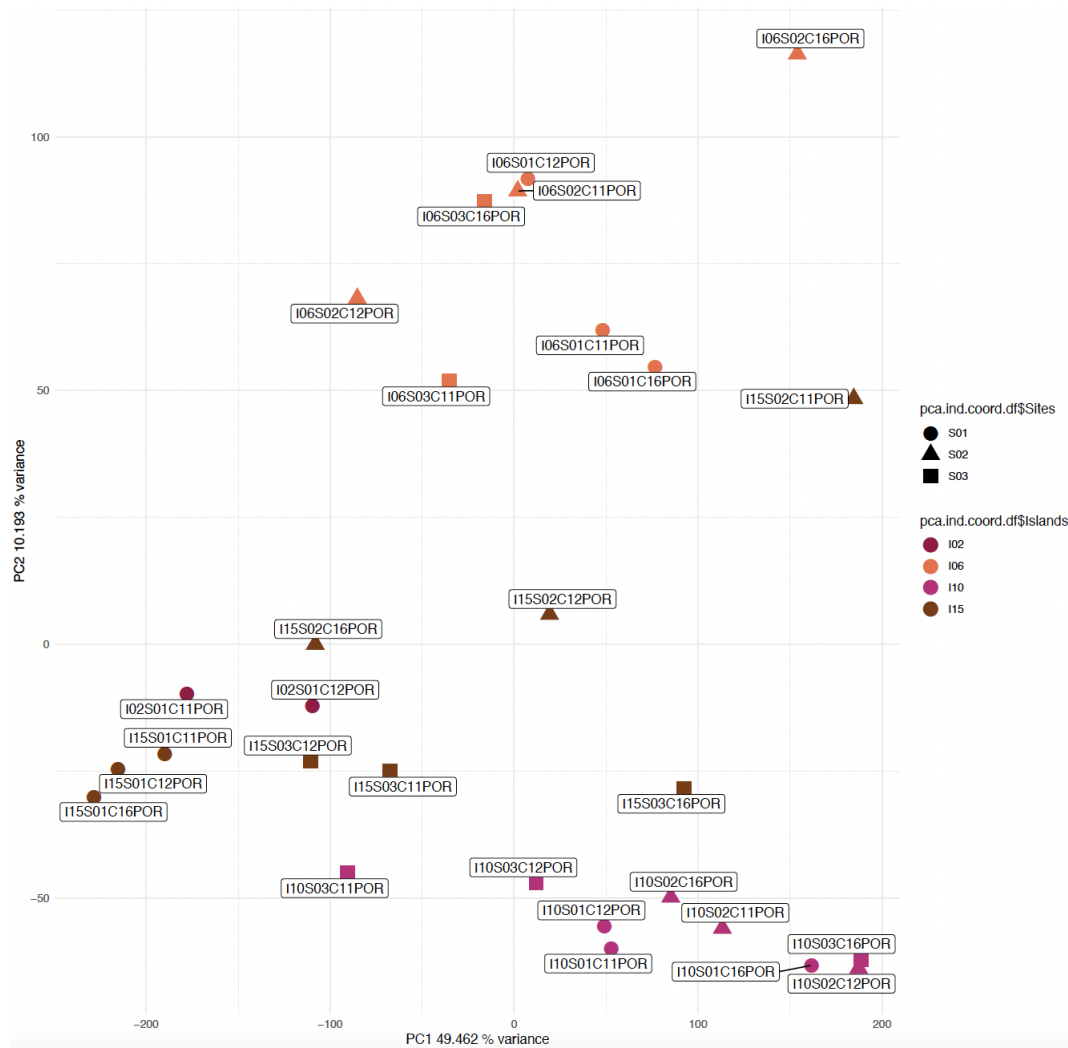


Figure III.4: PCA of gene expression normalized in TPM on each dual-transcriptomic samples. *Cladocopium C15* expression in *Porites sp.*. Each color indicates an island; Forms indicate sites; I02: Coiba, I06: Gambier, I10: Samoa and I15: Guam.



### III.2.1.2 Expression level of Symbiodiniaceae in dual-transcriptomic samples - Pearson correlations

Then the calculation of the Pearson correlations on gene expression levels between each pair of samples were done on the different datasets (**Figures III.5, III.6 and III.7**).

These methods added a supplemental information on the expression profiles of the data. Indeed, I obtained higher correlations between samples from the same island and weaker correlations between samples from different islands.

Furthermore, Pearson coefficients intra-sites and inter-sites showed similar coefficients for the different groups determined above.

From these results, it seemed that samples from the same site should be considered as biological replicates, because their correlation coefficients are very close to 1. Observations of the Pearson correlations between each island showed similar results. I obtained high correlations closed to 1 between samples from the same island.

However, some differences may exist between colonies with two different symbiont genera in *Pocillopora sp.* These results were expected due to the fact that these genera are very different.

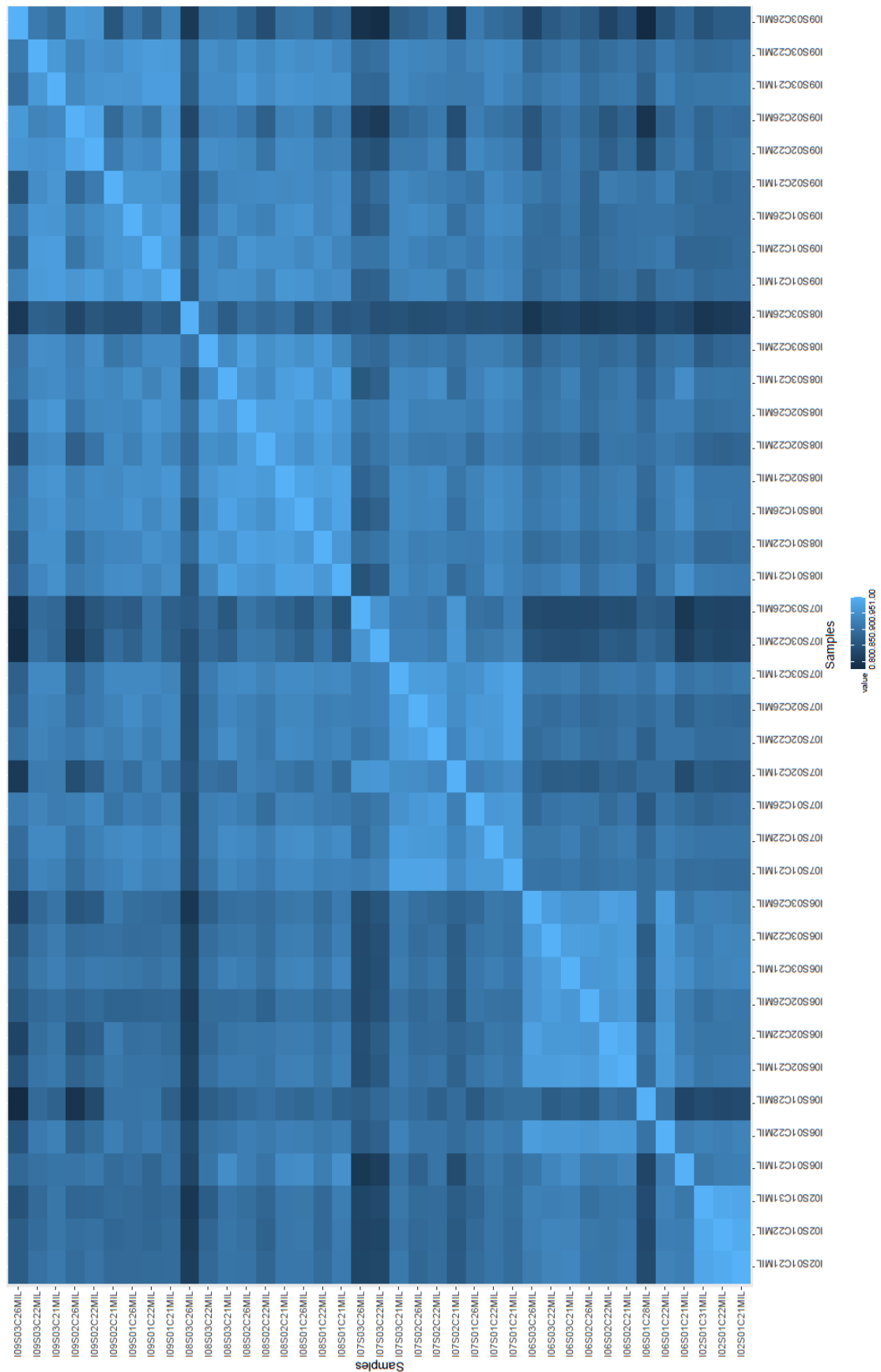


Figure III.5: Pearson correlations on gene expression levels normalized in TPM on *Symbiodinium A3* in *Millepora sp.*

102: Coiba, 106: Gambier, 107: Moorea, 108: Cook and 109: Niue.

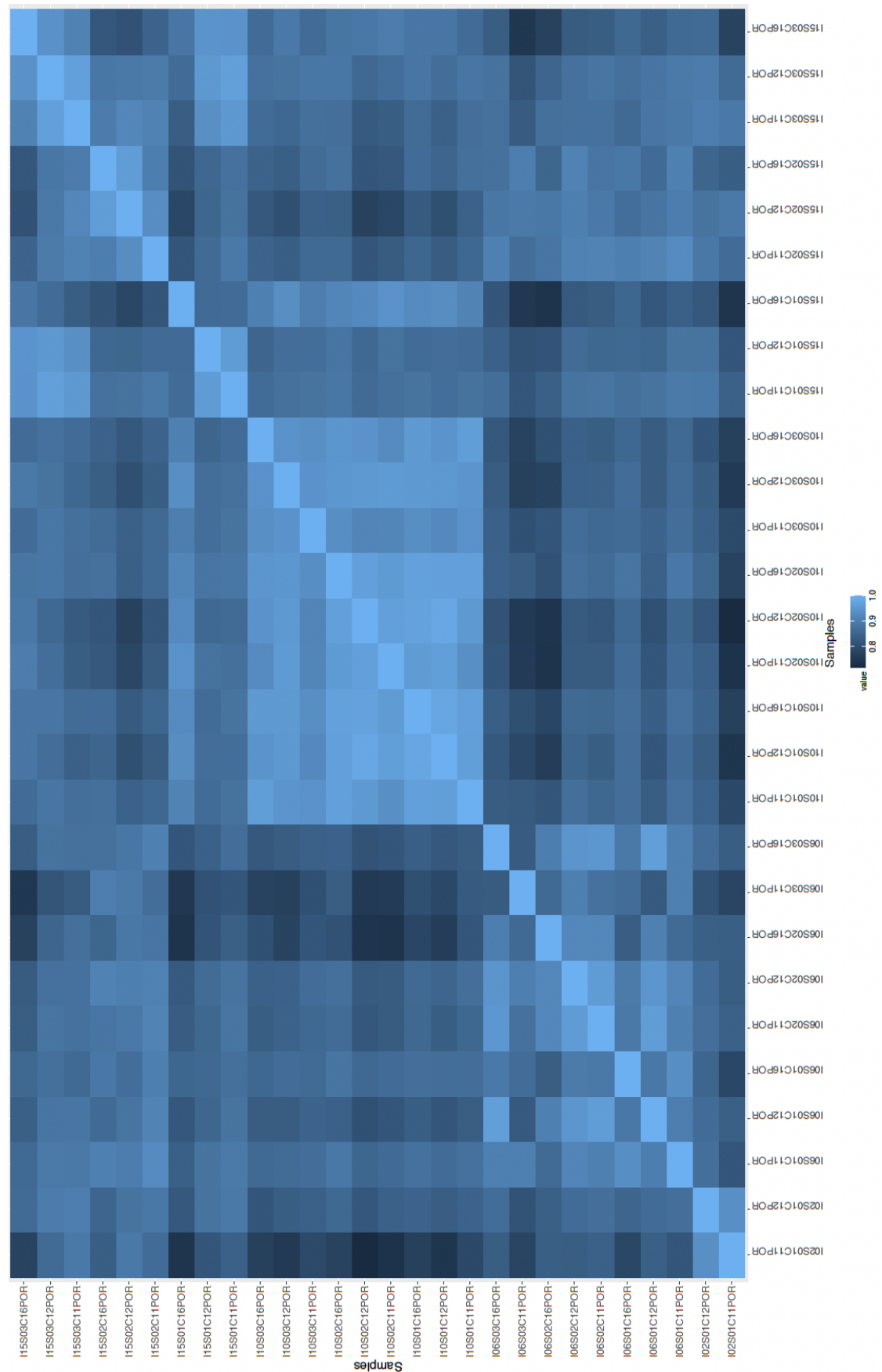


Figure III.6: Pearson correlations on gene expression levels normalized in TPM on *Cladocopium C15* in *Porites sp.*

I02: Coiba, I06: Gambier, I10: Samoa and I15: Guam.

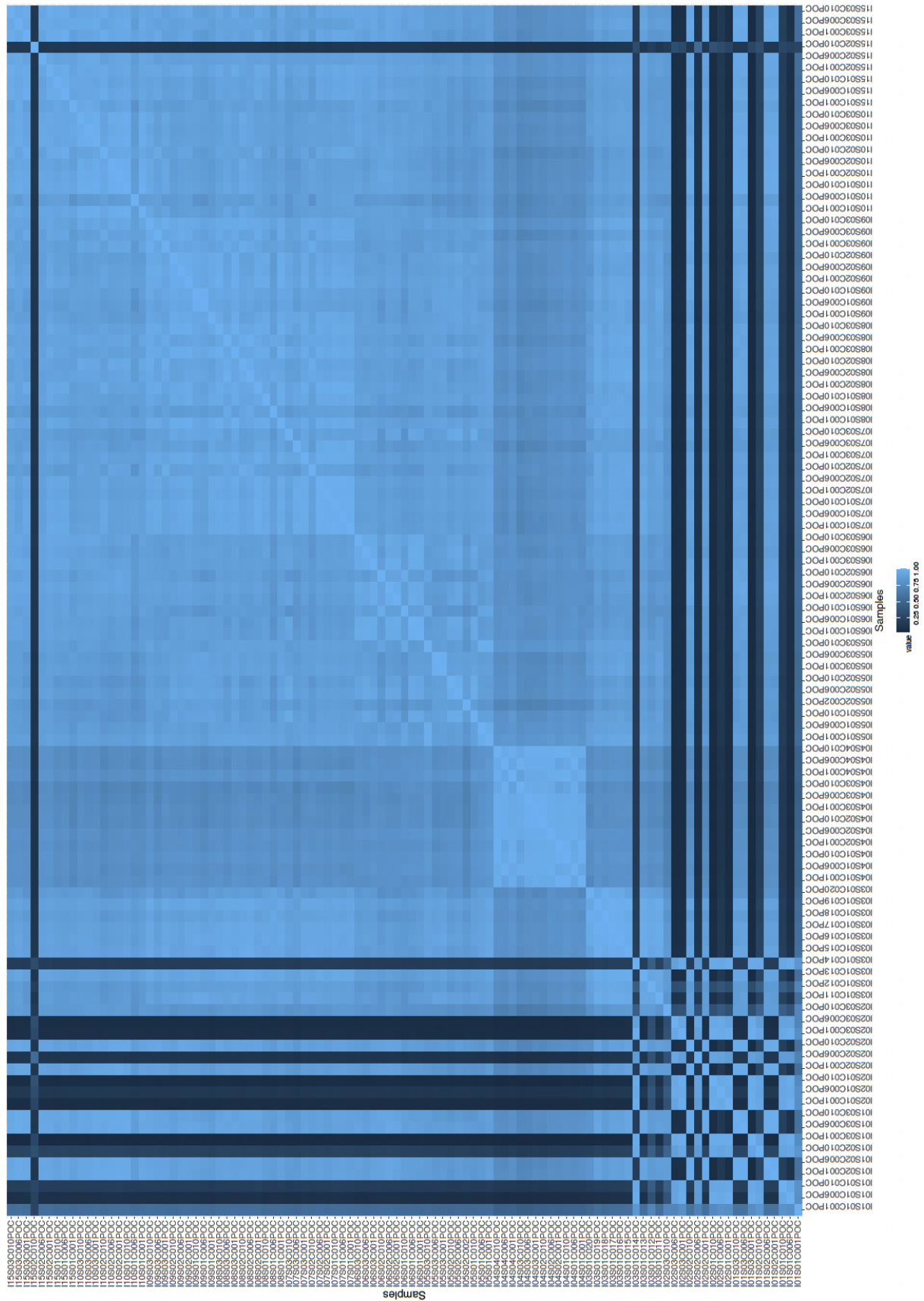


Figure III.7: Pearson correlations on gene expression levels normalized in TPM on *Cladocopium C1* in *Pocillopora sp.*

I01: Island de las perlas, I02: Coiba, I03: Malpelo, I04: Rapa Nui, I05: Ducie, I06: Gambier, I07: Moorea, I08: Cook, I09: Niue, I10: Samoa and I15: Guam.

## III.2.2 Environmental adaptation of Symbiodiniaceae across Pacific Ocean

### III.2.2.1 Environmental adaptation of *Symbiodinium A3* in the Pacific Ocean

*Symbiodinium A3* is presented in *Millepora sp.* samples, in Coïba, Gambier, Moorea, Aitutaki and Niue islands.

The A3 reference genome contains a total of 87.464 predicted genes from predicted transcripts. A total of 64.962 of them were aligned with more than 98% identity on the dual-transcriptomic *Millepora sp.* samples, to detect differentially expressed genes between the different islands.

I obtained the number of differentially expressed genes using sites as biological replicates (A) and islands as biological replicates (B) (**Figure III.8**).

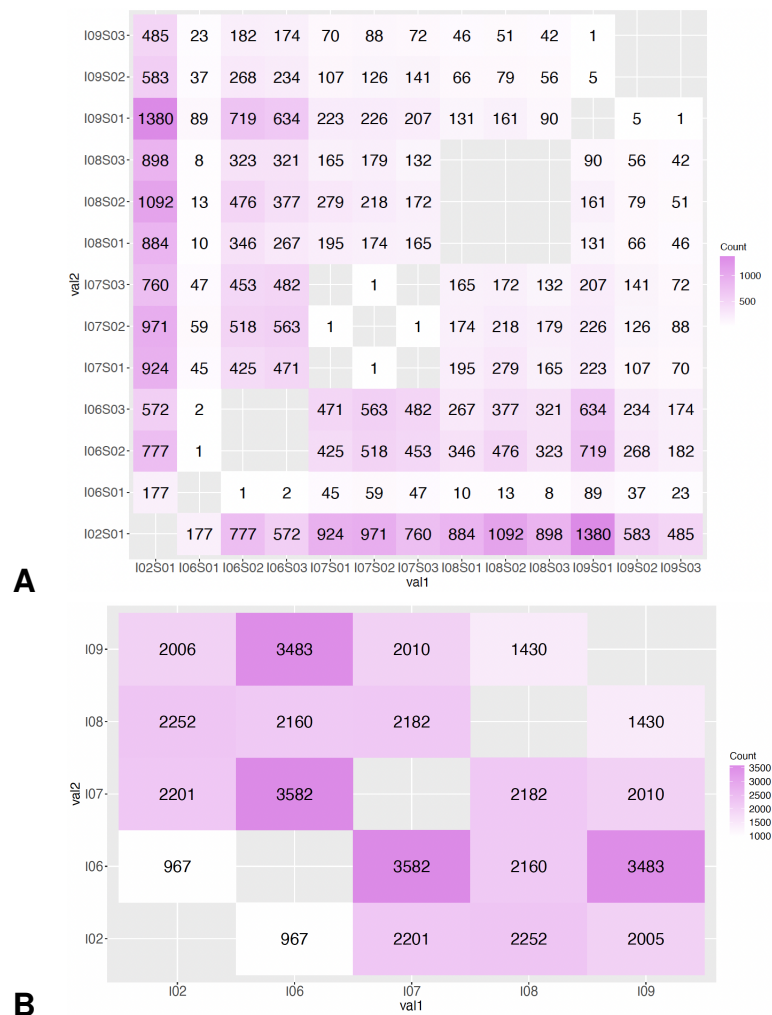


Figure III.8: Number of differentially expressed genes in *Symbiodinium A3*, from *Millepora sp.* samples. (A) DEG with sites as biological replicates (B) DEG with islands as biological replicates. I02: Coïba, I06: Gambier, I07: Moorea, I08: Cook and I09: Niue.

The number of differentially expressed genes using the sites as biological replicates (**Figure III.8.A**), showed a significant number of DEG between samples from Coïba and those from other islands. This difference can be explained by the fact that the coral host is different from the other samples. This result suggested that *Symbiodinium A3* in *Millepora intricata* are either genetically different from *Symbiodinium A3* in *Millepora platyphylla* and carry a specific expression pattern to be able to live in symbiosis with these different species.

By looking the number of DEG within the same island, it appeared a very low number of DEG between the sites of the same island. These results were expected because the sites on the same island have very similar environmental conditions. These results were also observed in the first analyses of *Symbiodinium A3* expression levels.

The DEG of *Symbiodinium A3* between the islands (**Figure III.8.B**) showed the islands of Aitutaki and Niue close in terms of expression profiles, because they have a low number of DEG compared to other comparisons (1.430 DEG). The number of DEG is lower between the islands of Coïba and Gambier compared to the other islands (967 DEG). *Symbiodinium A3* on these two islands also seem to be similar. These results of the differential expression analyses confirmed the expression profiles observed previously with the Pearson correlations and with the PCA analyses. It also appeared in these results that the use of sites as biological replicates in the same island were not conclusive. There are very few DEG between the sites within the same island.

### III.2.2.2 Environmental adaptation of *Cladocopium C15* in the Pacific Ocean

One clade of Symbiodiniaceae was detected in *Porites sp.* samples. Metatranscriptomic reads were aligned on *Cladocopium C15*. The C15 genome contains a total of 22.508 CDS and dual-transcriptomic reads were aligned with a total of 20.274 CDS with more than 98% identity.

I obtained the number of differentially expressed genes using sites as biological replicates (A) and islands as biological replicates (B) (**Figure III.9**).

The number of differentially expressed genes using the sites as biological replicates (**Figure III.9.A**) gave a low number of DEG between Gambier and Coïba islands. They seemed different of Samoa and Guam islands. The other profiles have a certain homogeneity in terms of DEG.

The number of DEG within the same island gave a very low number of DEG. These results were expected because the sites on the same island have very similar environmental conditions.

There is a high number of DEG between Gambier and Samoa (1.796 DEG). In the other comparisons the number of DEG is low (less than 1.000 DEG). This means that the environmental conditions have a low influence of the gene expression of *Cladocopium C15* in *Porites sp.* samples in these four different islands.

It also appeared that the use of sites as biological replicates were not conclusive, because there is very few DEG between the sites within the same island.

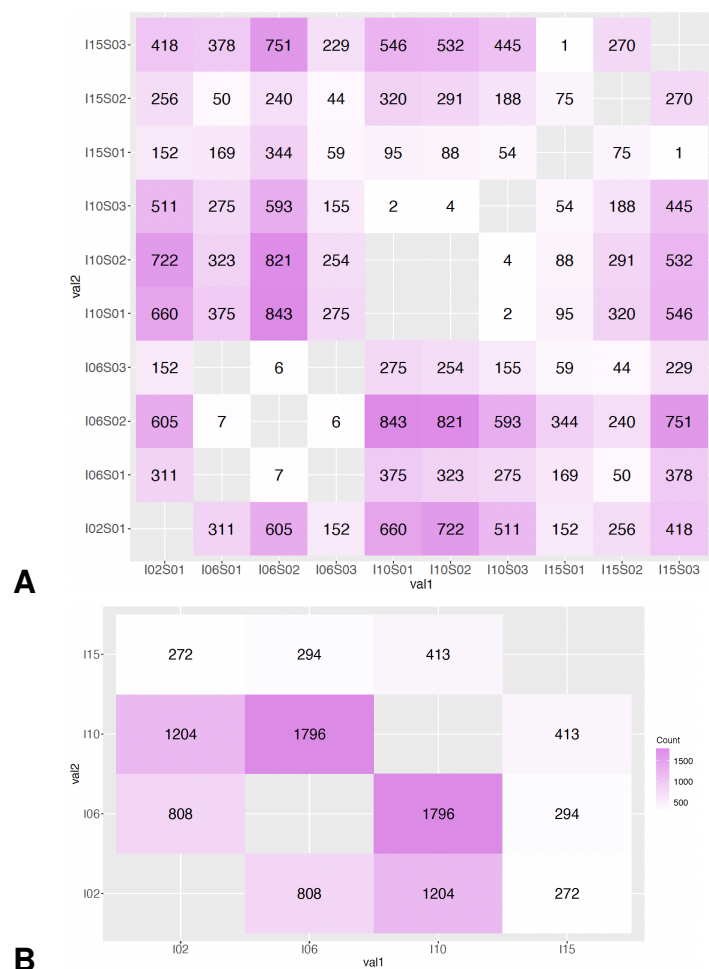


Figure III.9: Number of differentially expressed genes in *Cladocopium C15*, from *Porites sp.* samples. (A) DEG with sites as biological replicates (B) DEG islands as biological replicates. 102: Coiba, 106: Gambier, 110: Samoa and 115: Guam.

### III.2.2.3 Environmental adaptation of *Cladocopium C1* in the Pacific Ocean

Two different clades of Symbiodiniaceae were detected in *Pocillopora sp.* samples. Meta-transcriptomic reads were aligned on *Cladocopium C1* and/or *Durusdinium D2*. *Durusdinium D2* was found in the first islands (Las Perlas, Coïba and Malpelo islands) and one sample from Guam.

The C1 genome CDS contains a total of 35.913 CDS and a total of 24.305 CDS were aligned at 98% identity. From these results, the goal was to detect differentially expressed genes of *Cladocopium C1* between the different islands, after removing samples with *Durusdinium D2*.

The number of differentially expressed genes using the sites as biological replicates (**Figure III.10.A**), appeared that Rapa Nui island once again has a different expression profile from the other islands. Some islands seemed close like the islands of Aitutaki and Niue. However, with these large numbers of samples and therefore of comparisons, it was more difficult to distinguish using the sites as biological replicates, the profiles between them. Again, looking at the number of DEG within the same island it appeared a very low number of DEG between the sites of the same island.

There is a very high number of DEG between the island of Rapa Nui and the other islands (**Figure III.10.B**). These results again, confirmed a different expression profile for the samples belonging to this island. Once again, I observed a very strong proximity between Aitutaki and Niue islands, with only 6 DEG between these two islands.





### III.2.2.4 Environmental adaptation of Symbiodiniaceae across the Pacific Ocean

The link between differentially expressed genes and environmental parameters were estimated using a CCA<sup>5</sup> visualization (**Figures III.11, III.12 and III.13**).

DEG of *Symbiodinium A3* in *Millepora sp.* samples and their links with the available environmental parameters (Gambier, Aitutaki and Niue islands) (**Figure III.11**) were done using a CCA. There is a separation of Gambier from the other two islands, and this separation could be related to the environmental parameters of chlorophyll a, Particulate Organic Carbon (POC), O<sub>2</sub> and salinity concentrations.

Then, DEG of *Cladocopium C1* in *Pocillopora sp.* samples, and their links with the available environmental parameters (Rapa Nui, Ducie, Gambier, Aitutaki, Niue and Samoa islands) were done (**Figure III.12**). It was observed that the island of Rapa Nui is once again different from the other islands. There is also a separation between islands such as Gambier and Samoa islands in a first group and Aitutaki, Niue and Ducie islands in another group. The distinction between these two groups would be related to the chlorophyll a, POC and luminosity (PAR).

Finally, DEG of *Cladocopium C15* in *Porites sp.* samples and their links with the available environmental parameters (Gambier and Samoa islands) were done (**Figure III.13**). It was difficult with only these two islands to determine an analysis that can explain the expression profiles of *Cladocopium C15* genes in these islands. The observations showed environmental parameters which potentially separate these two islands. It would be the salinity and the O<sub>2</sub> concentrations for the island of Gambier. This island seems more impacted by these parameters than the other island. However, due to the little data that I had, it was difficult to conclude on this last analysis.

---

<sup>5</sup>Canonical Correlation Analysis

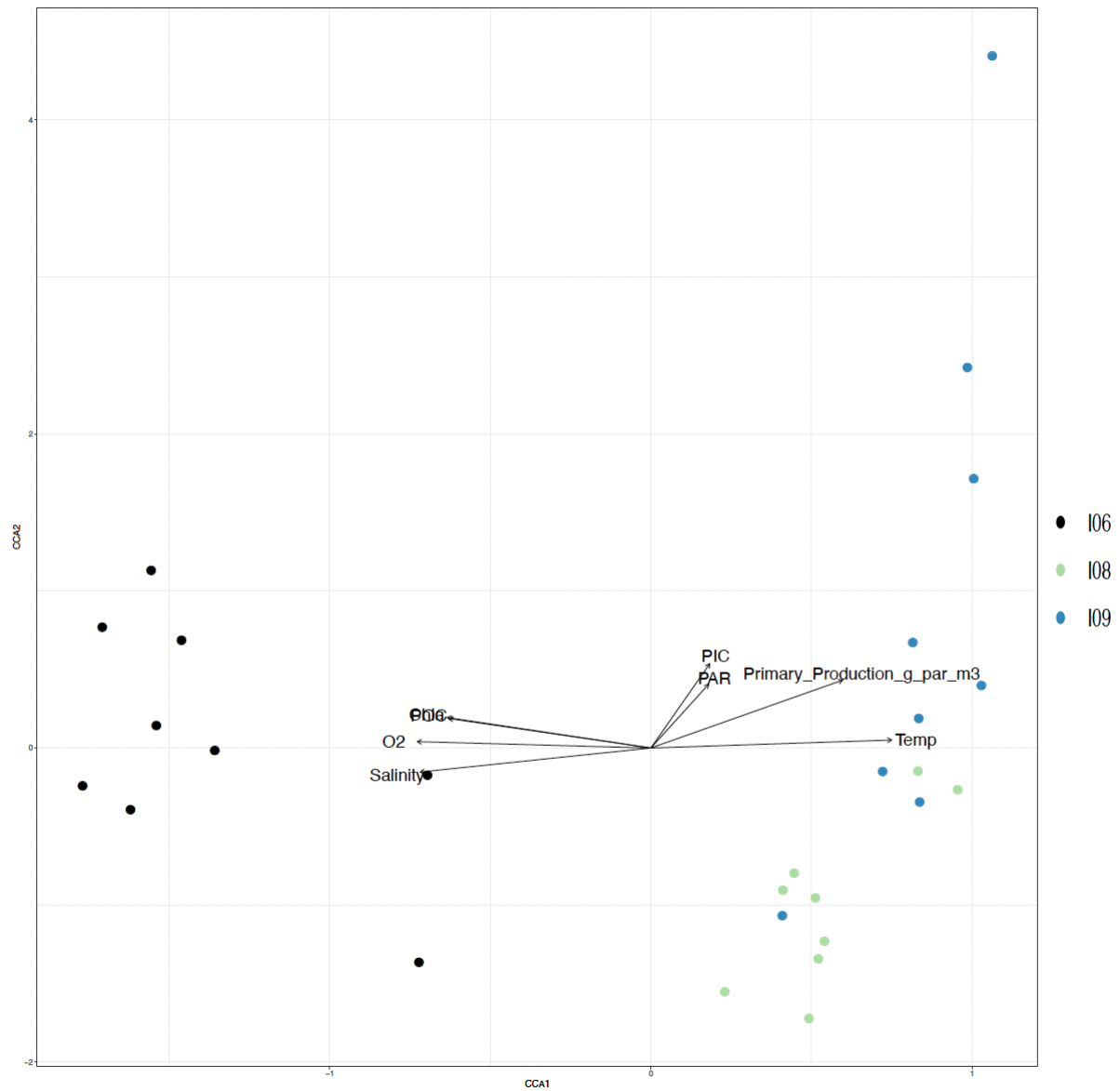


Figure III.11: CCA of gene expression from DEG normalized in TPM. *Symbiodinium A3* DEG in *Millepora sp.* samples; Each color indicates an island; Arrows indicate environmental parameters available and chosen for the CCA. I06: Gambier, I08: Cook and I09: Niue.

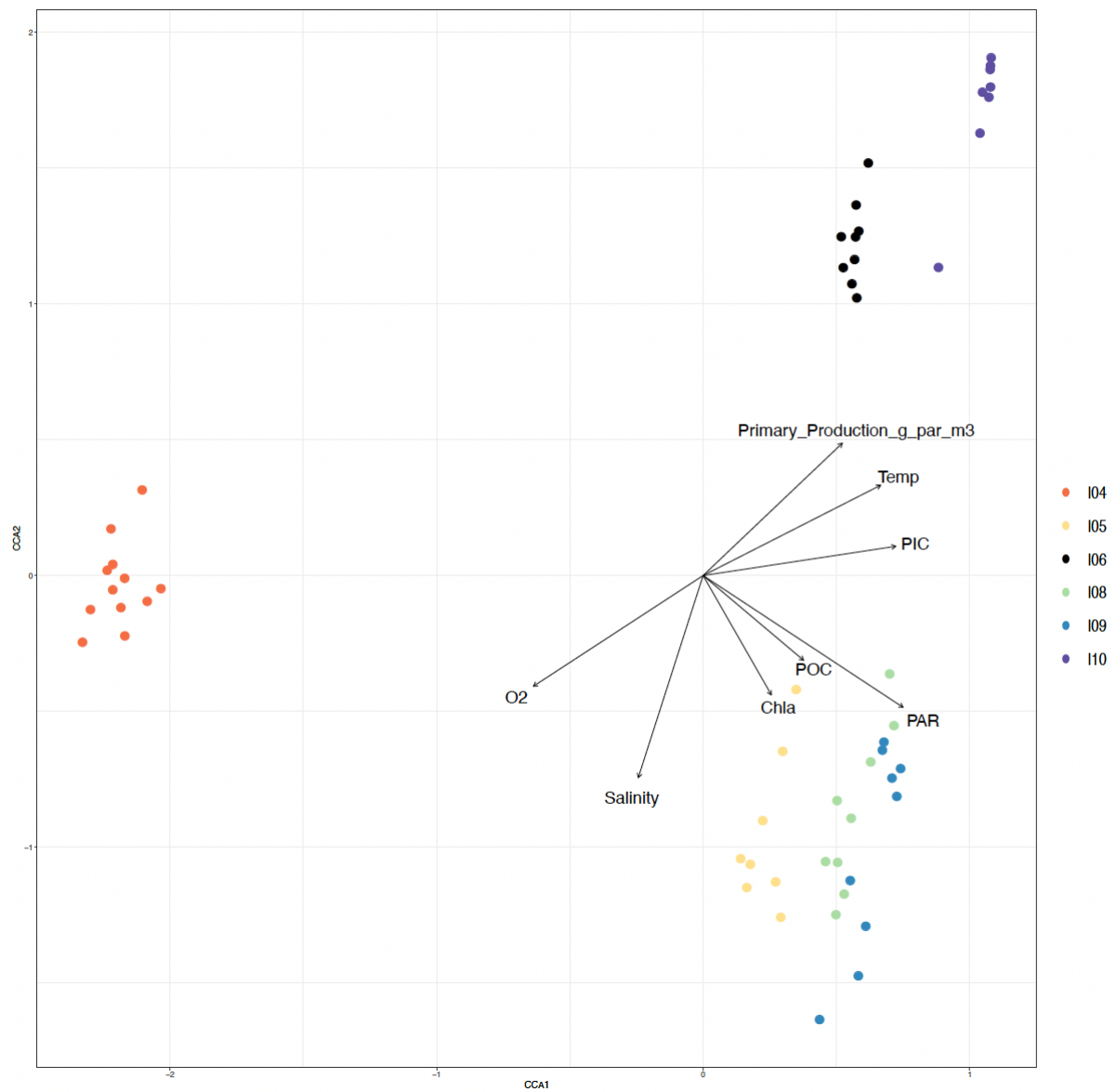


Figure III.12: CCA of gene expression from DEG normalized in TPM. *Cladocopium C1* DEG in *Pocillopora sp.* samples; Each color indicates an island; Arrows indicate environmental parameters available and chosen for the CCA. I04: Rapa Nui, I05: Ducie, I06: Gambier, I08: Cook, I09: Niue and I10: Samoa.

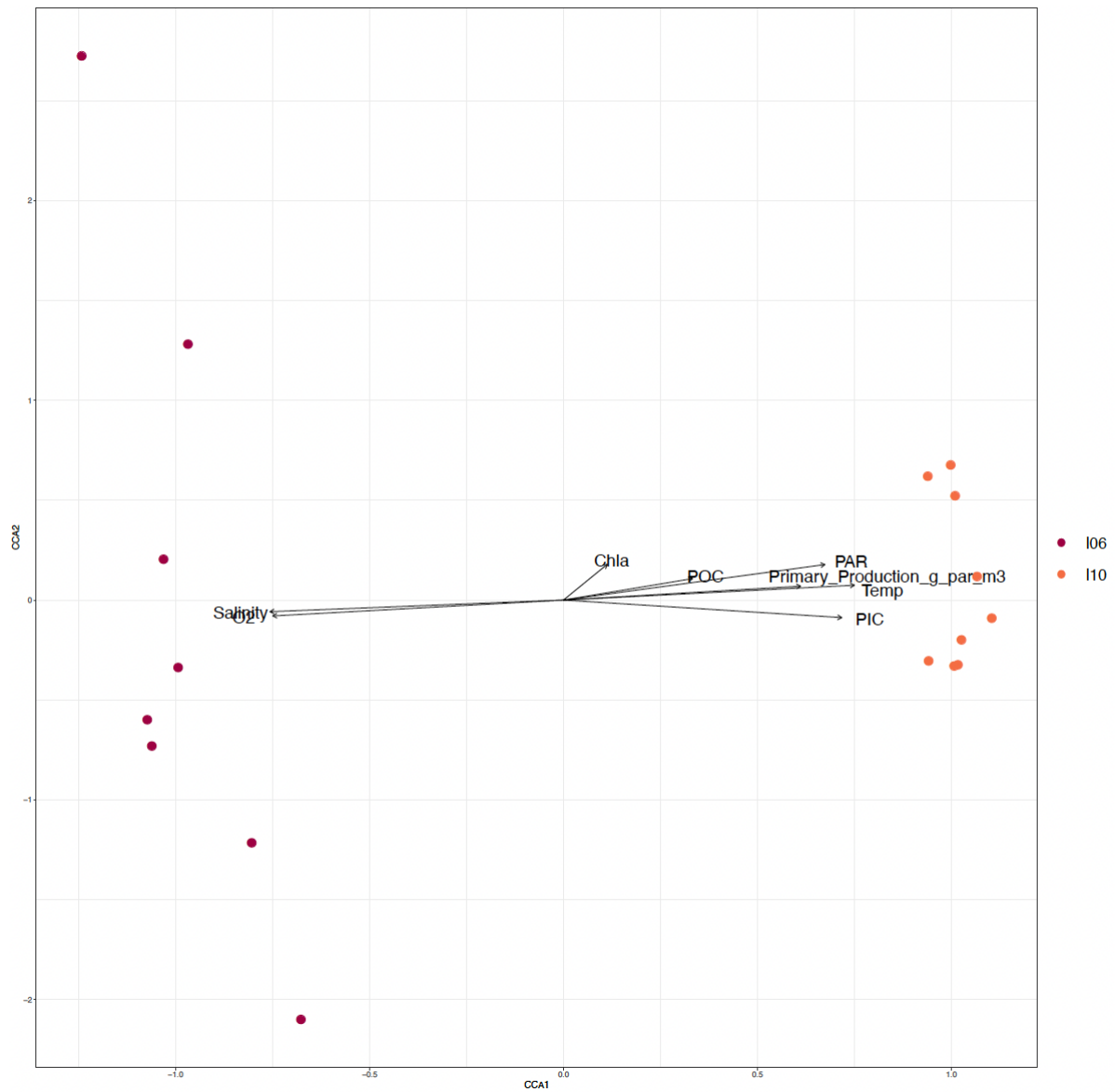


Figure III.13: CCA of gene expression from DEG normalized in TPM. *Cladocopium C15* DEG in *Porites sp.* samples; Each color indicates an island; Arrows indicate environmental parameters available and chosen for the CCA. I06: Gambier and I10: Samoa.

### III.2.2.5 Genetic adaptation of *Cladocopium C1* across the Pacific Ocean

In a final differential expression analysis, I took in account only the diversity of the symbiont and then did differential expression analyses between the different *Cladocopium C1* lineages (Figure III.14).

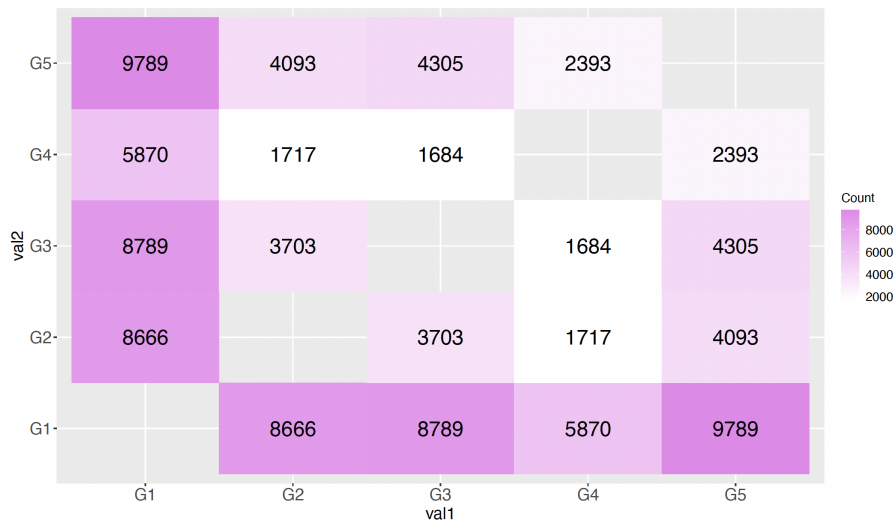


Figure III.14: Number of DEG between each *Cladocopium C1* lineages defined, in *Pocillopora sp.* samples.

G1: Lineage 1; G2: Lineage 2; G3: Lineage 3; G4: Lineage 4 and G5: Lineage 5.

It was observed that lineage 1 which represents the samples from Rapa Nui island has a very different expression profile from the other lineages (more than 5.870 DEG with the other lineages). Again, this lineage showed a different profile that could be linked to the genetic and the environmental adaptation of *Cladocopium C1* in these different conditions.

Then, I observed that the numbers of DEG in some lineages may be related to the fact that there are few samples in some lineages. These results were showed for the lineages 2 and 4 for example.

To conclude, these analyses of DEG between each *Cladocopium C1* lineage, gave some clues to the genetic adaptation of *Cladocopium C1* in different island, but due to some biases in these analyses it was difficult to interpret these different expression profiles between different lineages.

### III.2.3 Differentially expressed functions for the adaptation of Symbiodiniaceae in the Pacific Ocean

Following the results obtained, the functions associated with differentially expressed genes were a last part of the analyses. The aim of this part was to understand the metabolic pathways involved in the genetic and environmental adaptation of Symbiodiniaceae across the Pacific Ocean.

To do this, bioinformaticss methods on the different datasets were applied using several databases such as Gene Ontology (GO), Protein families (PFAM) or Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

The annotations of *Symbiodinium A3* and *Cladocopium C1* genomes were used. However, due to the lack of sufficient information for the *Symbiodinium A3* genome, a functional re-annotation was performed.

#### III.2.3.1 Expression and associated functions for the adaptation of *Symbiodinium A3* in *Millepora sp.* samples across the Pacific Ocean

First, I performed a re-annotation of the *Symbiodinium A3* genome due to the lack of annotations on the *Symbiodinium A3* genes. Then, to analyse the genome annotation, the most enriched PFAM2GO genes were selected

As expected, the major metabolic pathways of this genome are linked to general biological processes in Symbiodiniaceae. There are fatty acids metabolic pathway (glycolic process) and sugars pathway (glucose metabolic process, polysaccharide process, gluconeogenesis, fructose etc.), but also functions associated with photosynthesis (oxidation-reduction process) and nutrient transport (transmembrane transport and ion transport).

These results showed that the re-annotation of *Symbiodinium A3* genome gives results expected in term of functions and it also showed that this re-annotation was necessary to have more genes annotated in *Symbiodinium A3* reference.

In a second study, I looked at the PFAM2GO enriched in each comparison of DEG profiles, taking in account the genes over-expressed and under-expressed in any given comparison. The results showed:

- **Coïba vs Gambier:** The genes involved in DNA replication (DNA replication), photosyn-

thesis (oxidation-reduction process) and transport (location) are over-expressed in Coïba compared to Gambier. In comparison, the genes involved in carbohydrate processes are under-expressed in Coïba compared to Gambier.

- **Coïba vs Moorea:** The genes involved in translation, ammonium transport and photosynthesis are under-expressed in Coïba compared to Moorea, while the genes involved in carbohydrate processes are over-expressed in this comparison.
- **Coïba vs Aitutaki:** Looking at these two islands, genes are under-expressed in Coïba compared to Aitutaki in the metabolic pathways of nutrient transport (ammonium, nitrogen, ion, metal). In comparison, the genes are over-expressed in Coïba compared to Aitutaki in the process of carbohydrates and on lipid synthesis.
- **Coïba vs Niue:** Similar results from the other comparisons with Coïba are observed with the island of Niue.
- **Gambier vs Moorea:** Between these two islands, there is an under-expression of the genes involved in photosynthesis for Gambier compared to Moorea.
- **Gambier vs Aitutaki:** Comparison of these two islands results showed genes that are under-expressed and over-expressed in the carbohydrate process.
- **Gambier vs Niue:** The comparison of these two islands gave genes over-expressed and under-expressed in general metabolic pathways of Symbiodiniaceae.
- **Moorea vs Aitutaki:** Stress-related genes seemed to be under-expressed in the island of Moorea unlike those of Aitutaki. Genes involved in DNA repair are also under-expressed in Moorea compared to Aitutaki.
- **Moorea vs Niue:** The comparison of these two islands gave results very similar to those observed between Moorea and Aitutaki islands. It appeared genes involved in the response to oxidative stress and DNA repair genes under-expressed in Moorea unlike Niue.
- **Aitutaki vs Niue:** There appeared few metabolic pathways emerging from this comparison that can be explained. These two islands are very close geographically which may explain these results.



These results showed an important complexity in the functional analyses of genes differentially expressed between two islands. These analyses did not take in account *Symbiodinium A3* lineages or the coral host lineages. A comparison of these two parameters could help to better understand the metabolic pathways involved in the environmental adaptation of *Symbiodinium A3* in the Pacific Ocean.

Similar results were analysed using exact Fisher tests and using PFAM instead of GO terms, showed similar results in the metabolic pathways involved in *Symbiodinium A3* adaptation to environment.

Then, photosynthesis and nutrients transports using PFAM2GO selection were analysed in each comparison (**Figures III.15 and III.16 and Table III.2**). The results showed that using a *priori* method gave less functions enriched in DEG comparisons. In fact, because I selected DEG and I reduced the number of DEG to analyse and therefore the number of DEG with annotations involved in the metabolic pathways selected.

Indeed, looking the DEG involved in photosynthesis (**Figures III.15 and III.16**), there are two comparisons with results. These comparisons concern the island of Moorea versus the islands of Gambier and Aitutaki. It appeared that Moorea has over-expressed genes involved in photosynthesis unlike these two islands. These results may explain the expression of some genes potentially essential in *Symbiodinium A3* in the island of Moorea compared to Aitutaki and Gambier islands to adapt in different environmental conditions.

DEG involved in nutrient transports gave also few results (**Table III.2**). There are seven comparisons with DEG selected in this metabolic pathway, and among these comparisons 3 pairwise islands (Coïba versus Gambier, Moorea and Aitutaki islands) were selected. It appeared that Coïba island has over-expressed genes involved in nutrient transports which are different from those in the other islands. In fact, Coïba samples contain *Millepora intricata* coral compared to the other islands where it was a *Millepora platyphylla* coral host. It appeared that genes involved in nutrient transports in these comparisons, are essential to *Symbiodinium A3* to adapt in two different islands.

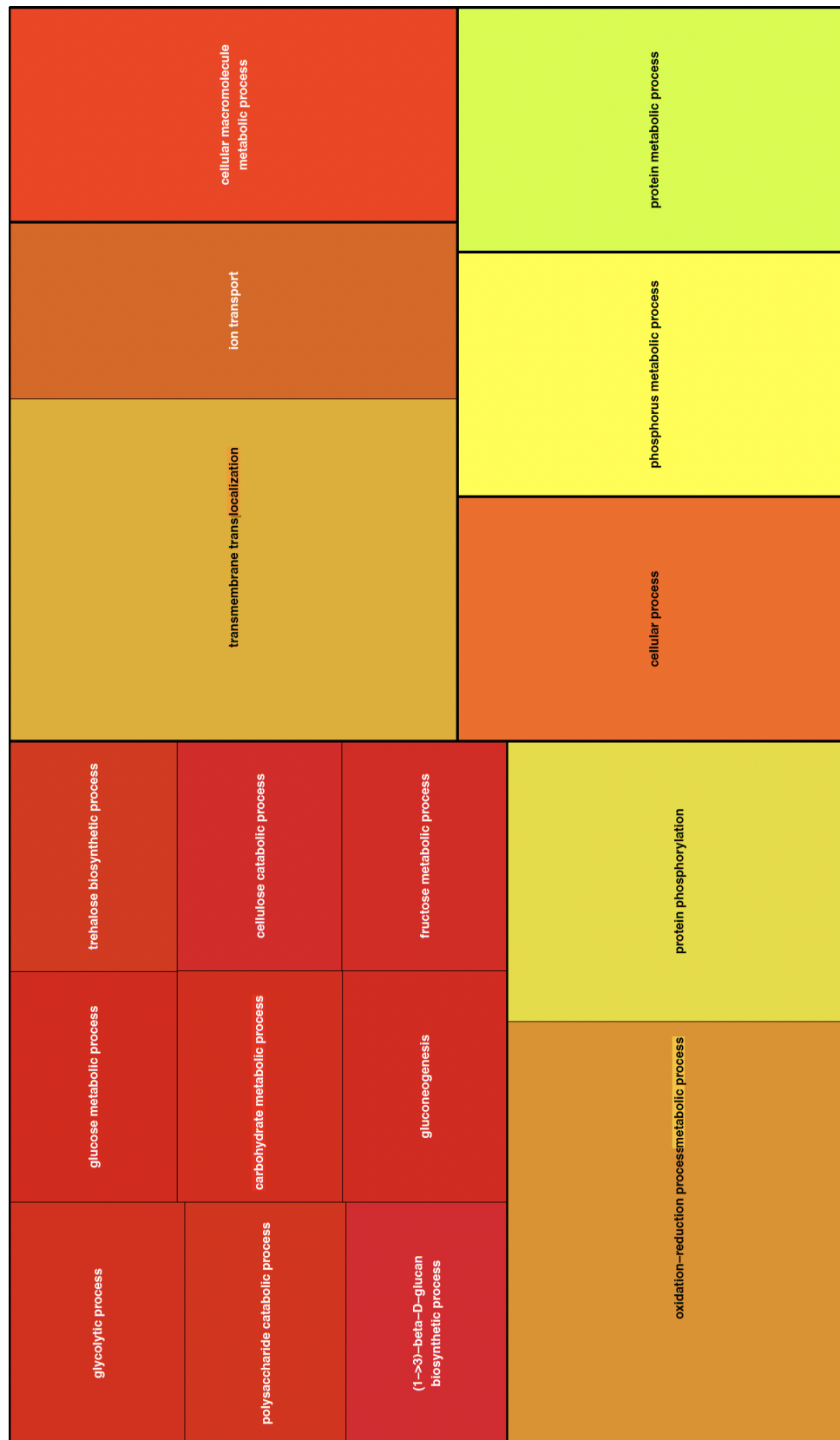


Figure III.15: The most enriched PFAM2GO in DEG of *Symbiodinium A3*, after selection of PFAM involved in photosynthesis metabolic pathway for down-regulated genes in Gambier compared to Moorea.

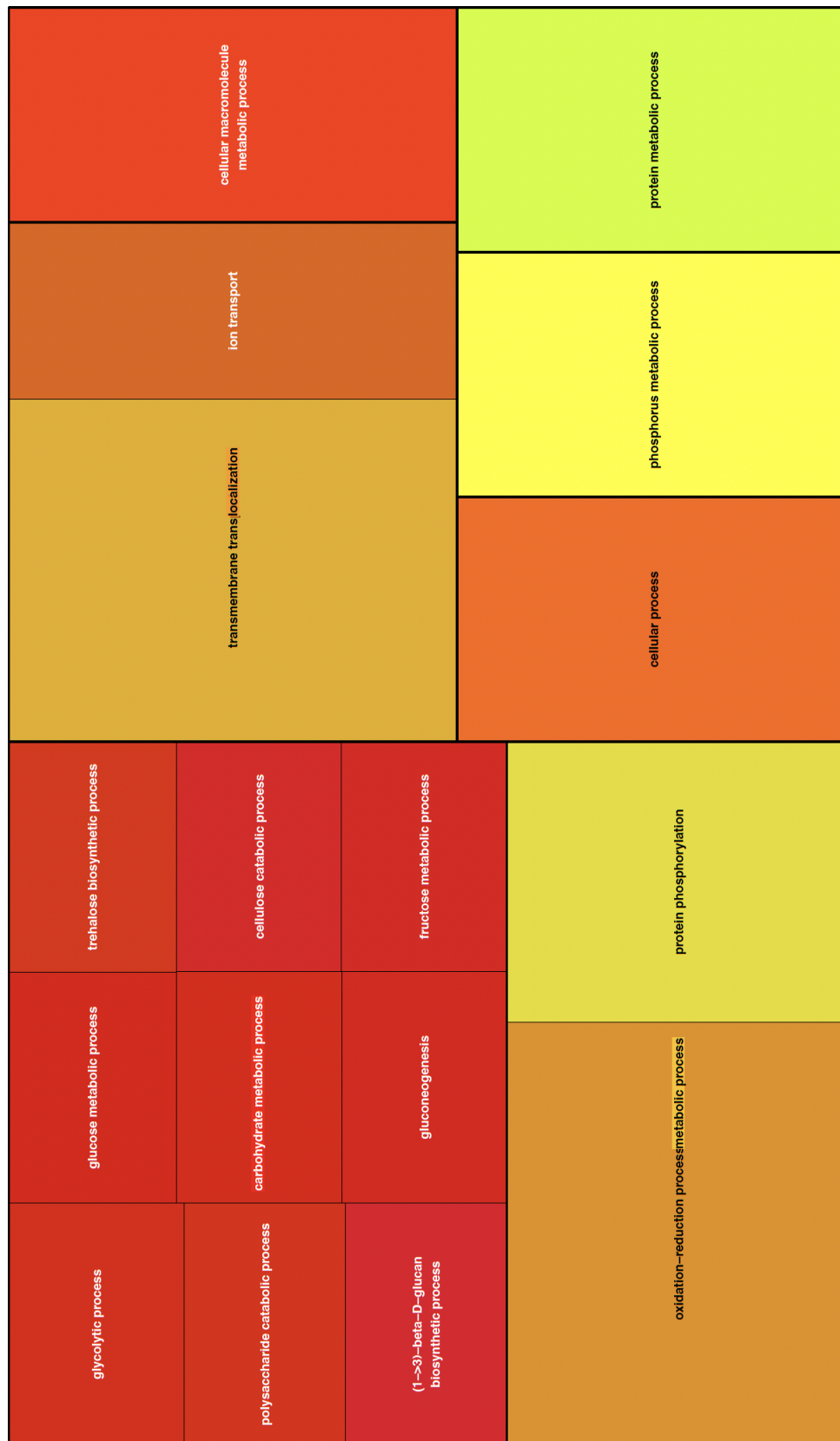


Figure III.16: The most enriched PFAM2GO in DEG of *Symbiodinium A3*, after selection of PFAM involved in photosynthesis metabolic pathway for Down-regulated genes in Moorea compared to Aitutaki.

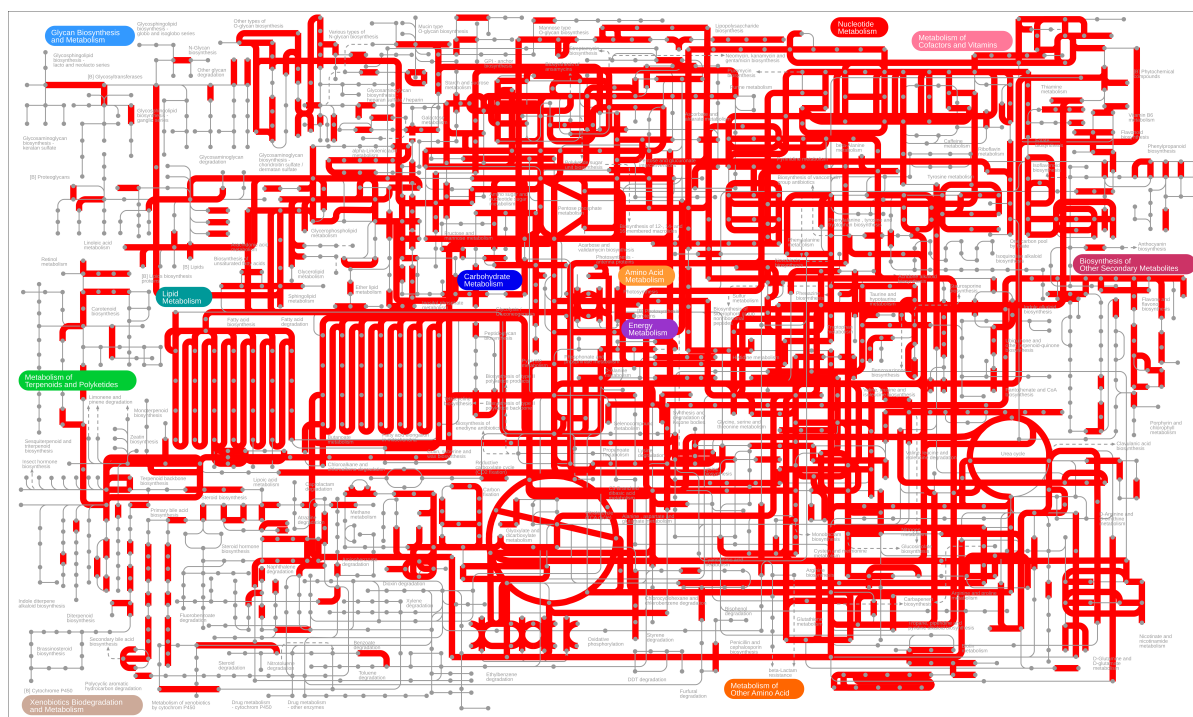
Table III.2: Number of DEG associated to GO terms involved in transport of nutrients, selected in *Symbiodinium A3*.

Conditions	ParentID	ParentGO	ChildrenID	ChildrenGO	Number of DEG
102vsl06-Down	GO:0051179	location	GO:0015696	ammonium transport	1
102vsl06-Up	GO:0051179	location	GO:0015696	ammonium transport	5
102vsl07-Down	GO:0051179	location	GO:0015696	ammonium transport	10
102vsl07-Up	GO:0051179	location	GO:0015696	ammonium transport	3
102vsl08-Down	GO:0051179	location	GO:0015696	ammonium transport	7
102vsl08-Up	GO:0051179	location	GO:0015696	ammonium transport	6

### III.2.3.2 Differentially expressed functions of *Cladocodium C1* in *Pocillopora sp.* samples across the Pacific Ocean

In this part, I performed an analysis of the *Cladocodium C1* annotation using the reference genome annotated and then I compared this annotation with the functions in the different *Cladocodium C1* lineages obtained.

First, a global view of the genes of *Cladocodium C1* was made using the iPath3 visualization tool, with associated KEGG (Figure III.17). Using this tool, it appeared that all the general metabolic pathways associated with *Cladocodium C1* are represented in the genome.

Figure III.17: KEGG enriched in *Cladocodium C1* genome, using iPath3 visualization.

In order, to compare whether the different lineages obtained in the article **Armstrong, Lê-Hoang *et al.***, a comparative analysis of the KEGG associated with the genes of each lineage was done (**ANNEXE XI - Figure 9**) using the same protocol.

As presented on each lineage, there are very few differences between the lineages enriched in KEGG in the different *Cladocopium C1* lineages and the global view of the *Cladocopium C1* (**Figure III.17**). Indeed, all of the general metabolic pathways of *Cladocopium C1* are represented for each of the different genetic lineages identically to the global view of the genome.

These results showed that using KEGG pathways to have a global point of view between the different lineages, should not influence the enrichment results, because all metabolic pathways are enriched independently of the *Cladocopium C1* lineages.

In an other part, I performed functional analyses on the DEG of *Cladocopium C1* taking in account the genetic lineage of the symbiont and the environment (islands). These results are presented in the article **Armstrong, Lê-Hoang *et al.* (submitted)**.

Additional analyses not presented in the article were also performed using an enrichment with the TopGO package. These results showed enrichment in various general metabolic pathways of *Cladocopium C1*, but because the number of samples influences the results of differential expression analyses, for some comparisons it was difficult to conclude on these results with this method.

Moreover, with the high diversity of the GO terms in *Cladocopium C1* DEG, the results were difficult to interpret. Therefore, the same analyses were done using PFAM instead GO terms to have a clearer view of the metabolic pathways enriched in the different comparisons by taking in account *Cladocopium C1* lineages, host genetic and environment (islands) (**Armstrong, Lê-Hoang *et al.* submitted**).

In a last functional analyse, I carried out functional analyses of the DEG taking in account only the genetic lineages of *Cladocopium C1* (**APPENDIX XI - Table 7**).

The results of enrichment using Fisher's exact tests were filtered with an adjusted p-value of  $< 0.01$  and showed DEG enriched using PFAM between each different *Cladocopium C1* lineages. These enrichments showed PFAM associated with the different functions such as:

- Chlorophyll A-B binding protein.

- Cyclic nucleotide-binding domain.
- Galactose oxidase, central domain.
- Ion transport protein.
- Major Facilitator Superfamily.
- Major intrinsic protein.
- Papain family cysteine protease.
- Pyruvate flavodoxin/ferredoxin oxidoreductase.
- thiamine diP-bdg.
- Pyruvate phosphate dikinase.
- AMP/ATP-binding domain.
- Reverse transcriptase (RNA-dependent DNA polymerase).
- Serine carboxypeptidase

Looking at the associated comparisons, there is an enrichment of lineage 1 compared to other lineages. Indeed, this lineage emerges enriched in PFAM for the associated functions indicated and these functions could be explain some metabolic pathways involved in this lineage compared to the other *Cladocopium C1* lineages.

Lineages 2 and 4 also appeared with enriched PFAM. It was difficult to conclude on these two lineages because they contain very few samples, despite a global distribution in the general metabolic pathways of Symbiodiniaceae for these lineages.

### III.3 Transcriptomic plasticity and symbiont shuffling underpin *Pocillopora* acclimatization across heat-stress regimes in the Pacific Ocean

Specific studies on the *Pocillopora* sp. coral and its symbiont *Cladocopium* C1 were the subject of a study presented in the article **Armstrong, Lê-Hoang et al.**

In the framework of this article, I carried out the transcriptomic and genomic analyses of the symbiont *Cladocopium* C1. These analyses concerned the diversity of Symbiodiniaceae in the samples of *Pocillopora* sp., in a first time using the transcriptomic diversity of the symbiont affiliated to the two genera *Cladocopium* and *Durusdinium*. Then, I carried out a *Cladocopium* C1 population structure (**Chapter 2**).

After, the differential expression analyses associated with these different genetic lineages were done. The goal was to understand the genes involved in the environmental adaptation of *Cladocopium* C1. To do this, I carried out differential expression analyses associated with the genetic of *Cladocopium* C1 to understand its adaptation in different coral hosts.

From these genes and among other analyses (Variation Partitioning, DAPC and Eve analyses), I performed all the functional analyses of *Cladocopium* C1 genes.

## Transcriptomic plasticity and symbiont shuffling underpin *Pocillopora* acclimatization across heat-stress regimes in the Pacific Ocean

Eric J Armstrong<sup>1,2,3,\*‡</sup>, Julie Lê-Hoang<sup>1,2,‡</sup>, Quentin Carradec<sup>1,2</sup>, Jean-Marc Aury<sup>1</sup>, Benjamin Noel<sup>1</sup>, Julie Poulain<sup>1</sup>, Caroline Belser<sup>1</sup>, Corinne Da Silva<sup>1</sup>, Patrick Wincker<sup>1,2</sup>, Tara Pacific Consortium

‡ These authors contributed equally to this manuscript.

<sup>1</sup> Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France.

<sup>2</sup> Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GO-SEE, 3 rue Michel-Ange, 75016 Paris, France

<sup>3</sup> PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, Perpignan, France

\* Corresponding author: [armstrong@berkeley.edu](mailto:armstrong@berkeley.edu)

### ABSTRACT

The characterization of adaptation and acclimation capacities of coral holobionts is crucial for anticipating the impact of global climate change on coral reefs. Understanding the extent to which the coral host and its photosymbionts contribute to adaptive and/or plastic responses in the coral metaorganism is important. In this study, we highlight new and complex links between coral genomes, transcriptomes, and environmental features in Pocilloporid corals at basin-wide scale. We analyzed metagenomic and metatranscriptomic sequence data from *Pocillopora* colonies sampled from 11 islands across the Pacific Ocean in order to investigate patterns of gene expression in both the host and photosymbiont across an environmental gradient. Single nucleotide polymorphisms (SNPs) analysis partitioned coral hosts and algal photosymbionts into five genetic lineages each. We observed strong host-symbiont fidelity across environments except at islands where recent and/or historical heat stress may have induced a symbiont shift towards more heat-tolerant lineages in some colonies. Host gene expression profiles were strongly segregated by genetic lineage and environment, and were significantly correlated with several historical sea surface temperature (SST) traits. Symbiont expression profiles were less dependent on environmental context than the host and were primarily driven by algal genotype. Overall, our results suggest a three-tiered strategy underpinning thermal acclimatization in *Pocillopora* holobionts with 1) host-photosymbiont fidelity, 2) host transcriptomic plasticity, and 3) photosymbiont shuffling playing progressive roles in response to elevated SSTs. Our data provide a reference for the biological state of coral holobionts across the Indo-Pacific and demonstrate the power of disentangling environmental and genetic effects to provide new insights into corals' capacities for acclimatization and adaptation under environmental change.

Keywords: *Pocillopora*, *Symbiodiniaceae*, phenotypic plasticity, local adaptation, environmental stress response

### 1. INTRODUCTION

Coral reefs are ecologically and economically important ecosystems whose existence depends upon the mutualistic, photosymbiotic, association between certain Cnidarian hosts and their dinoflagellate symbionts. Breakdown of this association can lead to expulsion of algal symbionts



and ultimately result in coral mortality and reef loss (Hoegh-Guldberg 1999; Hughes *et al.* 2017, 2018). Recent environmental perturbations, particularly rising sea surface temperatures and increasing frequency of extreme heating events (i.e., marine heatwaves), are disrupting coral-dinoflagellate photosymbioses worldwide, increasing the frequency and severity of coral bleaching events (Hughes *et al.* 2018). There is therefore rising concern about the potential for heating-driven local extirpation of coral species and severe reef loss in the coming century. However, in some regions, corals exhibit higher than average heat tolerances (Osman *et al.* 2018) or show the ability to rapidly respond to and recover from acute thermal challenges (Savary *et al.* 2021). Repeated sublethal exposure to elevated sea surface temperatures can also drive positive acclimatization (phenotypic plasticity) and/or selection for thermotolerant genotypes (i.e., local adaptation; Middlebrook *et al.* 2008; Savary *et al.* 2021). However in other species, warm preconditioning had no or negative effects on holobiont performance during subsequent thermal challenges (Middlebrook *et al.* 2012; Schoepf *et al.* 2019). The capacity for thermal acclimatization and/or adaptation may therefore play an important role in determining potential ‘winners and losers’ among coral species in the face of global change (Wright *et al.* 2019; Savary *et al.* 2021). Determining whether acclimatization capacity can keep pace with projected environmental change and in which symbiotic assemblages, is therefore of paramount importance for predicting responses of the coral-dinoflagellate photosymbiosis under a warming climate.

Adaptation to the local thermal environment has been well-documented across numerous phyla (Hereford 2009; Sanford and Kelly 2011) including hermatypic corals (Kenkel *et al.* 2013; Kenkel and Matz 2017). In some cases, symbiont shuffling from thermally-sensitive to thermally-tolerant Symbiodiniaceae confers a higher resistance of the coral holobiont to heat stress (Cunning *et al.* 2015). In other cases, rapid thermal acclimation is achieved through altered gene expression in the host and/or symbiont (i.e., transcriptomic plasticity) resulting in increased whole organism (holobiont) thermal tolerance (Kenkel and Matz 2017; Savary *et al.* 2021). Coral holobiont thermal sensitivity is therefore dependent on an array of interacting drivers, including environmental history (Middlebrook *et al.* 2008; Safaie *et al.* 2018; Schoepf *et al.* 2020; Wall *et al.* 2021; Savary *et al.* 2021), endosymbiont community composition (Hoadley *et al.* 2019; Qin *et al.* 2019; Cunning and Baker 2020; Dilworth *et al.* 2021), and host genotype (Barshis *et al.* 2013; Dilworth *et al.* 2021; Drury *et al.* 2021). However, the relative role of each of these factors in determining holobiont expression *in situ* remains poorly resolved. A remaining challenge, therefore, is to better understand how environmental context affects the complex interplay between the host and symbiont and to what extent certain genotypes and or host/symbiont pairings might confer resistance to projected environmental change.

Understanding how corals will respond to future climate change requires prior knowledge regarding their capacities for acclimatization (e.g., holobiont flexibility and transcriptomic plasticity) as well as their capacities for adaptation. The mechanisms underlying these capacities as well as their limits and extent among different coral holobiotypes and the relative roles of each symbiotic partner remain under intense investigation. In addition, holobiont transcriptomic adaptation/acclimation can be difficult to assess in some cases because gene expression can be highly variable between polyps within a single coral colony (Drake *et al.* 2021). Finally, acclimation or adaptation of a coral holobiont at the local (reef) scale is often not

representative of capacities at larger (ocean) scales and therefore does not allow for accurate global projections of the survival or decline of a species or of reef ecosystems.

Given the variety of components involved in maintaining the coral-dinoflagellate mutualism against environmental perturbation, recent analyses have begun to adopt a more integrative and multidisciplinary approach to understanding the impacts of climate change on coral health. Integration of various analysis methods (genomic, transcriptomic, barcoding, imaging, etc.) is needed to draw a global picture of coral holobiont adaptation and acclimation capacities. The *Tara* Pacific expedition provides a unique opportunity to address these important questions because it allows for comparative analysis of holobiont regulation at ocean scale - covering many different local environments as well as different species (Planes et al. 2019). To this end, we investigated gene expression profiles of 102 colonies of *Pocillopora* spp. and their associated endosymbionts alongside environmental context data from 11 islands across the Pacific Ocean in order to assess the relative contributions of environmental and genetic factors in determining coral holobiont gene expression *in situ*. We used genome-wide gene expression profiling, and variation partitioning to examine the role of transcriptomic plasticity in local adaptation among various *Pocillopora* hosts and their associated dinoflagellate endosymbionts (*Symbiodiniaceae*) across environmentally distinct reef environments.

## 2. MATERIALS & METHODS

### 2.1. Site selection, coral colony sampling, and environmental metadata collection

A total of 102 *Pocillopora* spp. colonies from 11 islands (Islas de las Perlas, Coiba, Malpelo, Rapa Nui, Ducie, Gambier, Moorea, Aitutaki, Niue, Upolu, and Guam) were sampled as part of the *Tara* Pacific expedition (Planes et al. 2019). At each island, fragments of three coral colonies were collected from each of three reef sites yielding a total of ca. 9 colonies sampled per island (Table 1). Reef site sampling protocols are presented in detail in Gorsky et al. 2019. Fragments of sampled coral colonies were removed from the reef and brought back on board the *Tara* vessel for processing of DNA/RNA. Several environmental parameters were measured *in situ* at the time of collection and historical sea surface temperature (SST) data (including degree heating week data) for each sampling site were obtained as described by Gorsky et al. (2019; Table 1).

### 2.2. DNA/RNA isolation, library preparation, and sequencing

Coral fragments were processed to extract and isolate DNA/RNA as described by Belser et al. (*in prep*).

### 2.3. Host/Symbiodiniaceae diversity and *Cladocopium* population structure

#### 2.3.1. Host lineage assignment

We performed host lineage assignment as described by Armstrong et al. (*in prep*; Figure S1). Briefly, we identified a set of genome-wide single nucleotide polymorphisms (SNPs) from metagenomic reads mapped to the *Pocillopora meandrina* genomic reference (Aury et al. *in prep*) using the Genome Analysis Toolkit program (GATK, v3.7.0; Van der Auwera and O'Connor 2020). We followed the best practices guide for variant discovery with GATK which included indexing of the genomic reference (picardtools v2.6.0, CreateSequenceDictionary),

followed by identification of realignment targets (GATK RealignerTargetCreator) and realignment around detected indels (GATK, IndelRealigner). Variants were called for each colony individually (GATK, HaplotypeCaller) and resulting variant call files (VCFs) were merged into island-specific cohorts (GATK, CombineGVCFs) before performing joint genotyping across all 11 islands (GATK, GenotypeGVCFs) with polyploidy defined at 1 (Poplin *et al.* 2018). The resulting SNPs were filtered using VCFtools (v0.1.12; Danecek *et al.* 2011) to include only biallelic sites with minor allele frequencies  $\geq 0.05$ , quality scores  $\geq 30$ , and no missing data across colonies. These curated SNPs were then used to identify divergent host lineages using a coalescent analysis in RaxML.

### 2.3.2. *Cladocopium* lineage assignment

To determine which *Symbiodiniaceae* lineages were present in each coral colony we first used PCR-amplified ITS2 reads sequenced for each sample. ITS2 profiles were obtained with the SymPortal method (Hume *et al.* 2019) which uses the intragenomic diversity of *Symbiodiniaceae* ITS2 to define ITS2 type profiles based on consistent co-occurrence of intragenomic ITS2 variants across all samples. After assigning ITS2 profiles to each colony, we further analyzed the 83 coral colonies containing only a single *Cladocopium* ITS2 profile in order to determine their population structure across the sampled islands. SNPs from these 83 colonies were identified from metatranscriptomic reads previously aligned on the predicted coding sequences of *Cladocopium goreau* genome (see next paragraph) following the same best practices protocol as described for the host above. SNPs were filtered using VCFtools (v0.1.12; Danecek *et al.* 2011) to include only biallelic SNP with a quality score  $\geq 30$  and a coverage  $\geq 4$  in the 83 samples. The frequencies of the 3,712 SNP were clustered using the Hclust function in R with Complete Linkage clustering method.

### 2.4. Gene expression levels of *Pocillopora* and *Cladocopium* from metatranscriptomic reads

Metatranscriptomic reads (Illumina-generated 150-bp PE) were aligned to predicted coding sequences of the *Pocillopora meandrina* host reference genome (Aury *et al. in prep*) as well as to predicted coding sequences of the *Cladocopium goreau* genome (Liu *et al.* 2018b) and *Durusdinium D2* transcriptome (Ladner *et al.* 2012) using Burrows–Wheeler Transform Aligner (BWA-mem, v0.7.15) with the default settings (Li and Durbin 2009). Host- and symbiont-mapped reads were then sorted and processed using SAMtools v1.10.2 (Li *et al.* 2009) to generate respective bam files. A read was considered a host contig if its sequence aligned to the *P. meandrina* predicted coding sequence with  $\geq 95\%$  of sequence identity and with  $\geq 50\%$  of the sequence aligned. *Cladocopium* reads aligned with *Cladocopium goreau* coding sequences and *Durusdinium D2* transcriptome were mapped against each of the symbiont databases and assigned in the same manner as described above except with a more stringent assignment cutoff of  $\geq 98\%$  of sequence identity and with  $\geq 80\%$  of the read length of the sequence aligned. All assigned reads were further filtered to remove sequences which contained  $> 75\%$  of low-complexity bases and  $< 30\%$  high-complexity bases and resulting read alignments were visualized using the Integrated Genomics Viewer (Robinson 2011). Read counts were normalized as transcript per million (TPM).

## 2.5. Bioinformatic analyses

To evaluate the relative influence of the environment and genotype on coral host and *Cladocopium* expression profiles we used three complementary approaches, namely (1) quantitative partitioning of gene expression variance into the fraction attributable to each driver using a linear mixed model approach (variation partitioning), (2) discriminant analysis of principal components (DAPC), and (3) constrained correspondence analysis informed by both historical and *in situ* environmental drivers. We also performed a phylogenetically-informed ANOVA to identify genes whose expression variation was linked to lineage divergence within *Pocillopora* and *Cladocopium* C1 and which therefore represent potential adaptive loci between lineages.

### 2.5.1. Quantitative partitioning of gene expression variance

In the first approach, we used the `variancePartition` package in R (v 1.21.2; Hoffman and Schadt 2016; Hoffman and Roussos 2021) to partition the variance attributable to the environment, the host genotype, and the symbiont genotype in the *Pocillopora* and *Cladocopium* expression datasets. Prior to performing the variation partitioning analysis, a correction was applied to remove a bias due to the two different library preparation methods used for RNA sequencing following the procedure described by Hoffman *et al.* Three variables were defined as random effects: *Pocillopora* genetic lineage, *Cladocopium* genetic lineage, and the sampling island. To take into account the sampling effect on the variance for each gene we repeated the variation partitioning analysis 100 times with a random removal of two samples. Genes with a median explained variance > 50% for one of the three selected variables across these 100 rounds were selected for downstream analysis. For each selected gene, a z-score (the number of standard deviations from the mean expression of all samples) was calculated for each island or genetic clade. A z-score > 0.5 or < -0.5 was used to consider the gene as over or under-expressed respectively. These differentially expressed genes across all samples, hereafter referred to as top variant genes, were retained for functional enrichment analyses and were represented with the `upsetR` package (Conway *et al.* 2017).

### 2.5.2. Discriminant analysis of principal components

In the first approach, we used the `DAPC` package in R (Jombart 2008; Jombart *et al.* 2010) to determine which genes are best able to discriminate between pre-defined groups of samples (i.e., based on either their genetic lineage or on their environment of collection) and to assess how well these groups could be distinguished from one another based on similarity among their expression profiles. The purpose of DAPC is to find the linear combinations of genes which maximize the differences between pre-specified groups while minimizing the within-group variance thereby allowing for the determination of which genes most strongly discriminate between the pre-defined groups. Coefficients of these genes are called loadings and higher loading scores indicate a stronger discriminatory ability. The linear combinations of these genes' expression values are referred to as discriminant functions and serve to orient the groups in multi-dimensional space.

We performed three discriminant analyses each for the host and algal symbiont with the first model using the island of collection (i.e., the environmental effect) as the pre-defined

grouping factor, the second using coral genetic lineage and photosymbiont genetic lineage as the grouping factor for the coral and photosymbiont, respectively, (i.e., the primary genetic effect), and the third using the genetic lineage of the symbiont (photosymbiont for the host and vice versa) as the grouping variable (i.e., the symbiotic partner effect). The input for these analyses were the normalized expression data (i.e., variance stabilized transformed counts) for all genes with 10 or more counts in at least 90% of colonies (host  $n = 28,732$ ; symbiont  $n = 20,338$ ). Firstly, we performed an ordination analysis to extract principal Components (PCs). We then computed a-scores to determine the optimal number of PCs to retain under each clustering model for subsequent cluster identification using discriminant analysis (Jombart et al. 2010). A-score optimization resulted in our selecting to retain 12 PCs for the host (for all three models) and 11 PCs for the algal symbiont (all models). Group memberships were then independently predicted for the colonies based on DAPC scores.

Proportions of colonies correctly reassigned to their pre-defined groups, PCs and discriminant functions (DF) retained, and the overall proportion of variance explained by each of the clustering models are included in Table S9. The relative strength of environmental-, genetic-, and symbiont- impacts on *Pocillopora* and *Cladocopium* gene expression was assessed by comparing the proportion of colonies that were correctly re-assigned to their pre-defined groups under each clustering model. A higher proportion of colonies correctly reassigned to their pre-defined groups indicates greater power to discriminate between divergent expression profiles and, therefore, a greater influence of that grouping factor on gene expression. In addition to this assessment of grouping-factor influence we also identified genes whose expression profiles most strongly contributed to model discriminant axes. Genes with discriminant axis loading scores within the upper quartile of scores (i.e., top 25%) were considered as significant contributors to that axis, are referred to hereafter as discriminant genes, and were retained for analysis of functional enrichments. Because there was strong host-symbiont fidelity across all lineages, there was consequently significant overlap between discriminant genes recovered from the primary genetic and symbiotic partner DAPC models. We therefore further divided these discriminant genes into two subcategories: genes which were shared between the two models (shared discriminant genes) and genes which were unique to one of the two models (unique discriminant genes). We use these subcategories to distinguish between analyses of all discriminant genes and unique discriminant genes, respectively, throughout this manuscript.

### 2.5.3. Expression variance and evolution model (EVE)

The combination of RNA-seq and genomic datasets in this study allowed us to test whether variation in gene expression differed significantly among species (i.e., expression divergence) and/or within species across environments (i.e., expression diversity or plasticity). By treating gene expression levels as quantitative traits, we used the expression variance and evolution model (EVE) to parameterize the ratio ( $\beta$ ) of population (i.e., within-species) to evolutionary (i.e., among-species) expression variance (Rohlf and Nielsen 2015; Avila-Magaña et al. 2021). This, in turn, allowed us to detect genes which may be under selection or which exhibit high plasticity across environments.

In brief, the EVE model works by identifying genes whose  $\beta$ -values fall significantly above or below a given threshold established from genes which are not under strong selection.

In the presence of stabilizing or no selection, the  $\beta$ -value for a given gene will remain constant and this constant forms the threshold for identifying  $\beta$ -values that deviate from the 'no selection' assumption. However, if the variation in gene expression within-species is significantly greater than that among-species (i.e., a high  $\beta$ -value) this indicates that that gene shows high *diversity* (phenotypic plasticity) in gene expression within a lineage and therefore is identified as a potential plastic candidate gene. Conversely, when  $\beta$  is low, indicating a higher among-species expression variation than within-species variance, this indicates that that gene shows lineage-specific *divergence* in expression and may therefore represent a potential adaptive candidate gene. The EVE model uses a likelihood ratio test (LRT) to examine whether the  $\beta$ -value for a given gene deviates significantly from the  $\chi^2$  distribution of all tested genes which is expected under the null model. We identified genes with a  $\beta$ -value greater/lower than expected at a cutoff value of FDR < 0.1 as potential plastic/adaptive candidate genes in both *Pocillopora* and *Cladocopium C1*. These genes were retained for functional enrichment analyses and are hereafter referred to as diverse and divergent genes, respectively.

#### 2.5.4. Constrained correspondence analysis of environmental data

We also performed a constrained correspondence analysis (CCA) between the host/symbiont expression data and the historical SST data in order to visualize how expression data were structured with respect to potential environmental drivers. In order to avoid redundant environmental data we first identified highly correlated variables (Pearson  $R^2 \geq 0.70$ ) and kept only one of these for further analysis. We then used the `bioenv` function of the `Vegan` package in R (v2.5.6) to identify the best subset of environmental variables so that the Euclidean distances of the scaled environmental variables would have the maximum correlation with expression profile dissimilarities (Clarke and Ainsworth 1993). Permutation testing was then implemented in order to select the top environmental variables (i.e., those with the highest explanatory power at  $p \leq 0.01$ ) driving differentiation in overall gene expression profiles across the colonies sampled.

#### 2.5.5. Analysis of functional enrichments among genes of interest

Functional annotations for predicted coding sequences of the *Pocillopora* host genome were extracted from the genomic reference as described by Aury *et al.* (*in prep*). These included Interproscan (IPR) and Pfam protein family terms as well as corresponding Gene Ontology (GO) annotations attributed with the Interpro2GO correspondence table (version 2020/06/13; Mitchell *et al.* 2015). Functional annotation of *Cladocopium goreau* genes were recovered from the published genome (Liu *et al.* 2018b).

Functional enrichment among top variant (variation partitioning) and discriminant (DAPC) genes were analyzed in the host using the `GOSeq` (v1.40.0; Young *et al.* 2010) and `GO_MWU` (Wright *et al.* 2015) packages in R. For top variant genes, the number of GO annotations assigned to genes within that interest group was compared to the number of annotations assigned to the rest of the dataset, to evaluate whether any ontologies were more highly represented within the module than expected by chance (i.e., Fisher's exact test in `GOSeq`). However, for discriminant genes, because an additional quantitative measure of gene importance (i.e., loading score) was available for each gene, we performed a Mann-Whitney U Test (`GO_MWU` package) to identify enriched GO terms. Loading scores for each discriminant

function were normalized to the highest score for that function yielding values ranging from 0 (no significant contribution) to 1 (highest contribution). We then initiated the `gomwuStats` function of the `GO_MWU` package in module mode and the normalized loading scores used in place of the module-membership (kME) values in the Mann-Whitney U Test.

*Cladocopium* functional enrichment analyses were performed using independent Fisher's exact tests for each Pfam (protein family) domain between genes present in each gene-of-interest category (top variant, discriminant, etc.) as compared to all expressed genes of *Cladocopium*. P-values were corrected for multiple testing with the Benjamini & Hochberg method implemented in the `p.adjust` function in R. For each Pfam term, the number of annotations assigned to genes within an interest group (i.e., top variant and discriminant genes) was compared to the number of annotations assigned to the genome, to evaluate whether any ontologies were more highly represented within the module than expected by chance (Fisher's exact test).

### 3. RESULTS

#### 3.1. Lineage assignment and characterization of host-symbiont specificity

Among the 102 colonies sampled across the 11 islands (Figure 1A), we identified five distinct *Pocillopora* host lineages, hereafter named SVD1 to SVD5 and observed a strong geographic partitioning of host lineages across the Pacific basin (Figure 1B). *Pocillopora* SVD 4 was the sole species present among colonies sampled from the eastern Pacific (Las Perlas, Coïba, and Malpelo) and was also present on Ducie and Gambier. Host SVD 5 was the only species sampled on Rapa Nui and was also present on Ducie and Moorea. *Pocillopora* SVD1 was restricted to the central Pacific (Ducie, Moorea, and Gambier) whereas *Pocillopora* SVD2 and SVD3 were both present in the western Pacific (Aitutaki, Niue, Upolu, and Guam). The SVD2 lineage was also present on the island of Moorea in the central Pacific.

Using the ITS2 metabarcoding data processed with the SymPortal method (see paper number 6), we identified a total of 28 ITS2 profiles across *Pocillopora* colonies (Figure S2). Whereas the majority of colonies (84 in total) displayed association with *Cladocopium* symbionts, we identified 9 colonies containing only *Durusdinium* (D2 ITS2 sequence) and 9 colonies containing both *Cladocopium* (mainly C42a ITS2 sequence) and *Durusdinium* (D2 ITS2 sequence) in varying proportions. *Durusdinium*-containing colonies were restricted to *Pocillopora* SVD4 hosts in the Eastern Pacific (Las Perlas, Coiba, and Malpelo) and one sample in *Pocillopora* SVD3 on Guam (Figure 1B).

Clustering of these *Cladocopium* populations into robust lineages based on the ITS2 was not possible due to the high diversity and the intragenomic variability of the ITS2 sequences. Therefore, we performed a single nucleotide polymorphism (SNP) analysis on the 82 colonies containing only a single *Cladocopium* species to identify major photosymbiont lineages present in the *Pocillopora* samples. SNPs were called by aligning the metatranscriptomic reads on the coding sequences of the *Cladocopium goreau* genome resulting in the detection of 3,712 SNPs distributed across 1,354 transcripts within the 82 colonies (see methods). The hierarchical clustering of these SNPs revealed five lineages of *Cladocopium*, hereafter referred to as L1 to L5 (Figure 1B and Figure S3). We then compared the ITS2 profile distances between each pair of *Cladocopium*-containing *Pocillopora* colonies (see methods section 2.3.2.) to the SNP-based clustering of *Cladocopium* populations (Figure

S4). In general, we observed a high degree of overlap between the two identification methods across *Cladocopium* lineages. However, whereas L1 and L5 are clearly isolated based on the ITS2 profile distances, L2, L3 and L4 are more similar to one another in the principal coordinate analysis and would have been difficult to identify using only ITS2 profiles. The ITS2 profile distances indicate that *Cladocopium* lineages which were undefined in the SNP clustering method belong to L5.

Lineage 1 is genetically distant from all other lineages and restricted to Rapa Nui island. L2 and L4 are both restricted to the central Pacific (Ducie, Gambier and Moorea islands). L2 is specific to the SVD1 lineage of the *Pocillopora* host and L4 to *Pocillopora* SVD4. Finally, L3 and L5 are both present in the Western Pacific (Aitutaki, Niue, Upolu, and Guam) and are also host specific in this oceanic region (L3 in *Pocillopora* SVD3 and L5 in *Pocillopora* SVD2; Figure 1B). Interestingly, L5 is also present in SVD4 in the East Pacific. In general, we observed high host-symbiont specificity at basin-wide scale in all *Pocillopora* lineages except within *Pocillopora* SVD4, which is in symbiosis with *Cladocopium* C1 and *Durusdinium*, and *Pocillopora* SVD5 which hosted both *Cladocopium* L1 and L3 symbionts.

### 3.2. Identifying the main drivers of holobiont gene expression and functional divergences

#### 3.2.1. Variation partitioning analysis

The variation partitioning approach allowed us to identify genes whose expression levels were driven primarily by the sampling environment, the host genetic lineage, or the algal genotype (see methods). For the host, among the 1,821 genes with greater than 50% of their expression variation attributed to these three factors, 57% (1,042 genes) are linked to the environment, 33% (606 genes) are associated with the host genetic lineage, and 9.5% (173 genes) are attributed to the symbiont genetic lineage (Figure 2A). For the *Cladocopium* C1 photosymbiont, we identified 2,424 genes above the 50% explained variation threshold, with 80% of the genes (1,950 genes) best explained by the genetic lineage of the symbiont, and only 10% (251 genes) and 9% (223 genes) of the genes linked to the sampling island and the genetic clade of the *Pocillopora* host, respectively (Figure 2B).

We investigated functional enrichments among these top variant genes and the results are presented in Tables S1-S6. In the *Pocillopora* host, genes with expression dependent on the island of sampling were enriched in biological process ontologies related to lipid and carbohydrate metabolism (and 17/125 genes, respectively), ion transmembrane transport (16/82 genes) including transmembrane transport of ammonium (3/11 genes) and calcium (9/64 genes), as well as regulation of DNA repair (9/102 genes) and autophagy (3/11 genes; Table S1, Figure 2C). In *Cladocopium*, genes whose expression is dependent on the island of sampling were enriched in PFam domains associated with heat shock chaperones (HSP70, 2 genes), active transport of protons across membranes (E1-E2 ATPases, 3 genes), and oligosaccharide processing (Glucosidase II beta subunit-like, 2 genes; Table S2).

Host genes whose expression was dependent on the *Pocillopora* genetic lineage were enriched in biological processes related to carbohydrate (3/29 genes) and xenobiotic transport (3/77 genes), DNA recombination (8/156 genes), and response to ionizing radiation (1/1 gene; Table S3, Figure 2E). *Cladocopium* C1 genes whose expression was dependent on the photosymbiont lineage included PFam enrichments related to ion transport (57 genes) and



papain family cysteine proteases (10 genes) which are known to show increased expression in plants in response to multiple environmental stressors (Liu et al. 2018a; Table S4).

*Pocillopora* genes whose expression was dependent on the photosymbiont lineage were enriched in biological process ontologies related to ion transmembrane transport (6/256 genes) particularly of calcium (3/64 genes), nitric oxide biosynthesis (1/1 gene), and assembly of the mitochondrial respiratory chain complex IV (1/2 genes; Table S5). Among *Cladocopium* genes whose expression is dependent on the host, 14 are involved in carbohydrate metabolism (Galactose oxidase, Glucan synthase, Glycosyl hydrolase) and gluconeogenesis (Fructose-bisphosphate aldolase, Enolase, Phosphofructokinase-2). Thirteen calcium-transporter and calcium-dependant protein kinases that have been shown to play a role in the establishment of symbiosis (Rosic et al. 2015) are also dependent on the host lineage (Table S6).

### 3.2.2. Discriminant analysis of principal components (DAPC)

To determine the main factors controlling the gene expression variations in the host and the symbiont, a discriminant analysis of principal components (DAPC) was performed. DAPC is a multivariate analysis which transforms the data using a principal component analysis (PCA) then identifies clusters using a discriminant analysis. The number of principal coordinates (PCs) and discriminant functions (DF) retained, the proportion of colonies correctly reassigned to their *a priori* groups, and the overall proportion of gene expression variance explained were calculated for the island of sampling, the primary genetic lineage, and the genetic lineage of the symbiotic partner using the gene expression of the host or the gene expression of the photosymbiont (Table S9).

For the host, the 12 PCs retained after a-score maximization in the DAPC explained 56.68% of total gene expression variance (Figure S5A). Discriminant axis eigenvalue scores were roughly equivalent between DF1 and DF2 for all three grouping models (Figure S5B-D). For the photosymbiont, the 11 PCs retained explained 73.86% of total gene expression variance (Figure S5E). Discriminant axis eigenvalue scores were highest for DF1 for all three grouping models (Figure S5F-H).

Based on proportions of colonies correctly reassigned to pre-defined groups, DAPC revealed marginally stronger influence of genotype (both primary lineage and of the symbiotic partner) than of the environment for both the host and the symbiont. When colonies were assigned to pre-defined groups based on the environment (island of sampling) we were able to correctly reassign 96% and 91% of colonies based on *Pocillopora* and *Cladocopium* expression profiles, respectively (Table S9, Figure 3A and B). Unconstrained analyses also confirmed that the distinction between environments of sampling was significant for both the host and photosymbiont (Table S10, PERMANOVA  $p < 0.001$ ).

When colonies were assigned to pre-defined groups based on primary genetic lineage, proportions of correct reassignment were maximized at 100% for both the host and algal symbiont (Table S9, Figure 4A and B). These genetic distinctions were also confirmed to be significant in unconstrained analyses for both the host and algal symbiont (Table S10, PERMANOVA  $p < 0.001$ ). Similarly, because of the strong host-symbiont fidelity we observed, when colonies were grouped by the genetic lineage of their symbiotic partner, we were able to correctly reassign 99% of colonies in the host and 100% of colonies in the algal symbiont (Table S9, Figure 5A and B). These partner distinctions were also significant in unconstrained analyses

for both the host and algal symbiont (Table S10, PERMANOVA  $p < 0.001$ ). These results indicate that the main driver of gene expression is the genetic lineage for both the host and photosymbiont with geography (i.e., local environment) being secondary.

In the coral host under the environmental model, DAPC differentiated expression profiles by geography along DF1 and by environmental factors (especially historical SST) along DF2. Colonies from the Eastern Pacific (Isla de Las Perlas, Coiba, and Malpelo) clustered around positive values of DF1 whereas colonies from the Western Pacific (Aitutaki, Niue, Upolu, and Guam) clustered at the negative end of this axis. Along DF2, colonies from islands with historically elevated SST values (Upolu, Guam, Isla de Las Perlas, Coiba, and Malpelo) clustered around positive DF2 values and colonies from the coldest islands (Rapa Nui and Ducie) clustered around negative values (Figure 3A). In the *Cladocopium* photosymbiont under the environmental model, DF1 primarily separated colonies from Rapa Nui from all others. The DF2 axis followed a similar pattern to that observed in the host, namely that colonies were distinguished from one another on this axis based on historical SST of the island of sampling (Figure 3B).

Under the host primary genetic model, DF1 primarily distinguished *Pocillopora* SVD 1 colonies from all others. The DF2 axis served to separate the remaining four SVD clades from one another, with clades SVD 2 and 4 (sister groups) clustered near one another at negative DF2 values, clades SVD 3 and SVD 1 overlapping around DF2 equal to zero, and SVD5 clustering at positive DF2 values (Figure 4A). In *Cladocopium*, DF1 of the primary genetic model, distinguished Lineages 1 (the most divergent lineage) and 2 from all others. DF2 served primarily to distinguish Lineage 2 from all others (Figure 4B).

Finally, under the symbiotic partner model, host colonies containing *Cladocopium* L2 symbionts were strongly separated from all others along DF1 whereas DF2 primarily distinguished colonies containing L1 and L3 photosymbionts (negative DF2 values) from those containing L4 and L5 photosymbionts (positive DF2 values, Figure 5A). In *Cladocopium*, DF1 of the symbiotic partner model distinguished expression profiles from photosymbionts inhabiting SVD 1 corals from all others and DF2 distinguished between photosymbiont inhabiting SVD 2 and SVD 4 colonies (negative DF2 values) from those inhabiting SVD 3 and SVD 5 colonies (positive DF2 Values, Figure 5B).

We also investigated gene ontology enrichments among genes which contributed most strongly to the two discriminant functions in each model (i.e., discriminant genes). The results of these analyses are presented in Tables S1-S6. Under the environmental model, host genes which discriminated between islands along DF1, the “geographical axis”, were enriched in biological processes related to carbohydrate binding and transport (27 and 15 genes, respectively) and glycolipid biosynthesis (13 genes, Figure 3C). Host genes which discriminated along DF2, the “SST axis”, were enriched in processes related to protein ubiquitination and repair (32 and 2 genes, respectively) and mitochondrial respiratory chain complex assembly (4 genes, Figure 3D). Functional enrichments shared across both axes included processes related to ion transmembrane transport (28/45 genes for DF1/DF2, respectively), generation of precursor metabolites and energy (2/1/4 genes for DF1/DF2/shared, respectively), immune response (3/4/7 genes), response to oxidative stress (2/3/8 genes), and cytolysis in other organism involved in symbiotic interaction (3 shared genes, Figure 3C and E). In *Cladocopium*, genes whose expression strongly distinguished photosymbionts from Rapa Nui from all others

(DF1 contributing genes) were enriched in protein family (Pfam) domains related to chlorophyll A-B binding (35 genes) and ion transport (91 genes; Table S2). *Cladocopium* genes which contributed to DF2 were not significantly enriched in any Pfam domains (Table S2).

Under the primary genetic lineage model, DF1 identified *Pocillopora* genes which distinguished SVD 1 colonies from all other lineages. These genes were enriched in biological processes related to carbohydrate transport (14 genes), lipid biosynthesis (5 genes), immune response (8 genes), tumor necrosis factor-mediated signaling (3 genes), and regulation of response to virus (3 genes, Figure 4C). Host genes which contributed most strongly to expression profile discrimination along DF2 were enriched in processes related to cytolysis in other organism involved in symbiotic interaction (3 genes), ion transport (90 genes), and protein glycosylation and phosphorylation (26 and 197 genes, respectively, Figure 4D). Functional enrichments shared across both axes included processes related to calcium ion and xenobiotic transport (9/12/16 and 14/16/15 DF1/DF2/shared genes, respectively), lipid biosynthesis and transport (1/1/4 and 6/2/8 genes, respectively), and regulation of apoptosis (28/31/16 genes, Figures 4C and E). Host genes which were unique to the primary genetic lineage model (i.e., not shared with the symbiotic partner model) were principally involved in cytokinesis cell division. These genes were enriched in processes related to centriole replication (1 gene), positive regulation of cytokinesis (1 gene), DNA replication (1 gene), and sphingolipid metabolism (1 gene) along DF1 and processes related to mitosis and meiosis (2 and 1 genes, respectively) and glycerol-3-phosphate metabolism (1 gene) along DF2. In *Cladocopium*, DF1 of the primary lineage model was enriched in Pfam domains related to chlorophyll A-B binding (37 genes) and ion transport (104 genes; Table S4). *Cladocopium* genes which contributed to DF2 were enriched in Pfam domains related to ion transport (113 genes) and P-loop containing dynein motor region (13 genes, Table S4).

According to the symbiotic partner model, host genes which distinguished *Cladocopium* L2-containing colonies (DF1 genes) from all others were enriched in biological processes related to carbohydrate biosynthesis, binding, and transport (8, 15, and 28 genes, respectively), mitochondrial cytochrome c oxidase assembly (3 genes), tumor necrosis factor-mediated signaling (3 genes), and regulation of response to virus (3 genes, Figure 5C). Host genes which discriminated between colonies along DF2 were enriched in processes related to cytolysis in another organism involved in symbiotic interaction (4 genes), phospholipid biosynthesis (8 genes), and nucleosome assembly (8 genes, Figure 5D). Functional enrichments shared across both axes included processes related to xenobiotic and calcium ion transmembrane transport (11/20/14 and 12/10/16 DF1/DF2/shared genes, respectively), carbohydrate transport (4/5/10 genes), immune response (4/2/5 genes), lipid biosynthesis (1/1/4 genes), regulation of apoptosis (26/28/19 genes). Genes unique to this model and which contributed strongly to DF1 were enriched in processes related to the TCA cycle (1 gene), negative regulation of gluconeogenesis (1 genes), positive regulation of apoptosis (1 genes), regulation of autophagy (1 genes), and response to heat (1 genes). Unique genes contributing to DF2 were enriched in processes related to polysaccharide catabolism (1 genes), meiotic chromosome segregation and chromatin assembly/disassembly (1 gene each), base-excision repair (1 gene), and ammonium transport (2 genes). In *Cladocopium*, DF1 of the DAPC discriminated expression profiles of photosymbionts residing in SVD1 hosts from all others and genes contributing to this axis were enriched in Pfam domains related to chlorophyll A-B binding (37 genes) and ion

transport (104 genes; Table S6). Genes which contributed to separation of colonies along DF2 were enriched in Pfam domains related to ion transport (113 genes), and P-loop containing dynein motor regions (13 genes; Table S6).

### 3.2.3. *Expression variance and evolution model (EVE) diverse and divergent genes*

In the *Pocillopora* host, we identified a total of 8,378 potential divergent candidate genes. There were no diverse candidate genes identified in the coral host. Host genes whose expression profiles diverged significantly between SVD clades were enriched in biological processes related to G-protein coupled receptor signaling (692/2139 genes) including activation of the innate immune response (5/9 genes), ATPase-coupled transmembrane transporter activity (47/108 genes) including transport of calcium (28/64 genes) and xenobiotic compounds (35/77 genes), carbohydrate transport (13/29 genes), glycolipid biosynthesis (11/25 genes), and cytolysis of another organism involved in a symbiotic interaction (3/4 genes).

In the *Cladocopium C1* photosymbiont, we identified 146 and 17,846 potential diverse and divergent gene candidates, respectively. We observed no Pfam enrichments among photosymbiont genes displaying diverse within-lineage expression. However, *Cladocopium C1* genes with expression profiles that diverged significantly between lineages were enriched in Pfam domains related to ion transport (145 genes), chlorophyll A-B binding (39 genes), and cyclic nucleotide-binding (39 genes).

### 3.2.4. *Constrained correspondence analysis*

Using a constrained correspondence analysis (CCA), we identified nine environmental variables and one genetic variable from among those that were not strongly autocorrelated that best explained the dispersion among host gene expression profiles (Figure 6A, Table S11). Similarly, for the photosymbiont, we identified eight environmental variables and one genetic variable that best explained the dispersion among gene expression profiles (Figure 6B, Table S11).

For both the coral host and *Cladocopium C1* photosymbiont, mean climatological SST and primary genetic lineage were the two strongest contributors to gene expression variation between colonies. In the *Pocillopora* host, separation of gene expression profiles along constrained correspondence axis 1 (CCA1 Figure 6A) was primarily driven by temperature and pH. Colonies from islands with high pH values (e.g., Rapa Nui) clustered around positive CCA1 values whereas colonies from islands with historically elevated mean SST (e.g., Upolu) clustered around negative CCA1 values. In contrast, CCA2 was primarily tied to SST variability (Figure 6A). Positive CCA2 values were associated with islands with the highest maximum surface temperature anomalies (e.g., Las Perlas and Cobia) or those with large standard deviations in the frequency of elevated SST anomaly events (e.g., Gambier). Additionally, positive CCA2 values were strongly correlated with elevated environmental chlorophyll concentrations suggesting a link to potential eutrophication. Conversely, negative CCA2 values were associated with islands with long recovery periods between extreme heating events (i.e., large degree heating week recovery period, for example Rapa Nui) and/or long degree cooling week durations (e.g., Moorea).

In the *Cladocopium C1* photosymbiont, no single environmental variable was strongly correlated with CCA1, although pH and photosynthetically active radiation (PAR) both

contributed somewhat to gene expression separation along this axis (Figure 6B). Again, colonies from islands with high pH values (e.g., Rapa Nui) clustered around positive CCA1 values whereas those closer to the equator, and therefore with higher incident PAR, tended to cluster around negative CCA2 values (e.g., Upolu). As in the coral host, CCA2 was primarily associated with temperature variability, however the specific associations differed. For example, exposure to extreme cooling events played a more significant role in structuring photosymbiont expression profiles than in the coral host (Figure 6B). Colonies from islands with long recovery periods between cooling events (i.e., Gambier and Ducie) tended to cluster around negative CCA2 values. Conversely, colonies from islands with prolonged exposure to cooling events or with elevated SSTs and environmental phosphate concentrations (e.g., Aitutaki and Niue) clustered together around positive CCA2 values. These results suggest that whereas gene expression profiles in both the coral host and photosymbiont are primarily structured by climatological SST, it is the frequency of exposure to extreme heating/cooling events that drives substructuring of gene expression in the host/photosymbiont, respectively.

### 3.3. Differential gene expression following symbiont shuffling

According to unconstrained analysis, global gene expression profiles of colonies from the Eastern Pacific (Isla de Las Perlas, Coiba, and Malpelo) differed between islands (PERMANOVA,  $p \leq 0.05$ ), but not between *Cladocopium C1*- and *Durusdinium D2*-containing colonies (Figure S6, PERMANOVA  $p > 0.05$ ). In addition, intra-island comparison of expression profiles between colonies in the Eastern Pacific revealed only a small number of genes that were differentially expressed between C- and D-containing hosts. Because the DEG counts were dependent on the number of colonies sampled from each location, we also calculated colony-normalized averages using a repeated random subsampling approach. We observed  $294 \pm 363$ ,  $129 \pm 53$ , and  $110 \pm 15$  genes up-regulated in D-containing colonies ( $p \leq 0.001$  and a LFC  $\geq 2$ ) on the three islands, respectively (Figure S7A). We also observed  $171 \pm 164$ ,  $211 \pm 102$ , and  $200 \pm 59$  genes down-regulated in D-containing colonies ( $p \leq 0.001$  and a LFC  $\leq 2$ ) on the three islands, respectively (Figure S7B). Very few DEGs were shared between pairs of islands. The number of pairwise shared DEGs ranged from 3 (up-regulated in both Las Perlas and Coiba) to 9 (down-regulated in both Las Perlas and Coiba) and no gene was commonly differentially expressed across all three islands (Figure S7A and B).

We also examined functional enrichments among genes differentially expressed between C- and D-containing colonies and these results are listed in Table S2. In general, genes up-regulated in D-containing colonies on Las Perlas, were enriched in GO biological processes related to bioluminescence (GO:0008218), xenobiotic transport (GO:0042908), gluconeogenesis (GO:0006094), mitochondrial protein catabolic process (GO:0035694), base-excision repair (GO:0006284), and transmembrane transport (GO:0055085; Table S2). Genes down-regulated in D-containing colonies on Las Perlas, were enriched in GO biological processes related to calcium ion transmembrane transport (GO:0070588), and phospholipid biosynthetic processes (GO:0008654). Genes up-regulated in D-containing colonies on Coiba, were enriched in GO biological processes related to protein phosphorylation (GO:0006468) and proteolysis (GO:0006508; Table S2) whereas down-regulated genes were enriched in processes related to chaperone-mediated protein folding (GO:0061077), chaperone-mediated protein transport (GO:0072321), synaptic vesicle transport (GO:0048489), nuclear envelope

organization (GO:0006998), protein phosphorylation (GO:0006468), cell adhesion (GO:0007155), neuron projection development (GO:0031175), and intermediate filament cytoskeleton organization (GO:0045104). Finally, genes up-regulated in D-containing colonies on Malpelo, were enriched in GO biological processes related to bioluminescence (GO:0008218), regulation of signaling receptor activity (GO:0010469), vacuolar transport (GO:0007034), and xenobiotic transport (GO:0042908; Table S2). Genes down-regulated in D-containing colonies on Malpelo, were enriched in GO biological processes related to bioluminescence (GO:0008218), regulation of apoptotic process (GO:0042981), protein phosphorylation (GO:0006468), S-adenosylmethionine biosynthesis (GO:0006556), regulation of pH (GO:0006885), notch signaling pathway (GO:0007219), MyD88-dependent toll-like receptor signaling pathway (GO:0002755), proteolysis (GO:0006508), signal transduction (GO:0007165), protein ubiquitination (GO:0016567), and netrin-activated signaling pathway (GO:0038007).

#### 4. DISCUSSION

In this study, we identify and discuss the high host-photosymbiont fidelity we observed among *Pocillopora* lineages across environments. We also investigate instances of potential symbiotic flexibility and/or breakdown which we observed among corals living on reefs which have historically experienced elevated sea surface temperatures (SSTs). We propose that these mechanisms represent a three-tiered strategy of thermal acclimatization/adaptation in *Pocillopora* corals and discuss the implications of this strategy for continued persistence of these holobionts under future ocean warming.

##### 4.1. *Pocillopora* corals exhibit high symbiotic fidelity with distinct *Cladocopium* lineages

Within *Pocillopora*, it is well known that a large diversity of Symbiodiniaceae species are found in symbiosis with *Pocillopora* colonies (LaJeunesse et al. 2007, 2010; Stat et al. 2009; Cunning et al. 2013). In total, we observed five lineages of *Cladocopium C1* and one strain of *Durusdinium (D2)* in symbiotic association with *Pocillopora* host colonies across the Pacific ocean (Figure 1B). Of the five *Pocillopora* lineages that we sampled, three exhibited near perfect symbiotic fidelity with distinct lineages of *Cladocopium* (i.e., one photosymbiont lineage for each host; Figure 1B) and this fidelity persisted regardless of the environment of sampling. For example, *Pocillopora* SVD 1 colonies, present on Ducie, Gambier, and Moorea, were always found in symbiosis with *Cladocopium C1* L2 symbionts. Similarly, *Pocillopora* SVD 2 colonies, present on several islands in the Western Pacific, always hosted *Cladocopium C1* L5 as their primary photosymbiont. Colonies of *Pocillopora* SVD 3 were present in both the Central and Western Pacific and were, save for one instance which we discuss in section 4.5 below, always found in association with *Cladocopium C1* L3. Strong fidelity between a single coral host and a single Symbiodiniaceae lineage has been widely observed across a variety of scleractinian species (Pinzón and LaJeunesse 2011; Baums et al. 2014; Thornhill et al. 2014; Howells et al. 2020). Our findings therefore support the general consensus that fidelity to a single clade of symbiont is the dominant pattern in most scleractinian corals (Baird et al. 2007).

Despite this overarching tendency towards symbiotic fidelity, we also detected two broad instances of imperfect specificity (i.e., symbiont flexibility) which warrant further discussion. In the first case, we observed a switch in symbiotic association between colonies of the same host

lineage inhabiting different environments. For example, *Pocillopora* SVD 4 colonies present on Rapa Nui (mean climatological SST 22.8 °C) were found to be in symbiosis with *Cladocopium* C1 L1 symbionts. However, when present on Ducie or Moorea (climatological SSTs of 24.2 °C and 27.6 °C, respectively), this same host lineage was found to be associated with *Cladocopium* C1 L3 symbionts (Figure 1B). Similarly, host SVD 4 colonies present in the Eastern Pacific (mean climatological SST of  $27.6 \pm 1.1$  °C for the three islands) were primarily in symbiosis with *Cladocopium* C1 L5 (although *Durusdinium* D2 was also present, see section 4.5, below). However, SVD 4 colonies from the Central Pacific (mean SST of  $24.6 \pm 0.5$  °C for the two islands) contained only *Cladocopium* C1 L4 photosymbionts. This observation suggests that both these coral hosts display flexibility in their symbiotic associations with *Cladocopium* across their range and that this flexibility may be dependent on the local thermal environment (Berkelmans and van Oppen 2006; Abrego et al. 2009). In both instances, photosymbionts belonging to lineages L3 and L5 seem to have been preferentially selected for by *Pocillopora* hosts when present on islands with historically elevated SSTs. In the second broad case of symbiotic flexibility, we observed a switch from *Cladocopium* C1 to *Durusdinium* D2 as the dominant photosymbiont in two *Pocillopora* lineages. This breakdown in symbiotic fidelity is discussed in further detail in section 4.5 below, but was particularly associated with colonies inhabiting islands which have historically experienced severe and/or prolonged warming.

#### 4.2. Expression adaptation and plasticity among coral and algal symbiont lineages

We used two approaches - variation partitioning and discriminant analysis of principal components (DAPC) - to assess the relative impact of the local environment, the primary genetic lineage, and the symbiotic partner on gene expression in both the *Pocillopora* host and the *Cladocopium* photosymbiont. Both analyses indicated a stronger influence of genetic lineage on photosymbiont gene expression than on the host (Figure 2B). Because we observed such high host-photosymbiont fidelity across the sampled colonies, it was difficult to identify which genetic factor (host or symbiotic partner lineage) was the primary driver of gene expression patterns. Given that the variation partitioning analysis indicated that for each partner (host and photosymbiont) their proper primary genetic lineage was the more influential genetic factor, we used those phylogenies, respectively, for informing our phylogenetic ANOVAs.

*Cladocopium* C1 genes whose expression varied according to the photosymbiont lineage were enriched in PFAm enrichments related to ion transport (57 genes) and papain family cysteine proteases (10 genes) which are known to show increased expression in plants in response to multiple environmental stressors (Liu et al. 2018a; Table S4). When grouped by genetic lineage of the host, *Cladocopium* C1 gene expression profiles showed very little variation between colonies within groups suggesting a strong regulation of photosymbiont gene expression within a host lineage regardless of the environment (Figure 5B). In addition, photosymbiont expression profiles strongly mirrored the *Pocillopora* host phylogeny with *Cladocopium* from closely-related sister taxa in the host (e.g., SVD 3/5 and SVD 2/4) displaying more similar gene expression profiles than those inhabiting distantly related host lineages (Figure 5B). Again, this pattern would seem to suggest very strong, and potentially highly conserved, regulation of photosymbiont gene expression by the coral host.

Host genes whose variation in expression was principally attributed to the host genetic lineage were involved in biological processes related to regulation of proteolysis, including

salvaging of amino acids, as well as genes involved in 'cellular response to abiotic stimulus', including 'regulation of the innate immune response', 'defense response to virus', and 'response to ionizing radiation' (GO:0071479, and GO:0010212; Table 2). This latter category was particularly enriched among genes overexpressed in host SVD1 present in the central Pacific.

By considering expression profile variation within its proper evolutionary context, we tested for evidence of expression-level adaptation among the five coral lineages identified in this study. We used the expression variance and evolution (EVE) model (Rohlf and Nielsen 2015; Avila-Magaña et al. 2021), also referred to as Phylogenetic ANOVA, to identify genes with higher variance among than within lineages (divergent expression genes) and vice versa (diverse expression genes). The former set of genes display lineage-specific shifts in baseline expression suggestive of directional selection, potentially as a result of adaptive front-loading and/or altered gene copy number, and are therefore potential candidates for expression level adaptation. Conversely, the latter set of genes display higher expression variation within a lineage than among species suggesting either that regulation of these genes is highly conserved (i.e., low variation across lineages) and/or that these genes are highly responsive to local stimuli. These genes are therefore candidates for expression level conservation and/or plasticity.

In both partners we observed more divergent than diverse genes thus indicating that the majority of responsive genes displayed higher among-lineage gene expression variation than within-lineage differences. This further implies that although host and photosymbiont expression profiles differ significantly between lineages, within a lineage, profiles are relatively fixed. The influence of the genetic lineage is particularly strong for the *Cladocopium* photosymbiont in which very few genes appear linked to changes in the external environment (Figure 2B). We interpret this pattern as a signal that, in the photosymbiont, prior selection (i.e., adaptive mechanisms) appear to outweigh plastic responses to the local environment and that a given lineage of *Cladocopium* displays a gene expression profile that is largely "fixed" by its genotype with limited plastic potential. This genetic influence is also strong in the *Pocillopora* host (Figure 4A). However, host expression profiles were also more responsive to the local environment (Figures 2A and 3A) indicating that plastic response likely play a larger role in host acclimatization to the environment than in the photosymbiont.

#### 4.3. Coral host transcriptomic plasticity may buffer the photosymbiont from environmental perturbation

Despite our finding that genotypic/adaptive effects may, in general, outweigh plastic responses among *Pocillopora* corals, we did observe a significant response to the environment among lineages. This was especially true for the coral host in which both the variation partitioning and DAPC approaches indicated a strong impact of the environment on coral host gene expression profiles.

Within the *Pocillopora* host, genes whose variation in expression was principally driven by differences in the local environment were involved in biological processes related to lipid metabolism/modification, cellular response to endogenous stimuli (especially growth factors), and signal transduction (Table 2). We interpret this differential sensitivity to the local environment as a signature of host-driven regulation of the photosymbiont micro-environment. Overall, we observed a reduced effect of the island of sampling on *Cladocopium C1* gene



expression within a colony as compared to that of the coral host. These results also reflect those described by Leggat *et al.* (2011) who observed a significantly more robust transcriptional response in *Acropora aspera* coral hosts than in their associated photosymbionts following a simulated bleaching event. When exposed to acute heat stress, few photosymbiont genes showed significant shifts in expression level as compared to strong up-regulation of several host genes including heat shock proteins.

We interpret this combination of high host transcriptional plasticity and relatively low photosymbiont transcriptional variation as an indication of host-driven buffering of its hosted *Cladocopium* symbionts from environmental perturbation. Additionally, the high similarity we observed in *Cladocopium* expression profiles when hosted by closely-related *Pocillopora* lineages (Figure 5B) may further indicate strong regulatory control of photosymbiont expression by the host coral. Under this model, the majority of transcriptional acclimatization to the local environment is therefore driven by gene expression changes within the coral host. These responses likely serve to maintain symbiosome homeostasis and thereby ensure continued maintenance of host-specific photosymbiont expression states and continued regulation of photosymbiont productivity across environments. This also suggests that the primary mechanism of environmental adaptation in the photosymbiont is through selection of *Cladocopium* genotypes (i.e., high symbiont fidelity), either through active selection by the host or through winnowing of unfit populations within a colony, rather than through plastic, acclimatory, responses.

#### 4.4. Elevated SST drives convergent expression among hosts and may select for heat-resistant algal genotypes

Constrained correspondence analysis of expression variation between islands, indicated that the strongest environmental factor contributing to colony gene expression was historical SST (Figure 6, Table S11). This is unsurprising given the primacy of temperature as an abiotic driver in ectothermic organisms generally (Somero *et al.* 2017) and given its specific, disruptive, effects on regulatory homeostasis within scleractinian corals (Hughes *et al.* 2018; Wall *et al.* 2021). In addition, discriminant analysis of principal components of host gene expression indicated that one of the two top discriminant functions was tightly linked to historical SST (DF2). This axis clearly distinguished colonies from Rapa Nui and Ducie islands (the coldest islands) from those inhabiting the Central Pacific (Gambier, Moorea, Aitutaki, and Niue, intermediate SSTs) and those inhabiting the the Western and Eastern Pacific (Upolu, Guam, Isla de Las Perlas, Coiba, and Malpelo, highest SSTs). Colonies from different genetic lineages and from geographically distant locations nevertheless displayed similar expression profiles along this discriminant axis, particularly when present on islands with historically elevated and/or low SSTs. This suggests a convergent plastic response to thermal challenge across *Pocillopora* lineages. Interestingly these same colonies also displayed high fidelity for three photosymbiont lineages - *Cladocopium* L2 and L5 as well as *Durusdinium* D2 (Figure 1B).

In terms of coral host transcriptomic plasticity, we observed an up-regulation of genes traditionally associated with the environmental stress response (ESR) in colonies present on historically warm reefs. For example, host lineages from warm islands showed elevated expression of a 70-kDa heat shock protein, Heat Shock Protein Family A (Hsp70) Member 12A-like. This class of chaperones is known to be involved in response to thermal stress in

scleractinian corals (Franzellitti et al. 2018). Because SSTs were not elevated on these islands at the time of sampling and because these colonies had not experienced an extreme heating event in the previous three weeks before sampling, we interpret this elevated expression of ESR genes among these colonies to be indicative of gene frontloading, a potential signal of local thermal adaptation. In addition, we observed enrichment among environmentally-responsive genes in xenobiotic transporters including both glutathione and lipid peroxidase activities. These functions may represent up-regulation of membrane-bound GSTs (MAPEG), an evolutionarily distinct class of enzymes that detoxify xenobiotic compounds and ameliorate oxidative stress (Kammerscheit et al. 2019).

In addition to transcriptomic plasticity within the cnidarian host, corals have many additional pathways permitting acclimatization to thermal challenges. At the colony level, thermal response integrates across the different components of the coral holobiont (coral host, algal photosymbiont, and associated microbiome/virome) which may respond at different timescale and/or through different physiological mechanisms. For example, whereas coral genotype is fixed within a colony, photosymbiont community composition can shift over time thereby altering the prevalence of the dominant genotype of the algal partner. Such shifts in the photosymbiont community can occur over short time scales, in response to acute thermal challenges for example (Claar et al. 2020), or over longer term periods as a result of competitive exclusion between photosymbiont lineages within a colony (Howells et al. 2020). Both mechanisms underpinning symbiont shift can indicate acclimatization of the coral holobiont to its local (thermal) environment.

In our dataset, we observed three photosymbiont lineages which appear to have been preferentially selected for in *Pocillopora* colonies inhabiting islands with historically elevated SSTs. Whereas non-selective mechanisms could account for this pattern. Dispersal limitations can result in the dominance of particular photosymbionts on certain reefs, particularly if those reefs are geographically isolated. However, dispersal limitation alone seems insufficient as an explanation for the high photosymbiont specificity we observed in this study in part because we observe the same photosymbiont communities present on reefs ca. 14,000 km apart across the Pacific basin (e.g., Coiba and Guam). The high holobiont fidelity we observed under elevated SSTs may therefore truly represent a signature of differential selection in response to local thermal conditions (i.e., local thermal adaptation).

Although we did not directly assess photosymbiont or colony fitness across environments in this study, we did examine holobiont gene expression profiles which provide some context regarding the relative physiological states of colonies within a given environment. For example, in the symbiotic partner DAPC model, one of the genes which significantly discriminated between *Pocillopora* colonies containing different *Cladocopium C1* photosymbionts was a DNAJ chaperone protein. This gene was also among the candidate adaptive genes whose expression was highly divergent across *Pocillopora* lineages. We observed reduced expression of this protein in colonies from the coldest island (Rapa Nui). However, we also observed elevated expression of this protein in colonies containing *Cladocopium L2* symbionts in the Central Pacific (all SVD 1 corals on Ducie, Gambier, and Moorea, Figure 1B and C). Interestingly this gene was only upregulated on these islands in SVD1/L2 holobiont colonies; SVD4/L4, SVD2/L5, and SVD5/L3 holobiont colonies from the same islands did not exhibit similar upregulation of this chaperone protein. This gene was also

not strongly upregulated among SVD5/L3 colonies from islands in the Western Pacific that have historically experienced much higher SSTs (and which were warmer at the time of sampling) than those of the Central Pacific. This indicates that while this DNAJ chaperone protein may play an important role in thermal acclimatization in the SVD1/L2 *Pocillopora* holobiont, it is not a significant contributor to thermal acclimatization in the SVD5/L3 holobiont.

In addition, although they represent only a small fraction of the top variant genes, *Pocillopora* genes with expression dependent on the symbiont lineage were involved in several important biological processes including those potentially related to regulation of calcification, symbiont photosynthesis, and to the scavenging of reactive oxygen species (Table 2). For example, genes with expression variation linked to symbiont lineage were highly enriched in functions related to transport of ions across cellular membranes including ATPase-coupled transmembrane transporters which are known to play a role in host calcification (Barott et al. 2015b, 2020; Zoccola et al.) and as carbon-concentration mechanisms in marine photosymbioses thereby regulating *Symbiodiniaceae* photosynthesis (Barott et al. 2015a, b; Tresguerres et al. 2017; Armstrong et al. 2018).

Finally, as discussed in section 4.2 previously, the low expression variation we observed between *Cladocopium* lineages inhabiting warm reefs may itself reflect convergent evolution in response to environmental selection. Photosymbiont expression profiles may be less affected by the local environment simply because only certain symbiont genotypes are found in warm environments. That is to say, algal expression profiles are similar because of shared ancestry/symbiont fidelity due to historical selection for heat-tolerant symbiotypes. When taken together, the strong correlation between historical SST and photosymbiont community composition, the presence of similar algal genotypes on extremely distant reefs, as well as the reduced expression of host chaperone proteins relative to colonies containing other symbiont communities all strongly suggest the presence of a selective signature in these photosymbiont lineages, thereby implying environmental specialization (i.e., local adaptation) for increased heat tolerance.

#### 4.5. Idiosyncratic host expression profiles associated with symbiont shuffling under thermal challenge

For example, both *Cladocopium*- and *Durusdinium*-dominant *Pocillopora* colonies were found in the Eastern and Western Pacific and this appearance coincides with historically elevated SST at those reefs (see section 3.5 below). One interpretation of these observations is that they provide evidence for thermally-driven symbiont switching/shuffling in these colonies (e.g., shuffling of dominant symbionts or recolonization after bleaching). However, because *Pocillopora meandrina* symbionts are generally acquired via vertical transmission, with oocytes being “seeded” with the maternal symbiont community (Marlow and Martindale 2007), it is also possible that these C- and D-containing colonies represent true, evolutionarily stable, holobiotypes which persist across generations on reefs which experience frequent and/or severe thermal stress.

The known switch between the thermo-sensitive *Cladocopium* to the thermo-tolerant *Durusdinium* (LaJeunesse et al. 2010), observed here mainly in the eastern Pacific, suggests that *Pocillopora* colonies may have recently experienced acute heat stress just prior to sampling.

In the Eastern Pacific, we observed both *Cladocopium*- and *Durusdinium*-dominant colonies of *Pocillopora* SVD 4. These colonies were present on reefs which have historically experienced frequent, extreme thermal stress providing evidence for ‘symbiont switching’ as a thermal acclimation strategy in this host. However, symbiont switching had very little effect on host gene expression *in situ* and host expression appears driven more by the environment (both at the site and island scale) than by symbiont lineage. Our data for *Pocillopora* show that switching between *Cladocopium* and *Durusdinium* algal communities had only a weak, and inconsistent, impact on host expression *in situ*. This stands in contrast to previous studies which have reported a significant role of the photosymbiont in modulating coral host gene expression (DeSalvo et al. 2010). For example, a recent study in *Montastraea cavernosa* reported that colonies containing *Durusdinium* symbionts displayed elevated expression of ESR genes even under ambient conditions relative to colonies containing *Cladocopium* symbionts (Cunning and Baker 2020). These results suggest that symbiotype significantly alters host expression even prior to environmental perturbation.

On average, we found 129 genes differentially expressed between C- and D-containing *Pocillopora* colonies in the Eastern Pacific (Table S2). In total, only 4% of gene expression variation between Eastern Pacific coral colonies was attributed to differences in symbiotype and this was not significant (PERMANOVA,  $p > 0.05$ ). In contrast, in *M. cavernosa* colonies, up to 14% of expression variation between individuals could be attributed to the switch between dominant symbiotypes (Cunning and Baker 2020). In our dataset, the environment played a much more significant role in determining *Pocillopora* gene expression with 24% of expression variation being explained by sampling island and, unlike symbiotype, reef environment contributed significantly to overall host gene expression in these colonies (PERMANOVA,  $p = 0.001$ ; Figure S5). Additionally we observed very little to no overlap in differentially expressed genes between C- and D-containing *Pocillopora* colonies among the three Eastern Pacific islands (Figure S6). This implies that there was no common influence of *Durusdinium* symbionts on host expression across reefs. Finally, we observed high gene expression variation among *Durusdinium*-containing *Pocillopora* colonies from within the same reef (Figures S5 and S6) suggesting that, even at small spatial scales, other factors were more important than symbiotype in determining host gene expression.

One explanation for the greater similarity in gene expression between C- and D-containing colonies in *Pocillopora* SVD 4 relative to *M. cavernosa* may be related to the influence of both the local thermal environment and variation in host genotype on host expression in our dataset. We observed D-dominant *Pocillopora* colonies only on reefs which have historically experienced frequent, severe heating events (Isla de las Perlas, Coiba, Malpelo, and Guam). We therefore suggest that both C- and D-containing colonies displayed elevated expression of ESR genes on these reefs explaining why we don’t observe the specific, ‘thermal priming’ influence of *Durusdinium* on the host. In contrast, the *M. cavernosa* colonies sampled by Cunning and Baker (2020) were reared under controlled conditions and represented ramets of the same genet. These controlled factors in the *M. cavernosa* colonies may have allowed for maximization of the ‘symbiotype signal’ relative to the environmental and genetic influences on host gene expression. Thus, it’s possible that the similarity in host expression we observed in *Pocillopora* between C- and D-containing colonies is a result of physiological canalization (e.g., a universal ESR) in the host which partially obscured the putative front

loading of these genes in colonies containing *Durusdinium* as observed in *M. cavernosa* under controlled conditions.

Despite the weak influence of the dominant symbiotype on overall *Pocillopora* host expression in the Eastern Pacific, we did observe a small set of genes that were differentially expressed between C- and D-containing hosts. These included genes associated with ‘chaperone-mediated protein folding’ (GO:0061077), ‘protein ubiquitination’ (GO:0016567), and ‘regulation of apoptosis’ (GO:0042981) which were among the most significantly enriched biological process terms associated with genes down-regulated in D-containing colonies relative to C-containing hosts in some reefs (Table S2). These results provide indirect support for the Adaptive Bleaching Hypothesis in that coral colonies containing the more thermally tolerant *Durusdinium* symbiont show reduced expression of ESR as compared to colonies containing *Cladocopium* symbionts inhabiting the same reef.

#### 4.6. Three-tiered strategy of thermal acclimatization in *Pocillopora* holobionts

When examined together, our metagenomic and metatranscriptomic datasets seem to imply the existence of three, potentially sequential, strategies for thermal acclimatization in *Pocillopora* holobionts: 1) selection of and high fidelity for heat tolerant *Cladocopium* lineages in warm environments, 2) high host transcriptomic plasticity which buffers photosymbiont against thermal challenge, and 3) symbiont shuffling and replacement of *Cladocopium* with *Durusdinium*. However, the effectiveness of these strategies or their relative thresholds of initiation may differ between *Pocillopora* lineages. For example, we observed a general trend towards symbiont fidelity among *Pocillopora* lineages (Figure 1B) and this fidelity appears linked to local sea surface temperatures. Similar high host-photosymbiont fidelity has also been reported in corals inhabiting some of the hottest reefs in the world in the Persian Gulf (Howells et al. 2020). This tendency for high symbiotic fidelity may therefore represent the first step in a strategy of thermal acclimatization in the coral metaorganism whereby heat tolerant photosymbiont lineages are preferentially selected/maintained under elevated SST.

Despite the trend towards high symbiont fidelity, we did observe instances of symbiont flexibility in SVD 4 and SVD 5 hosts. In both cases, this flexibility was characterized by the transition of photosymbiont communities towards two putatively heat tolerant lineages (*Cladocopium* C1 L3 and L5) suggesting that some *Pocillopora* lineages can initiate plastic shifts in symbiont community composition as a potential strategy for coping with environmental warming. However, the physiological effects of this symbiont flexibility may not be equivalent among lineages. For example, when present on the coldest island (Rapa Nui) host SVD 5 corals showed strong fidelity with *Cladocopium* L1. However, when present on warmer islands in the central Pacific, SVD 5 corals were always found in symbiosis with *Cladocopium* L3, a putative heat tolerant lineage. Despite this shift in symbiont community, however, host expression of a DNAJ chaperone protein was extremely elevated in L3-containing SVD 5 colonies relative to other lineages in the same environments. Conversely, when present in the central Pacific, host SVD 4 colonies showed strong fidelity with *Cladocopium* L4 and exhibited relatively high DNAJ expression. DNAJ was down-regulated, however, when this same lineage was present on some of the warmest reefs sampled in this study (eastern Pacific) where the SVD 4 corals were found in symbiosis with another putatively thermo-tolerant photosymbiont, *Cladocopium* L5. Although the high DNAJ expression observed in SVD 5 corals in the central Pacific could represent a

positive acclimatization response (i.e., gene frontloading) accompanying the symbiont flexibility, it could also be indicative of the onset of thermal stress in these colonies. In this case, the high DNAJ expression we observed in SVD 5 colonies relative to SVD 4 colonies present on warmer reefs may signal that the SVD 5 holobionts are already at or near their upper heat tolerance limits and may therefore only be able to rely on symbiont shuffling to resist future warming. Evidence from other coral species, however, suggests that symbiont shuffling may only be effective as a “last ditch” strategy to weather acute extreme heating events and that this strategy, while increasing survival in the short-term, may negatively impact resilience in colonies under future warming events (Claar et al. 2020).

## 5. CONCLUSION

Overall, we observed high symbiotic fidelity among *Pocillopora* lineages across environments with evidence of potential selection for heat-resistance photosymbiont lineages on islands with historically elevated sea surface temperatures. We reveal that host gene expression profiles are strongly segregated by genetic lineage and environment, and were significantly correlated with several historical sea surface temperature (SST) traits. In contrast, *Cladocopium C1* expression profiles were primarily driven by algal genotype and displayed low phenotypic plasticity across environmental gradients. Overall, our data suggest the existence of a three-tiered strategy underpinning thermal acclimatization in *Pocillopora* holobionts with strong selection for specific photosymbiont lineages in certain environments (i.e., host-photosymbiont fidelity), high host transcriptomic plasticity acting as an environmental buffer, and a ‘last ditch’ strategy of photosymbiont shuffling playing progressive roles in response to elevated SSTs. Our study has important implications for continued coral reef conservation and prediction of coral responses to future ocean warming. Our examination of the mechanisms underpinning acclimatization among lineages across an environmental gradient and our identification of potential candidate loci under thermal selection provide useful baseline metrics for informing future manipulative experiments focused on physiological mechanisms underlying these observations. Investigating the extent to which different thermal acclimatization strategies, including both transcriptional and compositional changes, are used among lineages will allow us to better manage coral reefs under future ocean warming.

## REFERENCES

- Abrego D, van Oppen MJH, Willis BL (2009) Onset of algal endosymbiont specificity varies among closely related species of *Acropora* corals during early ontogeny. *Mol Ecol* 18:3532–3543. doi: <https://doi.org/10.1111/j.1365-294X.2009.04276.x>
- Armstrong EJ, Roa JN, Stillman JH, Tresguerres M (2018) Symbiont photosynthesis in giant clams is promoted by V-type H<sup>+</sup>-ATPase from host cells. *J Exp Biol*. doi: 10.1242/jeb.177220
- Avila-Magaña V, Kamel B, DeSalvo M, Gómez-Campo K, Enríquez S, Kitano H, Rohlfs RV, Iglesias-Prieto R, Medina M (2021) Elucidating gene expression adaptation of phylogenetically divergent coral holobionts under heat stress. *Nat Commun* 12:5731. doi: 10.1038/s41467-021-25950-4
- Baird AH, Cumbo VR, Leggat W, Rodriguez-Lanetty M (2007) Fidelity and flexibility in coral symbioses. *Mar Ecol Prog Ser* 347:307–309. doi: 10.3354/meps07220
- Barott KL, Venn AA, Perez SO, Tambutté S, Tresguerres M (2015a) Coral host cells acidify symbiotic algal microenvironment to promote photosynthesis. *Proc Natl Acad Sci*

- 112:607–612. doi: 10.1073/pnas.1413483112
- Barott KL, Perez SO, Linsmayer LB, Tresguerres M (2015b) Differential localization of ion transporters suggests distinct cellular mechanisms for calcification and photosynthesis between two coral species. *Am J Physiol-Regul Integr Comp Physiol* 309:R235–R246. doi: 10.1152/ajpregu.00052.2015
- Barott KL, Venn AA, This AB, Tambutté S, Tresguerres M (2020) Regulation of coral calcification by the acid-base sensing enzyme soluble adenylyl cyclase. *Biochem Biophys Res Commun* 525:576–580.
- Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR (2013) Genomic basis for coral resilience to climate change. *Proc Natl Acad Sci* 110:1387–1392. doi: 10.1073/pnas.1210224110
- Baums IB, Devlin-Durante MK, LaJeunesse TC (2014) New insights into the dynamics between reef corals and their associated dinoflagellate endosymbionts from population genetic studies. *Mol Ecol* 23:4203–4215. doi: <https://doi.org/10.1111/mec.12788>
- Berkelmans R, van Oppen MJH (2006) The role of zooxanthellae in the thermal tolerance of corals: a ‘nugget of hope’ for coral reefs in an era of climate change. *Proc R Soc B Biol Sci* 273:2305–2312. doi: 10.1098/rspb.2006.3567
- Claar DC, Starko S, Tietjen KL, Epstein HE, Cunning R, Cobb KM, Baker AC, Gates RD, Baum JK (2020) Dynamic symbioses reveal pathways to coral survival through prolonged heatwaves. *Nat Commun* 11:6097. doi: 10.1038/s41467-020-19169-y
- Clarke KR, Ainsworth M (1993) A method of linking multivariate community structure to environmental variables. *Mar Ecol Prog Ser* 92:205–219. doi: 10.3354/meps092205
- Conway JR, Lex A, Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33:2938–2940. doi: 10.1093/bioinformatics/btx364
- Cunning R, Baker AC (2020) Thermotolerant coral symbionts modulate heat stress-responsive genes in their hosts. *Mol Ecol* 29:2940–2950. doi: 10.1111/mec.15526
- Cunning R, Glynn PW, Baker AC (2013) Flexible associations between *Pocillopora* corals and *Symbiodinium* limit utility of symbiosis ecology in defining species. *Coral Reefs* 32:795–801. doi: 10.1007/s00338-013-1036-y
- Cunning R, Silverstein RN, Baker AC (2015) Investigating the causes and consequences of symbiont shuffling in a multi-partner reef coral symbiosis under environmental change. *Proc R Soc B Biol Sci*. doi: 10.1098/rspb.2014.1725
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. doi: 10.1093/bioinformatics/btr330
- DeSalvo MK, Sunagawa S, Fisher PL, Voolstra CR, Iglesias-Prieto R, Medina M (2010) Coral host transcriptomic states are correlated with *Symbiodinium* genotypes. *Mol Ecol* 19:1174–1186. doi: 10.1111/j.1365-294X.2010.04534.x
- Dilworth J, Caruso C, Kahkejian VA, Baker AC, Drury C (2021) Host genotype and stable differences in algal symbiont communities explain patterns of thermal stress response of *Montipora capitata* following thermal pre-exposure and across multiple bleaching events. *Coral Reefs* 40:151–163. doi: <https://doi.org/10.1007/s00338-020-02024-3>
- Drake JL, Malik A, Popovits Y, Yosef O, Shemesh E, Stolarski J, Tchernov D, Sher D, Mass T (2021) Physiological and Transcriptomic Variability Indicative of Differences in Key Functions Within a Single Coral Colony. *Front Mar Sci* 8:685876. doi: 10.3389/fmars.2021.685876
- Drury C, Bean N, Harris C, Hancock J, Hucekba J, Christian Martin H, Roach T, Quinn R, Gates RD (2021) Intrapopulation adaptive variance supports selective breeding in a reef-building coral. *bioRxiv* 2021.05.21.445206. doi: 10.1101/2021.05.21.445206

- Franzellitti S, Airi V, Calbucci D, Caroselli E, Prada F, Voolstra CR, Mass T, Falini G, Fabbri E, Goffredo S (2018) Transcriptional response of the heat shock gene *hsp70* aligns with differences in stress susceptibility of shallow-water corals from the Mediterranean Sea. *Mar Environ Res* 140:444–454. doi: <https://doi.org/10.1016/j.marenvres.2018.07.006>
- Gorsky G, Bourdin G, Lombard F, Pedrotti ML, Audrain S, Bin N, Boss E, Bowler C, Cassar N, Caudan L, Chabot G, Cohen NR, Cron D, De Vargas C, Dolan JR, Douville E, Elineau A, Flores JM, Ghiglione JF, Haëntjens N, Hertau M, John SG, Kelly RL, Koren I, Lin Y, Marie D, Moulin C, Moucherie Y, Pesant S, Picheral M, Poulain J, Pujo-Pay M, Reverdin G, Romac S, Sullivan MB, Trainic M, Tressol M, Troublé R, Vardi A, Voolstra CR, Wincker P, Agostini S, Banaigs B, Boissin E, Forcioli D, Furla P, Galand PE, Gilson E, Reynaud S, Sunagawa S, Thomas OP, Thurber RLV, Zoccola D, Planes S, Allemand D, Karsenti E (2019) Expanding Tara Oceans Protocols for Underway, Ecosystemic Sampling of the Ocean-Atmosphere Interface During Tara Pacific Expedition (2016–2018). *Front Mar Sci* 6:750. doi: 10.3389/fmars.2019.00750
- Hereford J (2009) A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. *Am Nat* 173:579–588.
- Hoadley KD, Lewis AM, Wham DC, Pettay DT, Grasso C, Smith R, Kemp DW, LaJeunesse TC, Warner ME (2019) Host–symbiont combinations dictate the photo-physiological response of reef-building corals to thermal stress. *Sci Rep* 9:9985. doi: <https://doi.org/10.1038/s41598-019-46412-4>
- Hoegh-Guldberg O (1999) Climate change, coral bleaching and the future of the world’s coral reefs. *Mar Freshw Res* 50:839–866. doi: 10.1071/MF99078
- Hoffman GE, Roussos P (2021) Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* 37:192–201. doi: <https://doi.org/10.1093/bioinformatics/btaa687>
- Hoffman GE, Schadt EE (2016) variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* 17:13.
- Howells EJ, Bauman AG, Vaughan GO, Hume BCC, Voolstra CR, Burt JA (2020) Corals in the hottest reefs in the world exhibit symbiont fidelity not flexibility. *Mol Ecol* 29:899–911. doi: <https://doi.org/10.1111/mec.15372>
- Hughes TP, Barnes ML, Bellwood DR, Cinner JE, Cumming GS, Jackson JBC, Kleypas J, van de Leemput IA, Lough JM, Morrison TH, Palumbi SR, van Nes EH, Scheffer M (2017) Coral reefs in the Anthropocene. *Nature* 546:82–90. doi: 10.1038/nature22901
- Hughes TP, Anderson KD, Connolly SR, Heron SF, Kerry JT, Lough JM, Baird AH, Baum JK, Berumen ML, Bridge TC, Claar DC, Eakin CM, Gilmour JP, Graham NAJ, Harrison H, Hobbs J-PA, Hoey AS, Hoogenboom M, Lowe RJ, McCulloch MT, Pandolfi JM, Pratchett M, Schoepf V, Torda G, Wilson SK (2018) Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science* 359:80–83. doi: 10.1126/science.aan8048
- Hume BCC, Smith EG, Ziegler M, Warrington HJM, Burt JA, LaJeunesse TC, Wiedenmann J, Voolstra CR (2019) SymPortal: A novel analytical framework and platform for coral algal symbiont next-generation sequencing *ITS2* profiling. *Mol Ecol Resour* 19:1063–1080. doi: 10.1111/1755-0998.13004
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405. doi: 10.1093/bioinformatics/btn129
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94. doi: 10.1186/1471-2156-11-94
- Kammerscheit X, Chauvat F, Cassier-Chauvat C (2019) From cyanobacteria to human, MAPEG-type glutathione-S-transferases operate in cell tolerance to heat, cold, and lipid peroxidation. *Front Microbiol* 10:2248. doi: 10.3389/fmicb.2019.02248

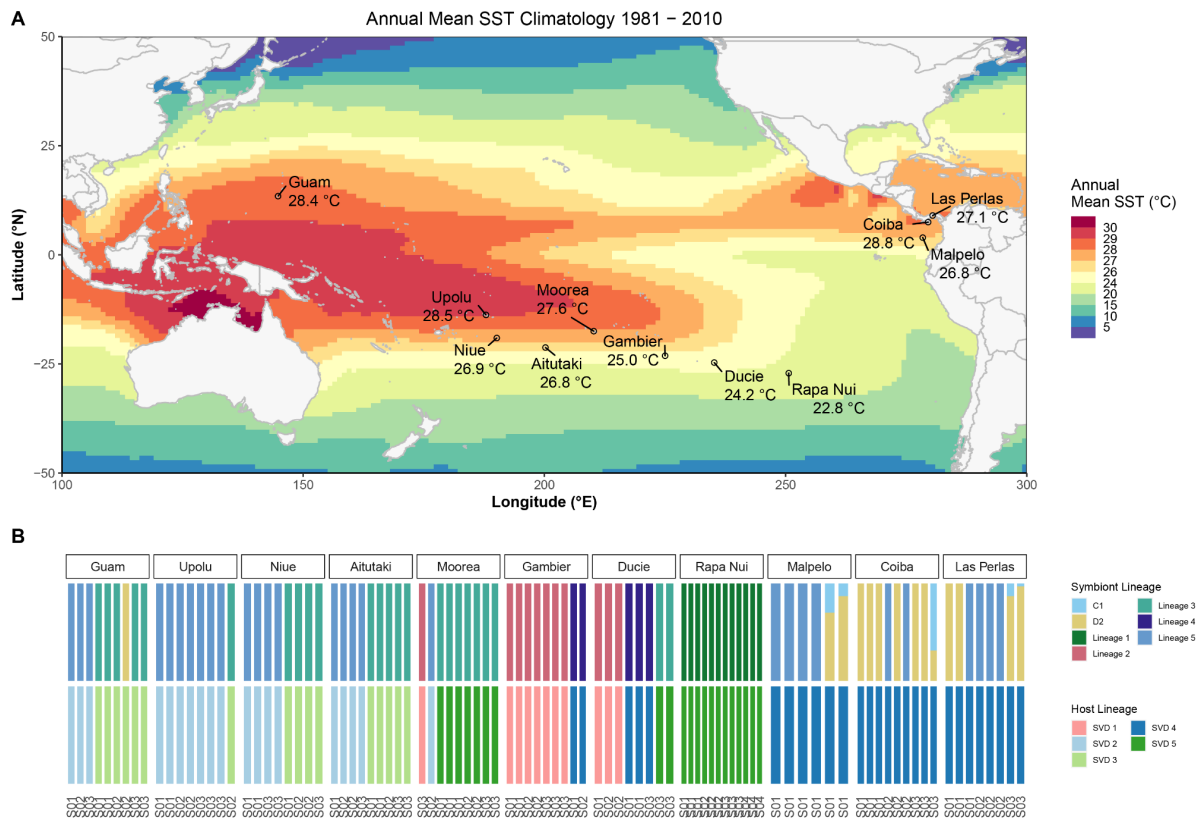


- Kenkel CD, Matz MV (2017) Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nat Ecol Evol* 1:0014. doi: 10.1038/s41559-016-0014
- Kenkel CD, Meyer E, Matz MV (2013) Gene expression under chronic heat stress in populations of the mustard hill coral (*Porites astreoides*) from different thermal environments. *Mol Ecol* 22:4322–4334. doi: 10.1111/mec.12390
- Ladner JT, Barshis DJ, Palumbi SR (2012) Protein evolution in two co-occurring types of *Symbiodinium*: an exploration into the genetic basis of thermal tolerance in *Symbiodinium* clade D. *BMC Evol Biol* 12:217. doi: 10.1186/1471-2148-12-217
- LaJeunesse TC, Reyes-Bonilla H, Warner ME (2007) Spring “bleaching” among *Pocillopora* in the Sea of Cortez, Eastern Pacific. *Coral Reefs* 26:265–270. doi: 10.1007/s00338-006-0189-3
- LaJeunesse TC, Smith R, Walther M, Pinzón J, Pettay DT, McGinley M, Aschaffenburg M, Medina-Rosas P, Cupul-Magaña AL, Pérez AL, Reyes-Bonilla H, Warner ME (2010) Host-symbiont recombination versus natural selection in the response of coral-dinoflagellate symbioses to environmental disturbance. *Proc R Soc B Biol Sci* 277:2925–2934. doi: 10.1098/rspb.2010.0385
- Leggat W, Seneca F, Wasmund K, Ukani L, Yellowlees D, Ainsworth TD (2011) Differential responses of the coral host and their algal symbiont to thermal stress. *PLoS ONE* 6:e26687. doi: 10.1371/journal.pone.0026687
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. doi: 10.1093/bioinformatics/btp324
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Liu H, Hu M, Wang Q, Cheng L, Zhang Z (2018a) Role of Papain-Like Cysteine Proteases in Plant Development. *Front Plant Sci* 9:1717. doi: 10.3389/fpls.2018.01717
- Liu H, Stephens TG, González-Pech RA, Beltran VH, Lapeyre B, Bongaerts P, Cooke I, Aranda M, Bourne DG, Forêt S, Miller DJ, van Oppen MJH, Voolstra CR, Ragan MA, Chan CX (2018b) *Symbiodinium* genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Commun Biol* 1:95. doi: 10.1038/s42003-018-0098-3
- Marlow HQ, Martindale MQ (2007) Embryonic development in two species of scleractinian coral embryos: *Symbiodinium* localization and mode of gastrulation. *Evol Dev* 9:355–367. doi: 10.1111/j.1525-142X.2007.00173.x
- Middlebrook R, Hoegh-Guldberg O, Leggat W (2008) The effect of thermal history on the susceptibility of reef-building corals to thermal stress. *J Exp Biol* 211:1050–1056. doi: 10.1242/jeb.013284
- Middlebrook R, Anthony KRN, Hoegh-Guldberg O, Dove S (2012) Thermal priming affects symbiont photosynthesis but does not alter bleaching susceptibility in *Acropora millepora*. *J Exp Mar Biol Ecol* 432–433:64–72. doi: 10.1016/j.jembe.2012.07.005
- Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong S-Y, Bateman A, Punta M, Attwood TK, Sigrist CJA, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas PD, Finn RD (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* 43:D213–D221. doi: 10.1093/nar/gku1243
- Osman EO, Smith DJ, Ziegler M, Kürten B, Conrad C, El-Haddad KM, Voolstra CR, Suggett DJ (2018) Thermal refugia against coral bleaching throughout the northern Red Sea. *Glob Change Biol* 24:e474–e484. doi: <https://doi.org/10.1111/gcb.13895>
- Pinzón JH, LaJeunesse TC (2011) Species delimitation of common reef corals in the genus *Pocillopora* using nucleotide sequence phylogenies, population genetics and symbiosis

- ecology. *Mol Ecol* 20:311–325. doi: <https://doi.org/10.1111/j.1365-294X.2010.04939.x>
- Planes S, Allemand D, Agostini S, Banaigs B, Boissin E, Boss E, Bourdin G, Bowler C, Douville E, Flores JM, Forcioli D, Furla P, Galand PE, Ghiglione J-F, Gilson E, Lombard F, Moulin C, Pesant S, Poulain J, Reynaud S, Romac S, Sullivan MB, Sunagawa S, Thomas OP, Troublé R, de Vargas C, Vega Thurber R, Voolstra CR, Wincker P, Zoccola D, the Tara Pacific Consortium (2019) The Tara Pacific expedition—A pan-ecosystemic approach of the “-omics” complexity of coral reef holobionts across the Pacific Ocean. *PLOS Biol* 17:e3000483. doi: [10.1371/journal.pbio.3000483](https://doi.org/10.1371/journal.pbio.3000483)
- Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E (2018) Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178. doi: [10.1101/201178](https://doi.org/10.1101/201178)
- Qin Z, Yu K, Chen B, Wang Y, Liang J, Luo W, Xu L, Huang X (2019) Diversity of Symbiodiniaceae in 15 Coral Species From the Southern South China Sea: Potential Relationship With Coral Thermal Adaptability. *Front Microbiol* 10:2343. doi: [10.3389/fmicb.2019.02343](https://doi.org/10.3389/fmicb.2019.02343)
- Robinson JT (2011) Integrative genomics viewer. *Nat Biotechnol* 29:3.
- Rohlf RV, Nielsen R (2015) Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. *Syst Biol* 64:695–708. doi: [10.1093/sysbio/syv042](https://doi.org/10.1093/sysbio/syv042)
- Rosic N, Ling EYS, Chan C-KK, Lee HC, Kaniewska P, Edwards D, Dove S, Hoegh-Guldberg O (2015) Unfolding the secrets of coral–algal symbiosis. *ISME J* 9:844–856. doi: [10.1038/ismej.2014.182](https://doi.org/10.1038/ismej.2014.182)
- Safaie A, Silbiger NJ, McClanahan TR, Pawlak G, Barshis DJ, Hench JL, Rogers JS, Williams GJ, Davis KA (2018) High frequency temperature variability reduces the risk of coral bleaching. *Nat Commun* 9:1671.
- Sanford E, Kelly MW (2011) Local Adaptation in Marine Invertebrates. *Annu Rev Mar Sci* 3:509–535. doi: [10.1146/annurev-marine-120709-142756](https://doi.org/10.1146/annurev-marine-120709-142756)
- Savary R, Barshis DJ, Voolstra CR, Cárdenas A, Evensen NR, Banc-Prandi G, Fine M, Meibom A (2021) Fast and pervasive transcriptomic resilience and acclimation of extremely heat-tolerant coral holobionts from the northern Red Sea. *Proc Natl Acad Sci* 118:e2023298118. doi: [10.1073/pnas.2023298118](https://doi.org/10.1073/pnas.2023298118)
- Schoepf V, Carrion SA, Pfeifer SM, Naugle M, Dugal L, Bruyn J, McCulloch MT (2019) Stress-resistant corals may not acclimatize to ocean warming but maintain heat tolerance under cooler temperatures. *Nat Commun* 10:4031. doi: [10.1038/s41467-019-12065-0](https://doi.org/10.1038/s41467-019-12065-0)
- Schoepf V, Jung MU, McCulloch MT, White NE, Stat M, Thomas L (2020) Thermally Variable, Macrotidal Reef Habitats Promote Rapid Recovery From Mass Coral Bleaching. *Front Mar Sci* 7:245. doi: [10.3389/fmars.2020.00245](https://doi.org/10.3389/fmars.2020.00245)
- Somero GN, Lockwood BL, Tomanek L (2017) *Biochemical Adaptation: Response to Environmental Challenges from Life’s Origins to the Anthropocene*, 9781605355641st edn. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts
- Stat M, Pochon X, Cowie R, Gates R (2009) Specificity in communities of *Symbiodinium* in corals from Johnston Atoll. *Mar Ecol Prog Ser* 386:83–96. doi: [10.3354/meps08080](https://doi.org/10.3354/meps08080)
- Thornhill DJ, Lewis AM, Wham DC, LaJeunesse TC (2014) Host-specialist lineages dominate the adaptive radiation of reef coral endosymbionts: adaptive radiation of symbiotic dinoflagellates. *Evolution* 68:352–367. doi: [10.1111/evo.12270](https://doi.org/10.1111/evo.12270)
- Tresguerres M, Barott KL, Barron ME, Deheyn DD, Kline DI, Linsmayer LB (2017) Cell Biology of Reef-Building Corals: Ion Transport, Acid/Base Regulation, and Energy Metabolism. In: *Acid-Base Balance and Nitrogen Excretion in Invertebrates*. Springer,
- Van der Auwera GA, O’Connor BD (2020) *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra.*, (1st Edition). O’Reilly Media, Inc.

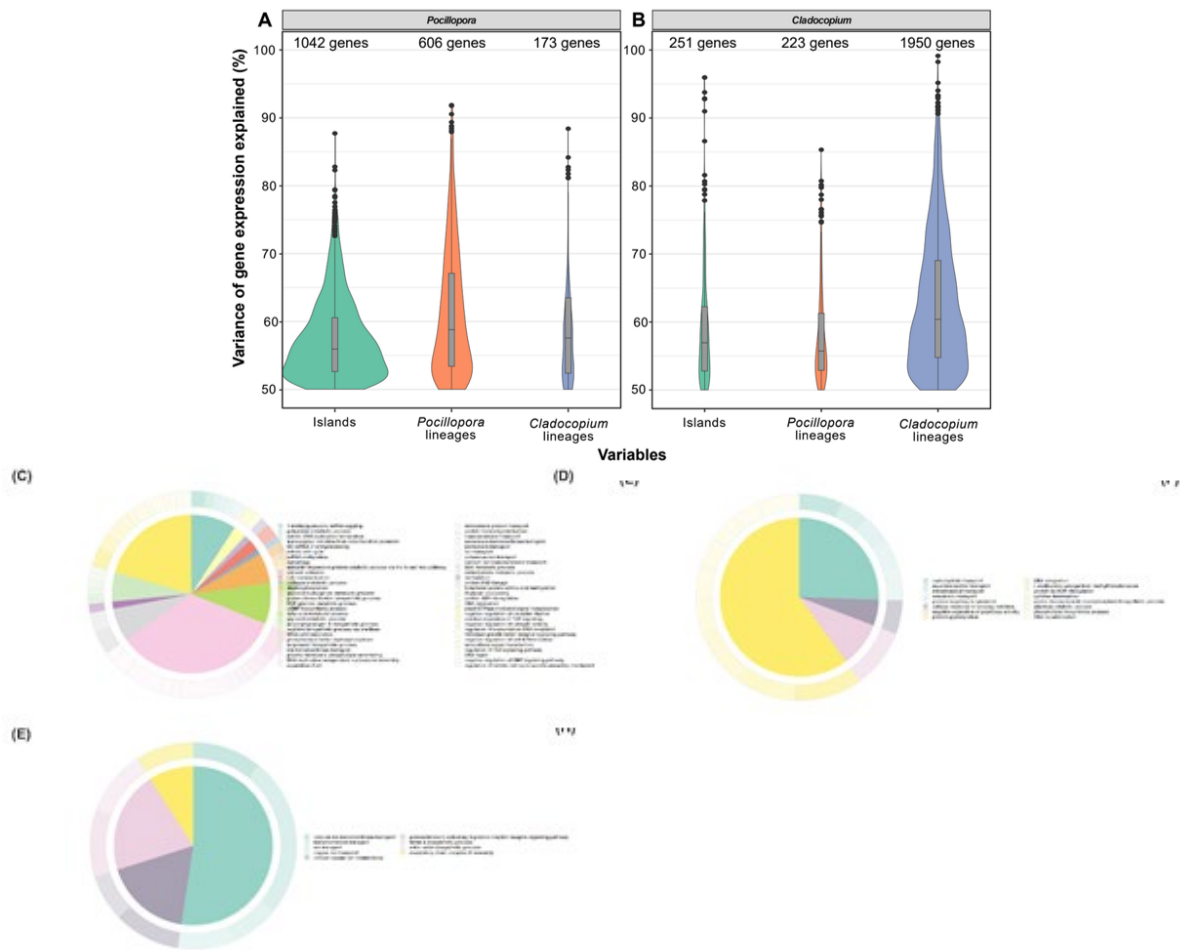
- Wall CB, Ricci CA, Wen AD, Ledbetter BE, Klinger DE, Mydlarz LD, Gates RD, Putnam HM (2021) Shifting baselines: Physiological legacies contribute to the response of reef corals to frequent heatwaves. *Funct Ecol* 35:1366–1378. doi: <https://doi.org/10.1111/1365-2435.13795>
- Wright RM, Aglyamova GV, Meyer E, Matz MV (2015) Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *BMC Genomics* 16:371.
- Wright RM, Mera H, Kenkel CD, Nayfa M, Bay LK, Matz MV (2019) Positive genetic associations among fitness traits support evolvability of a reef-building coral under multiple stressors. *Glob Change Biol* 25:3294–3304. doi: <https://doi.org/10.1111/gcb.14764>
- Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11:R14. doi: <https://doi.org/10.1186/gb-2010-11-2-r14>
- Zoccola D, Tambutte E, Kulhanek E, Puverel S, Scimeca J-C, Allemand D, Tambutte S Molecular cloning and localization of a PMCA P-type calcium ATPase from the coral *Stylophora pistillata*. *Biochim Biophys Acta* 10.

## FIGURES

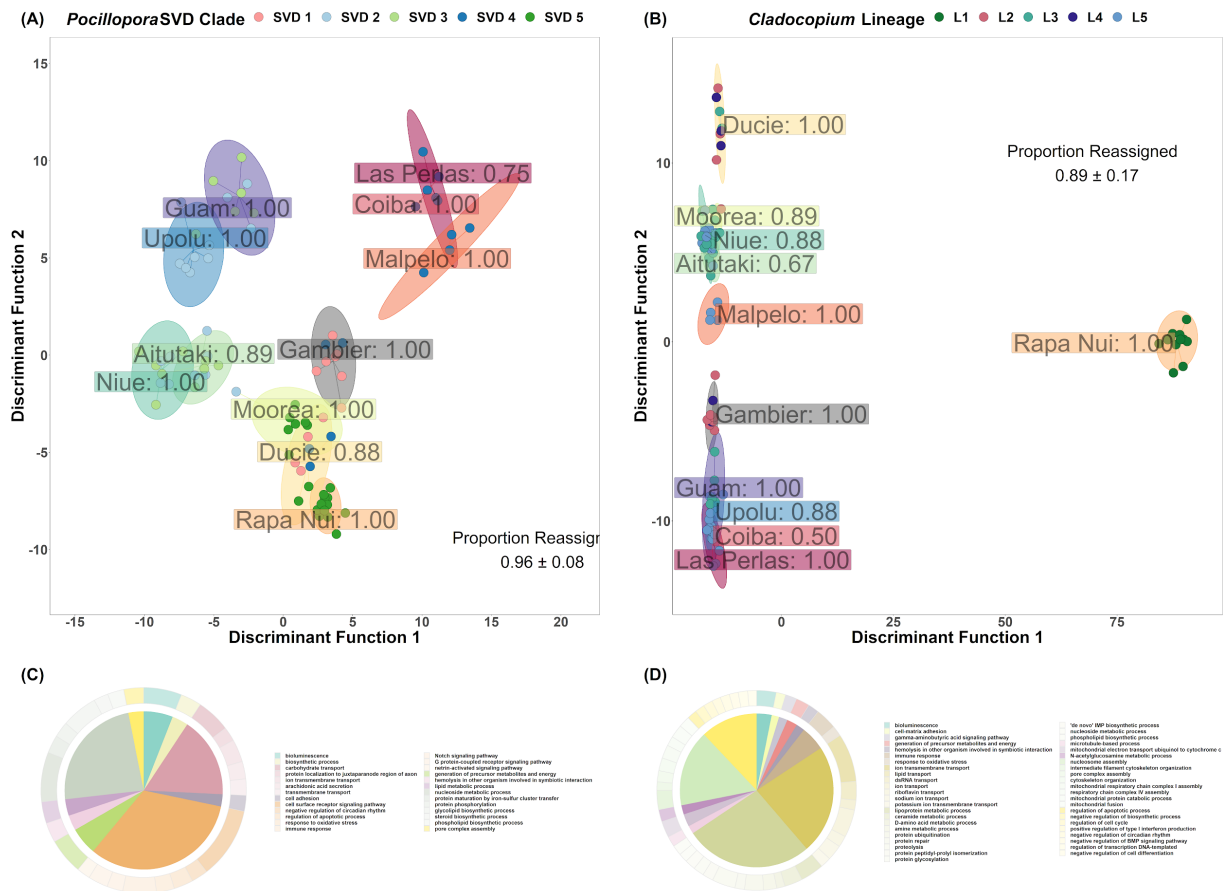


**Figure 1.** *Pocillopora* and Symbiodiniaceae lineages identified in each sample across the Pacific Ocean.

(A) Map showing the 11 islands sampled during the *Tara Pacific* expedition. (B) Lineage assignments of sampled colonies for both the dominant *Cladocopium* C1 genetic lineage (top, including proportions of *Durusdinium* D2 when it was also present) and *Pocillopora* host genotype (middle). Also displayed are historical mean sea surface temperatures (SST) at each sampling site (bottom). Lineages were identified from an analysis of single nucleotide polymorphisms (SNPs) performed using metagenomic reads for the *Pocillopora* host and on metatranscriptomic reads for the Symbiodiniaceae.

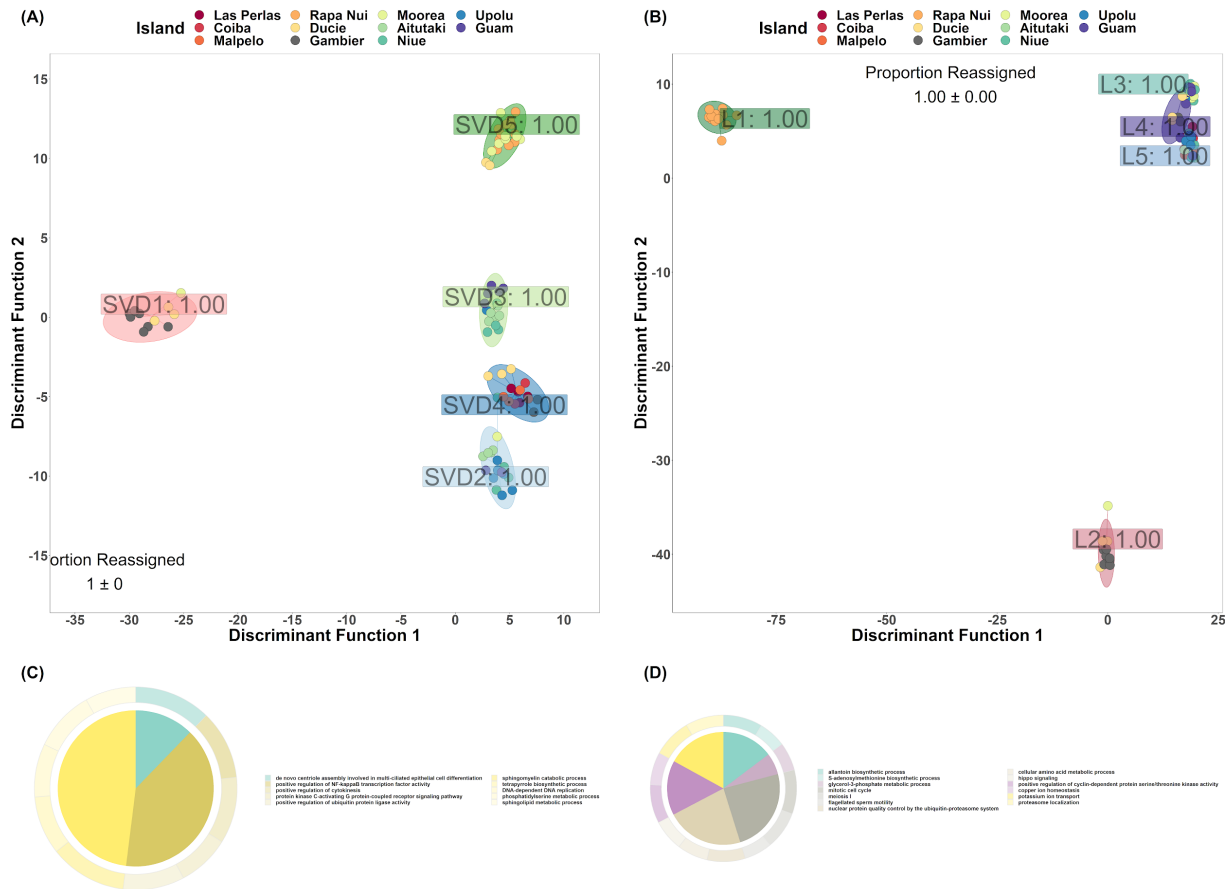


**Figure 2.** Contribution of environmental and genetic variables to *Pocillopora* and *Cladocopium* gene expression and biological process functional enrichments of top variant genes. Distribution of genes in which more than 50% of the variation in their expression is explained by one of three predictor variables: the sampling island, the *Pocillopora* genetic lineage, and the *Cladocopium* lineage. The number of genes in each distribution is indicated above each violin plot. A) Gene expression of *Pocillopora* across 101 samples. B) Gene expression of *Cladocopium* across 70 samples. Samples of Rapa Nui island and those containing *Durusdinium* symbiont were excluded. Host functional enrichments among top variant genes are shown for those best explained by C) island of sampling, D) the host genetic lineage, and E) the photosymbiont genetic lineage.



**Figure 3.** Discriminant analysis of principal components (DAPC) of gene expression data grouped by environment and biological process functional enrichments of discriminant genes in the *Pocillopora* host (left) and *Cladocopium* photosymbiont (right).

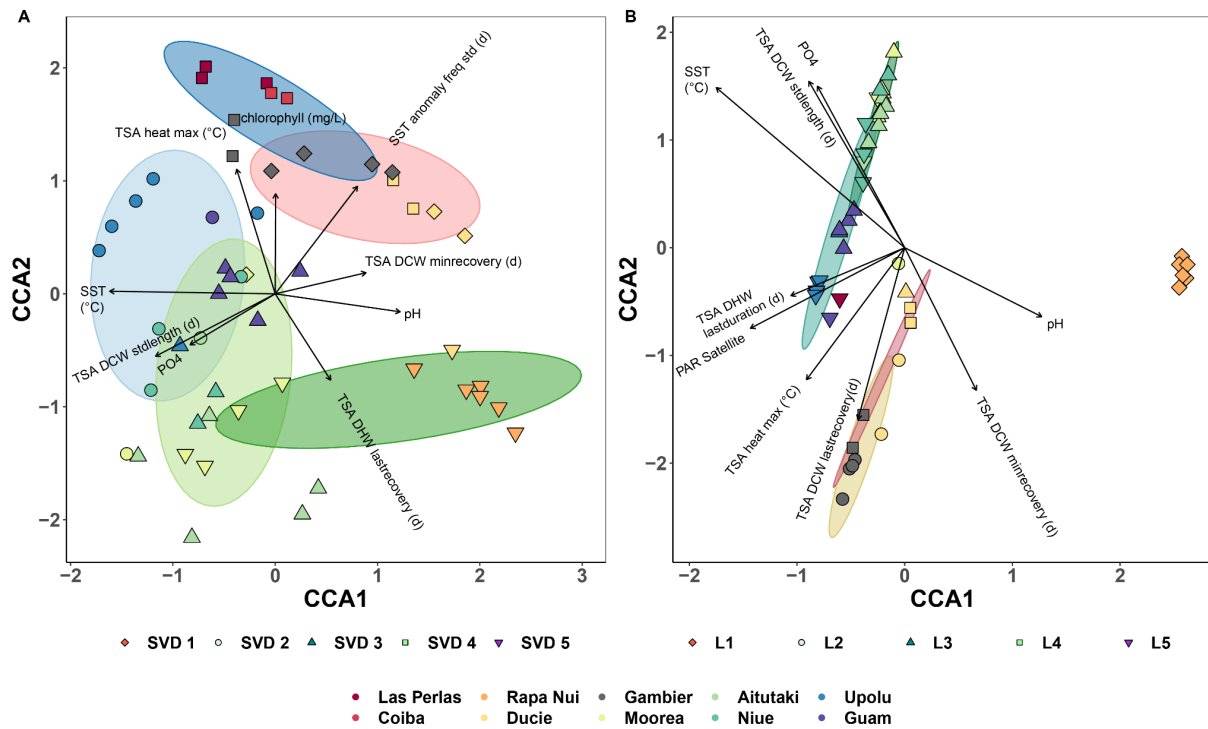
DAPC scatter plots showing expression profiles for the host (A) and algal symbiont (B) when colonies were grouped by sampling island. Points represent individual colony expression profiles and are colored by genetic lineage. Shaded ellipses denote 95%-confidence intervals around the group (island) mean. Group-specific proportions of correct reassignments are indicated within each cluster (boxed labels) and overall model proportions of correct reassignment (mean  $\pm$  standard deviation) are presented within each panel. Also shown are functional enrichments among host genes contributing to the two discriminant functions in the DAPC, DF1 and DF2, (C and D, respectively).



**Figure 4.** Discriminant analysis of principal components (DAPC) of gene expression data grouped by primary genetic lineage and biological process functional enrichments of discriminant genes in the *Pocillopora* host (left) and *Cladocopium* photosymbiont (right). DAPC scatter plots showing expression profiles for the host (A) and algal symbiont (B) when colonies were grouped by primary genetic lineage. Points represent individual colony expression profiles and are colored by island of sampling. Shaded ellipses denote 95%-confidence intervals around the group (lineage) mean. Group-specific proportions of correct reassignments are indicated within each cluster (boxed labels) and overall model proportions of correct reassignment (mean ± standard deviation) are presented within each panel. Also shown are functional enrichments among host genes contributing to the two discriminant functions in the DAPC, DF1 and DF2, (C and D, respectively).

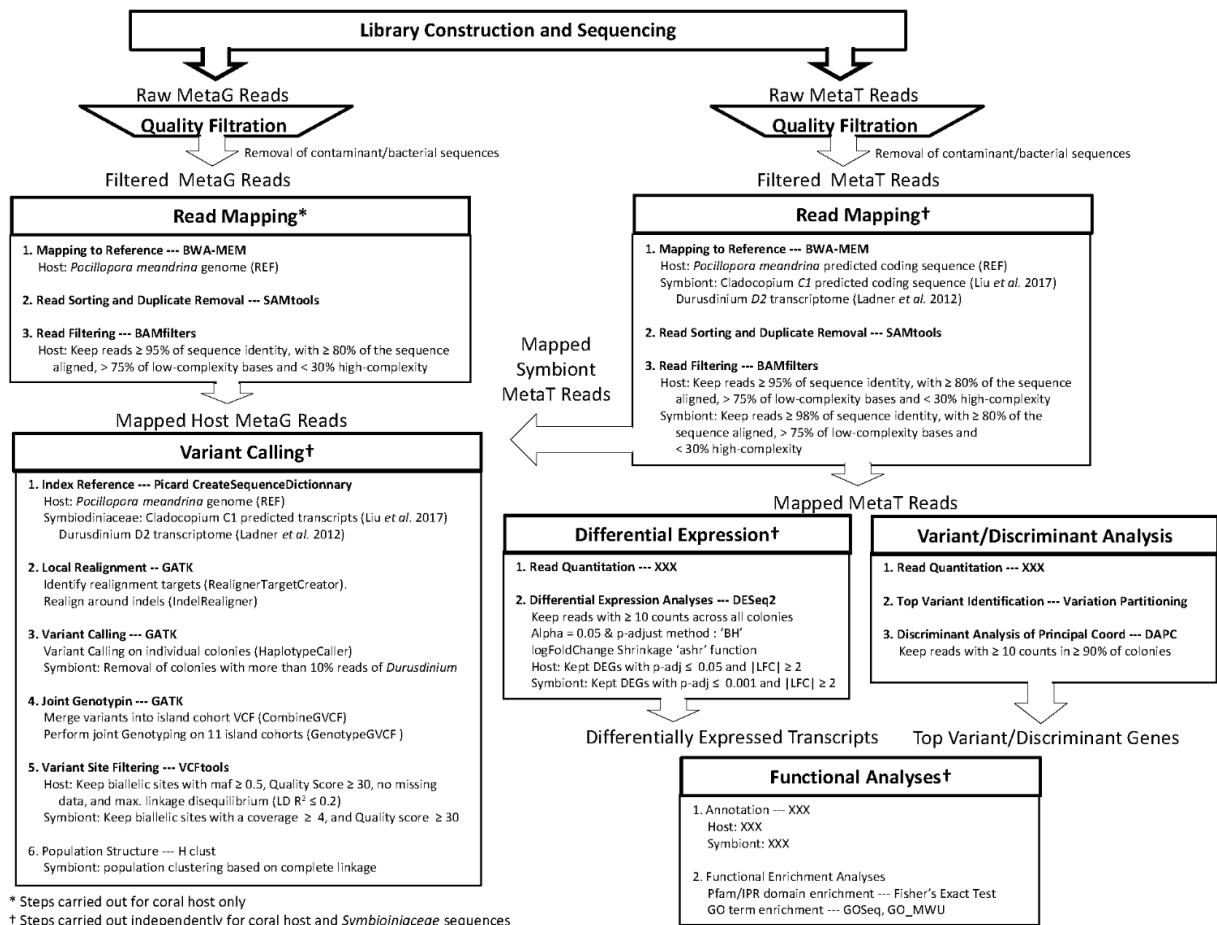






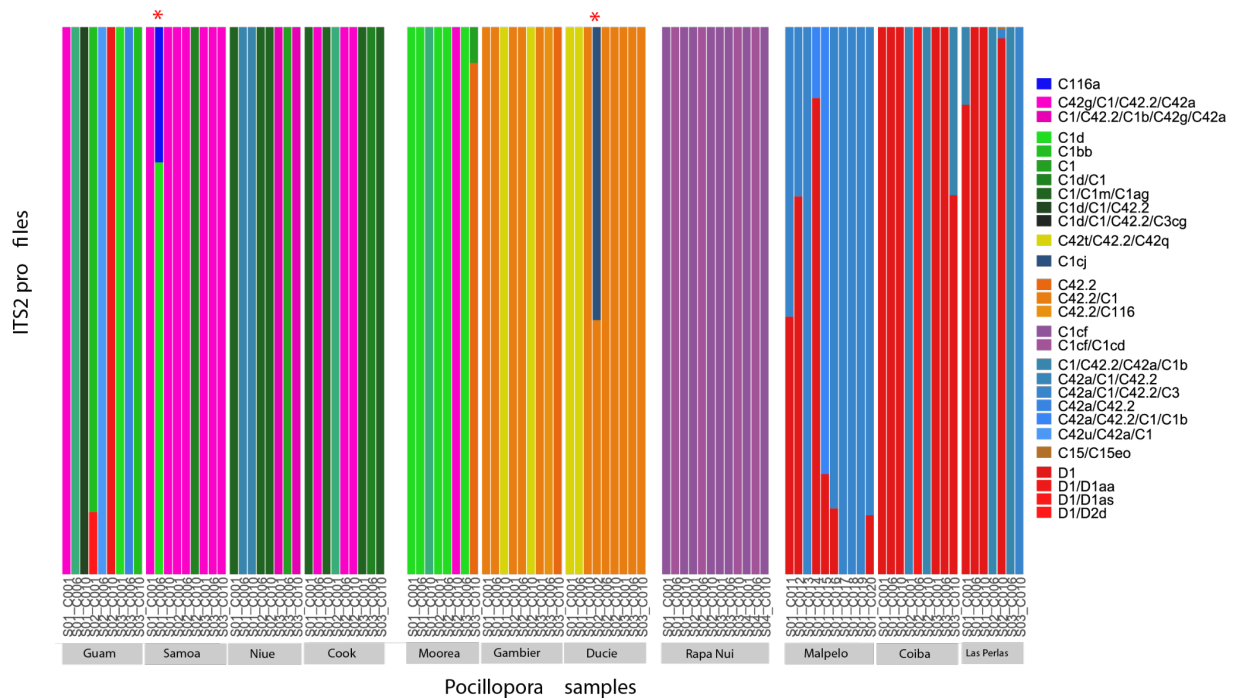
**Figure 6.** Constrained correspondence analysis (CCA) of *Pocillopora* spp. host and *Cladocopium* C1 symbiont gene expression profiles from colonies collected across the Pacific Ocean.

Scatter plots show (A) *Pocillopora* host and (B) *Cladocopium* C1 photosymbiont expression profiles for individual colonies with data colored by island of collection and clustered by genetic lineage (i.e., *Pocillopora* SVD clade and *Cladocopium* C1 genetic lineage, respectively; shaded ellipses). Length (strength) and direction (towards increasing values) of vector arrows denote the relative influence of the top explanatory environmental variables (permuted p-value  $\leq 0.01$ ) for each dataset.



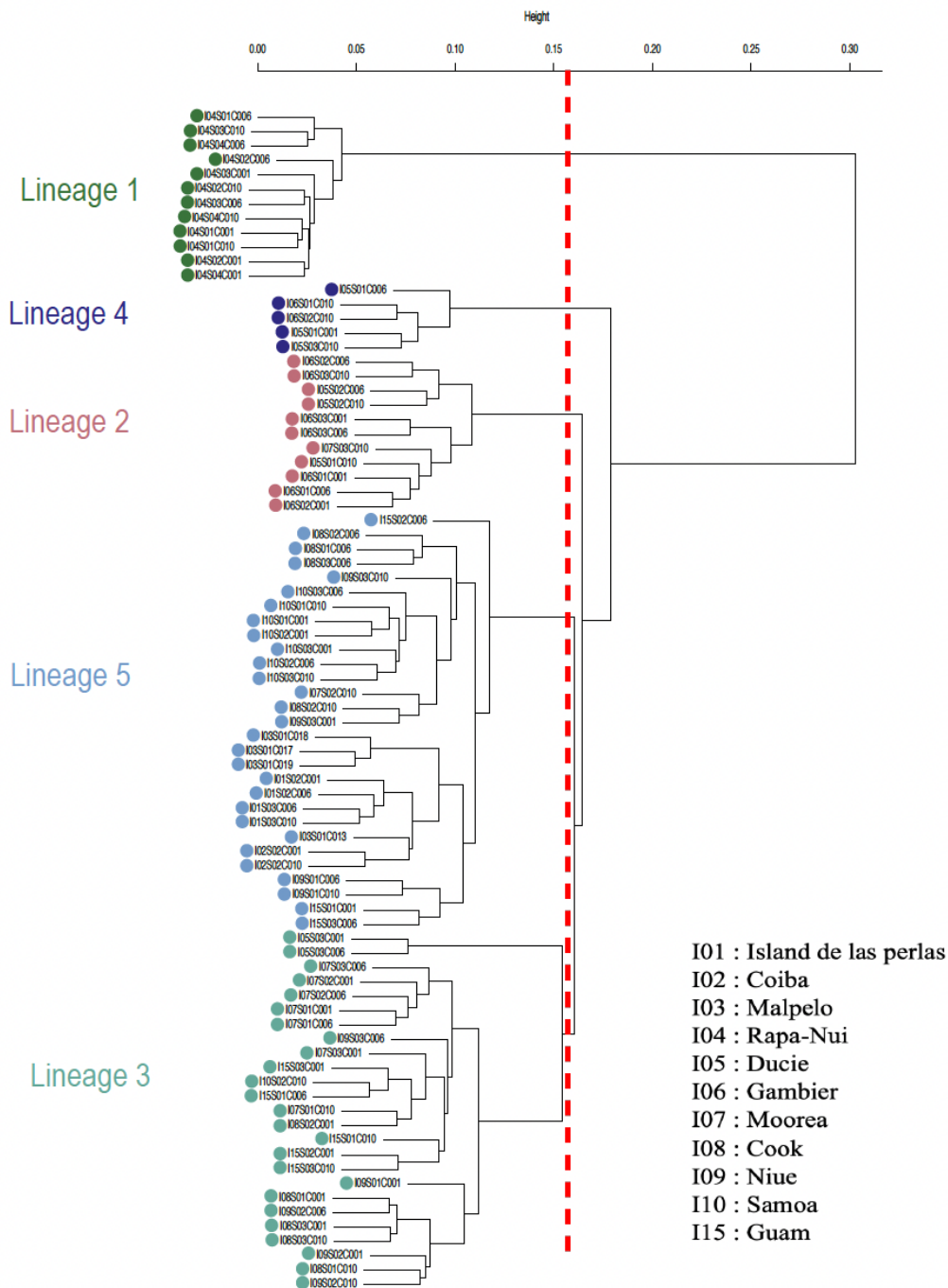
### Supplemental Figure 1. Bioinformatic workflow.

Summary of the bioinformatic workflow used to identify single nucleotide polymorphisms (SNPs) in the host and symbiont as well as to perform functional analyses.



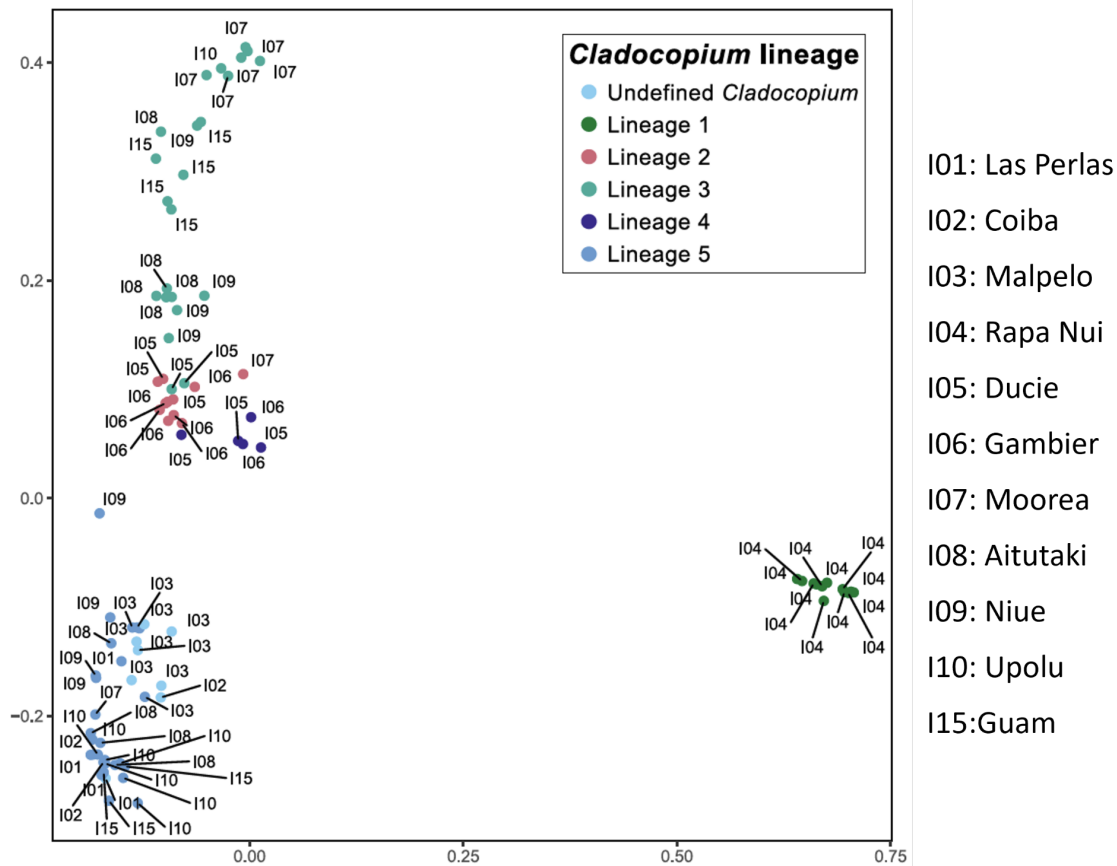
**Supplemental Figure 2.** Relative abundance of Symbiodiniaceae ITS2 profiles in *Pocillopora* colonies.

The abundance of each ITS2 profile was obtained from ITS2 reads analyzed with the SymPortal method (Hume *et al.* 2019). Each color corresponds to a different profile named by its most abundant ITS2 sequence. *Pocillopora* colonies indicated by an asterisk (top) contain 2 different *Cladocopium* ITS2 profiles and were removed from the population structure analysis.



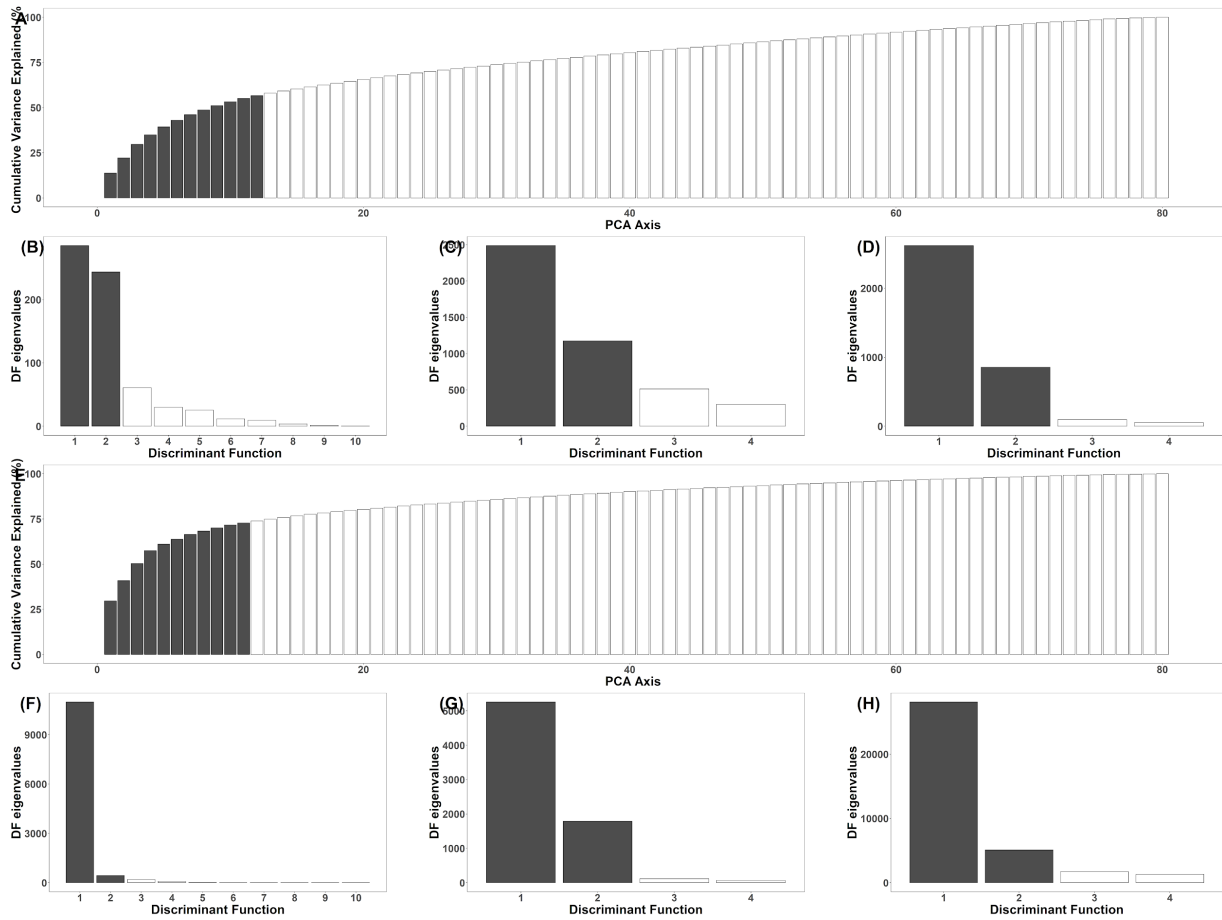
**Supplemental Figure 3.** Hierarchical clustering of *Cladocopium C1* populations in metatranscriptomic samples of *Pocillopora* spp.

Dendrogram representation of the hierarchical clustering of 3,712 SNPs detected by the mapping of metatranscriptomic reads on 1,354 genes of the *Cladocopium C1* reference genome (Liu et al. 2018b). 82 *Cladocopium* samples were clustered into 5 genetic lineages based on a tree height cutoff threshold (red line).



**Supplemental Figure 4:** *Cladocopium* ITS2 profiles distances compared to SNP-based clustering.

ITS2 profiles of *Pocillopora*-containing *Cladocopium* colonies were analyzed using the SymPortal method (Hume *et al.* 2019). Unifrac distances between each pair of samples were calculated and are represented in a principal coordinate analysis (PCA). Each dot represents a *Pocillopora* colony coloured according to the *Cladocopium* C1 lineage determined by the SNP frequencies on the genomic coding sequence.



**Supplemental Figure 5. DAPC Eigenvalues.**

Cumulative percentage of variance explained by the principal coordinate axes (PCA) of the DAPC model for the host (A) and photosymbiont (E). Shaded bars indicate the PCAs retained in each model respectively. Discriminant function eigenvalues for the DAPC models grouped by the environment (B,F), the primary lineage (C,G), and the lineage of the symbiotic partner (D,H) for the host and symbiont, respectively.

## III.4 Discussion and Conclusions

### III.4.1 Convergent expression patterns of Symbiodiniaceae across the Pacific Ocean in three different corals

Two methods were used to visualize the expression levels of the different datasets. These two methods are the PCA visualization and the calculation of Pearson correlations between pair of samples.

The visualization using PCA indicates the percentage of the variance explained for each principal component of the expression patterns. The results showed that for each datasets, the environment explains the different expression pattern, in *Symbiodinium A3*, *Cladocopium C1* and *Cladocopium C15* in *Millepora sp.*, *Pocillopora sp.* and *Porites sp.* respectively.

In addition to this explanation, the genetic of the symbiont is also an essential parameter to explain the gene expression patterns observed. These results for *Cladocopium C1* indicate that Rapa Nui island is very distinct from the other islands due to its different genetic lineages.

From these analyses, it is therefore essential to study the environmental parameters as well as the genetic lineages to explain the gene expression profiles of the Symbiodiniaceae. The results showed that independent of the genetic lineages for *Symbiodinium A3* and *Cladocopium C15*, the gene expression profiles of these samples are very similar when I compared samples between the same island and between sites of the same island. Even in different islands and sites in these datasets, the expression profiles using Pearson's correlations gave pair of samples correlated (Pearson's correlations higher than 0.7 between each pair). These results are similar for samples of *Cladocopium C1* in *Pocillopora sp.* samples, except for those containing two different genera (*Cladocopium C1* and *Durusdinium D2*).

These analyses also showed that samples from the same islands are correlated whereas samples from different island are weaker correlated. Moreover the correlations intra-sites were equivalent with the correlations inter-sites.

From these results it was deduced, that samples from the same site are enough closed to be used as biological replicates. It also appears that samples from different islands can't be used as biological replicates. It reflects the importance of the geographical location in the gene expression of Symbiodiniaceae.

Thus, with these 188 samples it therefore appeared essential to take in account the envi-

ronmental parameters and the genetic parameters to explain gene expression level of Symbiodiniaceae in the Pacific Ocean.

### III.4.2 *In situ* transcriptomic adaptation of Symbiodiniaceae at island scale

The differential expression analyses using DESeq2 pipeline revealed that only a few genes are differentially expressed between samples from the same island. It was a result expected because most of genes are expressed at the same level between samples of the same island. From these results, it appeared that Symbiodiniaceae should adapted to different environmental conditions at transcriptomic level in the different islands and so there is an *in situ* transcriptomic adaptation of Symbiodiniaceae between two different environmental conditions from the same island.

These results also showed that differentially expressed genes could not only be explained by the island parameter. Indeed, it appeared that some islands have very different patterns between the comparisons of each pairwise of islands. Given the high rate of differentially expressed genes, it appeared that the use of genetic to explain the adaptation of Symbiodiniaceae in the Pacific Ocean is essential.

The lack of environmental parameters to perform comparative analyses with the differentially expressed genes did not allow to go further in these analyses. In fact, the CCA obtained for each set of data did not made it possible to conclude on the profiles of these genes differentially expressed under each of the different conditions.

The analyses of differential expression patterns, taking in account only the genetic of the *Cladocopium C1* symbiont, allowed to distinguish certain patterns. Indeed, it appeared a distinction of lineage 1 compared to the others. These results made it possible to confirm the genetic distance of this lineage compared to the others and its importance to understand the genetic adaptation of Symbiodiniaceae across the Pacific Ocean.

However, these results were very strongly influenced by the number samples in each condition. In fact, lineages 2 and 4 have few differentially expressed genes because they have few replicates compared to the others.

It is therefore essential to better understand the adaptation of *Cladocopium C1* to take in account the genetic parameters and environmental parameters, avoiding biases of the bio-



informatic methods used.

### III.4.3 Gene functions characterization of *Symbiodinium A3* in different environmental conditions

The functional analyses of the differentially expressed genes of *Symbiodinium A3* in *Millepora* showed only a partial view of their functions. In this context, a new annotation of the *Symbiodinium A3* genome was necessary to better understand the functions associated with these differentially expressed genes.

These additional analyses allowed to take a new look of the functions associated with *Symbiodinium A3*. The results showed that the metabolic pathways associated with this genome are classical metabolic pathways found in Symbiodiniaceae such as photosynthesis, nutrient transport, the lipid synthesis and sugar pathway, etc.

Using these results, it was therefore possible to carry out a study of the functions associated with the differentially expressed genes of *Symbiodinium A3* in the different islands sequenced. Several gene functions associated with photosynthesis, nutrient transport, fatty acids and sugar pathways are differentially expressed between islands, which are classical metabolic pathways involved in *Symbiodinium A3*.

Stress-related metabolic pathways are differentially expressed between Moorea , Aitutaki and Niue islands. This result suggests that genes involved in the regulation of the oxidative stress are important for the environmental adaptation of *Symbiodinium A3* in these different conditions.

The GO terms analysis of Symbiodiniaceae genes was not conclusive. Indeed, GO terms are quite diverse and varied, which made global analyses complex. It is therefore necessary to perform filters on the datasets, which sometimes lead to the deletion of essential informations.

The use of metabolic pathways with *a priori* gave results with a bias due to the selection of the chosen metabolic pathways. Indeed, the selection of metabolic pathways for photosynthesis and nutrient transport gave only a partial view of the datasets.

There was a transcriptional regulation of nutrient transport and also photosynthesis which are essential metabolic pathways in Symbiodiniaceae to adapt in different environmental conditions.

The results on a global scale did not allow to define with confidence the functions linked to the adaptation of *Symbiodinium A3* in the Pacific Ocean. This is probably because several lineages of *Symbiodinium A3* are mixed in these analysis as shown for *Cladocodium C1* (Armstrong, Lê-Hoang *et al.*).

The population structure analysis of *Symbiodinium A3* was not carried out because the reference genome sequence was not available. It therefore appears necessary to continue these analyses taking in account the genetic lineages of *Symbiodinium A3* as well as the various environmental parameters and the genetic of the host in order to better understand the adaptation of this Symbiodiniaceae in the Pacific Ocean.

#### **III.4.4 Expression and associated functions for the adaptation of *Cladocodium C1* in *Pocillopora sp.***

In order to understand the adaptation of *Cladocodium C1* in the Pacific Ocean, I did differential expression analyses taking in account environmental parameters and distinct genetic lineages.

To compare the results on the different conditions, an analysis of the global genome of *Cladocodium C1* was carried out. This analysis was also compared by taking in account the different lineages of *Cladocodium C1* obtained. The results of the KEGG associated in these comparisons showed an overall enrichment of the genome for the different metabolic pathways of Symbiodiniaceae, independently of the genetic lineages. These results therefore indicated that the functions associated with the genes of the different lineages of *Cladocodium C1* are identical.

Analysis of genes differentially expressed for *Cladocodium C1* independently of genetic lineages were not conclusive. Indeed, a diversity of functions related to the GO terms which are diverse and varied on the global scale of these analyses were obtained.

Due to the richness of the comparisons between each pair of islands, it was difficult to identify functional profiles on a global scale. Manual study of these functions gave similar results. Moreover, the use of a manual search with *a priori* did not allow to obtain conclusive results on the functional diversity between the different islands.

Patterns emerged from the fact that lineage 1 is present exclusively on Rapa Nui island

unlike the others. It therefore appeared essential to use the genetic of the symbiont to carry out these functional analyses.

The use of different bioinformatics methods on GO, KEGG or PFAM gave similar results. These results showed that the use of different databases is essential to understand the functions associated with the DEG. However, there were bias and inconclusive results due to the study on a global scale. It appeared again essential to make a specific analysis of the DEG in *Cladocopium C1* to better understand its adaptation in different conditions.

The analyses of the functions associated with the DEG of the genetic lineages of *Cladocopium C1* made it possible to obtain metabolic pathways between different lineages. However, due to the fact that only one lineage is present for a single island (Rapa Nui, lineage 1), the results obtained were biased in the enrichment of the associated functions. Indeed, for the other lineages, the environmental parameters were not taken in account. It appeared essential to use them to understand the environmental adaptation of *Cladocopium C1* in the Pacific Ocean.

In addition, this method is biased by the number of samples present in each lineage. In fact, lineages 2 and 4 obtained in the **Chapter 2**, present few samples unlike the others, which indicates biased expression profiles due to this factor.

### III.4.5 Collaborations

The results obtained made it possible to create collaborations with laboratories from the *Tara* consortium.

A first collaboration concerned the bioinformatic protocols applied for the study of the data of the symbionts in the various corals. Indeed, the protocols that I developed, made it possible to extract the data of the symbiont from the dual-transcriptomic and dual-genomic samples in the different datasets. These results were used by Corentin Hochard in his thesis, to filter the coral samples and thus obtain samples with a minimum of Symbiodiniaceae contamination.

A second collaboration made it possible to look at the correlations between the telomeres of the corals and the symbiont. The results of alignments and filtrations for each coral sample, made it possible to create correlation matrices for the genes of Symbiodiniaceae and then, made the link with evolution of coral telomeres. These results were studied by Alice Rouan in

her thesis, to better understand the production and evolution of telomeres in coral organisms.

A final collaboration allowed to look at an overview of *Cladocopium C1* in a metabolomic point of view. The study of the population of *Cladocopium C1* and its biomass in each sample of *Pocillopora sp.* made it possible to carry out a study on the metabolomic of the coral *Pocillopora sp.* in the samples from the *Tara Pacific* expedition. These analyses were done by Mélanie Pousse in her thesis.



## Chapter IV

# Adaptation of Symbiodiniaceae in surrounding water across the Pacific Ocean

It has been shown that the coral symbionts are able to recolonize a coral tissue after a bleaching event. However, the capacity of Symbiodiniaceae to survive in the seawater is still poorly understood. The aim of this chapter was to answer the questions: Is there the same Symbiodiniaceae diversity within and outside the coral colony ? How does the symbiotic micro-algae adapt between the coral host tissues and the seawater environments ?

To answer to these questions, I studied the Symbiodiniaceae diversity using the bioinformatics methods applied in the **Chapter 2**. I performed a population structure of *Cladocopium C1* and then, I compared the results between the seawater and the dual-transcriptomic samples.

In addition to these studies, I analysed the potential contamination of the surrounding water samples by the coral host, to know if there is coral polyps in the seawater samples.

In a second part, I explored the number of differentially expressed genes between the two different environments to understand if the Symbiodiniaceae are strongly impacted or not by their expulsion of the coral host. The capacity of Symbiodiniaceae to survive in the seawater is crucial under recurrent coral bleaching event, and these analyses made it possible to obtain a list of DEG. These ocean-scale analyses had never been done before in the literature and they represent a major asset to understand Symbiodiniaceae adaptation in different environments

in the context of coral bleaching.

## IV.1 Materials and Methods

### IV.1.1 Coral surrounding water sampling to RNA sequencing

#### IV.1.1.1 Surrounding water sampling

During the *Tara* Pacific expedition, researchers sampled seawater surrounding *Pocillopora* sp. colonies. From these samples, I was interested to the symbiotic micro-algae living around coral colonies.

The sampling plan for Coral Surrounding Water (CSW) samples was the following: for each sampling site, 2 CSW were collected around 2 *Pocillopora* sp. colonies and one surface sample was collected (**Figure IV.1**).

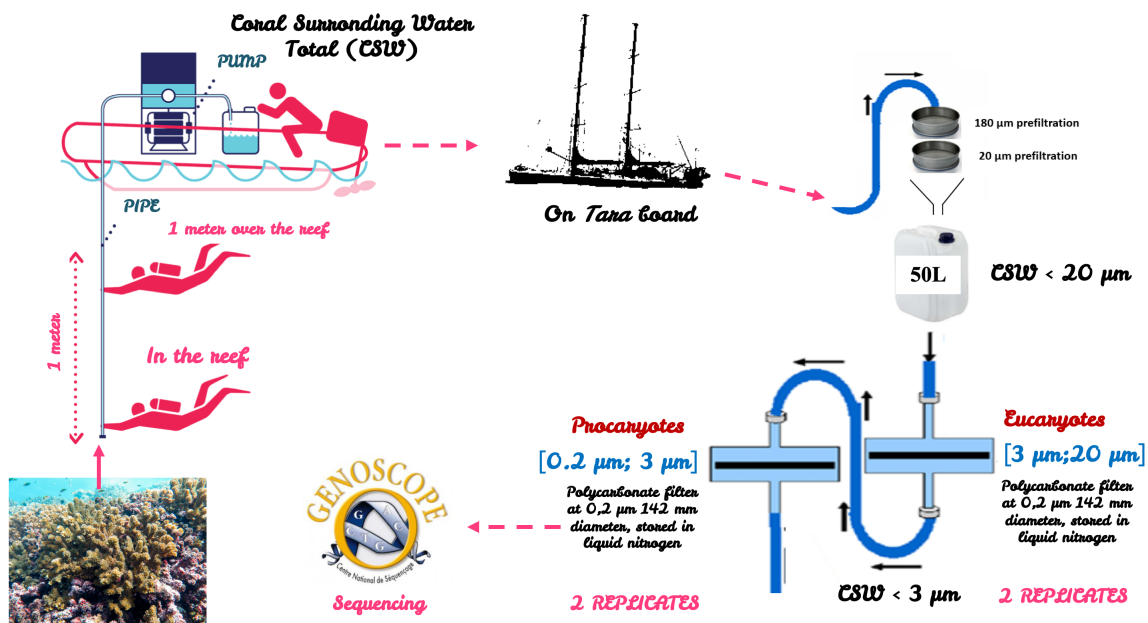


Figure IV.1: The *Tara* Pacific sampling plan for coral surrounding water samples in each sampling site.

50 liters of CSW were successively filtered with two nets of 180 μm and 20 μm meshes. Then, eucaryotes were retained on a 3 μm filter and procaryotes on a 0.2 μm filter and after, samples filtered were sent at the Genoscope sequencing platform.

Coral Surrounding Water samples were collected and sent to the Genoscope sequencing platform. All the CSW samples of all the *Tara* Pacific expedition were stored and sequenced at the Genoscope.

Samples were placed in 15mL Lysing Matrix A tube (reference: 116930050 by MP Biomed-





All the samples sequenced and used in this part are presented per island, site and environment (**Table IV.1**<sup>1</sup>). The number of available samples differ slightly from initial sequencing plan in some islands and sites where CSW samples were not sequenced. Surface samples of Las Perlas, Coïba, Gambier, Samoa and Guam islands were the only available samples, during my thesis.

#### **IV.1.2 bioinformatics methods to align transcriptomic reads on genome and transcriptome references**

To obtain Symbiodiniaceae diversity in CSW and SRF<sup>2</sup> samples, the same protocols that I developed before (**Chapter 2**) were applied on the two datasets, using Symbiodiniaceae genome and transcriptome references. For each dataset, I filtered samples as following:

- **CSW:**

%A percentage of identity over 70% and a coverage over 30 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept using all reference transcriptomes (**Table II.1**). A percentage of identity over 95% and a coverage over 50 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept using all reference transcriptomes (**Table II.1**). A percentage of identity over 98% and a coverage over 80 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept using reference CDS of *Cladocopium C1* (**Table II.2**) and reference transcriptome of *Durusdinium D2* (**Table II.1**).

**2: Surface samples:** A percentage of identity over 95% and a coverage over 50 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept using all reference transcriptomes (**Table II.1**).

To know if the water samples around the coral colonies were contaminated with coral polyps, I aligned CSW reads on *Pocillopora sp.* reference genome.

Indeed, because the surrounding water is close to the coral colonies, it is necessary for the analyses to know if Symbiodiniaceae hostpites is partially found in the water samples.

---

<sup>1</sup>DualT: Dual-Transcriptomic

<sup>2</sup>SuRface Samples

To do this, I used the classic alignment protocols that I developed previously (**Chapter 2**), using as reference the genome of *Pocillopora meandrina* sequenced and assembled at the Genoscope by the Production and the LBGB teams.

The filters used on this alignment were: A percentage of identity over 95% and a coverage over 50 %. Under 75% of low complexity reads were removed and over 30% high complexity reads were kept on CSW reads aligned against *Pocillopora* genome.

### **IV.1.3 Population structure of *Cladocopium C1* in coral surrounding water**

The same protocols as described in the **Chapter 2** were performed on transcriptomic CSW samples, to obtain a population structure of *Cladocopium C1* in these samples.

### **IV.1.4 bioinformatics methods to determine the genetic and the environmental adaptation of Symbiodiniaceae, in the seawater compared to the coral host**

In this part, I was interested to the differentially expressed genes of the micro-algae in the different samples from the *Tara* Pacific expedition (CSW versus DualT).

Thanks to the results obtained previously, on the diversity of symbionts and with the diversity of genomic and transcriptomic references in Symbiodiniaceae available, differential expression studies were performed on the two kind of samples using as describe in the **Chapter 3** the same protocols such as:

- Expression profiles of the *Tara* transcriptomic datasets: PCA and Pearson correlations.
- Determination of biological replicates: I applied the same analyses to determine environmental biological replicates on the datasets than in the **Chapter 3 - Figure III.1**. Then, I performed the same analyses to understand the genetic adaptation of *Cladocopium C1* seawater and coral on the datasets using the genetic population structure of *Cladocopium C1* (**Chapter 3**).
- After determining different conditions, I performed the classical protocol of differential expression analyses as presented before on the datasets (**Chapter 3**). %

- Following the results obtained, the functions associated with differentially expressed genes were analysed to understand the metabolic pathways involved in the genetic and environmental adaptation of *Cladocopium C1* in different environment *in situ* and *in vivo*, across the Pacific Ocean. To do this, I applied bioinformatics methods developed before on the different datasets (**Chapter 3**). Several databases such as Gene Ontology (GO), Protein families (PFAM) or Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were used to interpret the gene functions of *Cladocopium C1* genome.

## IV.2 Results

### IV.2.1 Symbiodiniaceae diversity in different environmental samples across the Pacific Ocean

In this chapter, I intended to compare the transcriptomic activity of Symbiodiniaceae within the coral host versus its activity in the surrounding water. The mandatory preliminary step was to understand if the Symbiodiniaceae genetic lineages were the same inside and outside the host.

#### IV.2.1.1 Symbiodiniaceae diversity in coral surrounding water using Symbiodiniaceae reference transcriptomes

To identify Symbiodiniaceae species present in CSW samples from the first 11 islands, I aligned reads simultaneously on all Symbiodiniaceae transcriptomes (**Figure IV.2**).

In the East Pacific (Las Perlas and Coïba islands and one sample in Guam), reads were aligned on two Symbiodiniaceae genera (*Cladocopium* and *Durusdinium*). In the Central and West Pacific, the reads are only aligned on *Cladocopium* genus. These results indicate that at least two species are present in these samples. From these results, *Cladocopium C1* reference genome and also *Durusdinium D2* reference transcriptome were used for the next analyses.

After having determined the species present in each CSW sample and used genomic and transcriptomic references available, I performed a specific alignment using the closest available genome or transcriptome available with stricter alignment parameters (minimum of 98% of identity).

Again for the CSW samples, I aligned simultaneously *Cladocopium C1* CDS and *Durusdinium-D2* transcripts to avoid non-specific read aligned (**Figure IV.3.A**). Thus, samples from Las Perlas, Coïba and 1 sample from Guam contain *Durusdinium D2* in CSW samples (**Figure IV.3.A**). In all other samples, the Symbiodiniaceae population is *Cladocopium C1*. Then, these results were compared to the Symbiodiniaceae present in coral colonies (**Figure IV.3.B**). These results showed the same Symbiodiniaceae diversity between the two different samples for the two different environments. It means that Symbiodiniaceae living inside and outside of the coral host are identical based on these transcriptomic analyses.

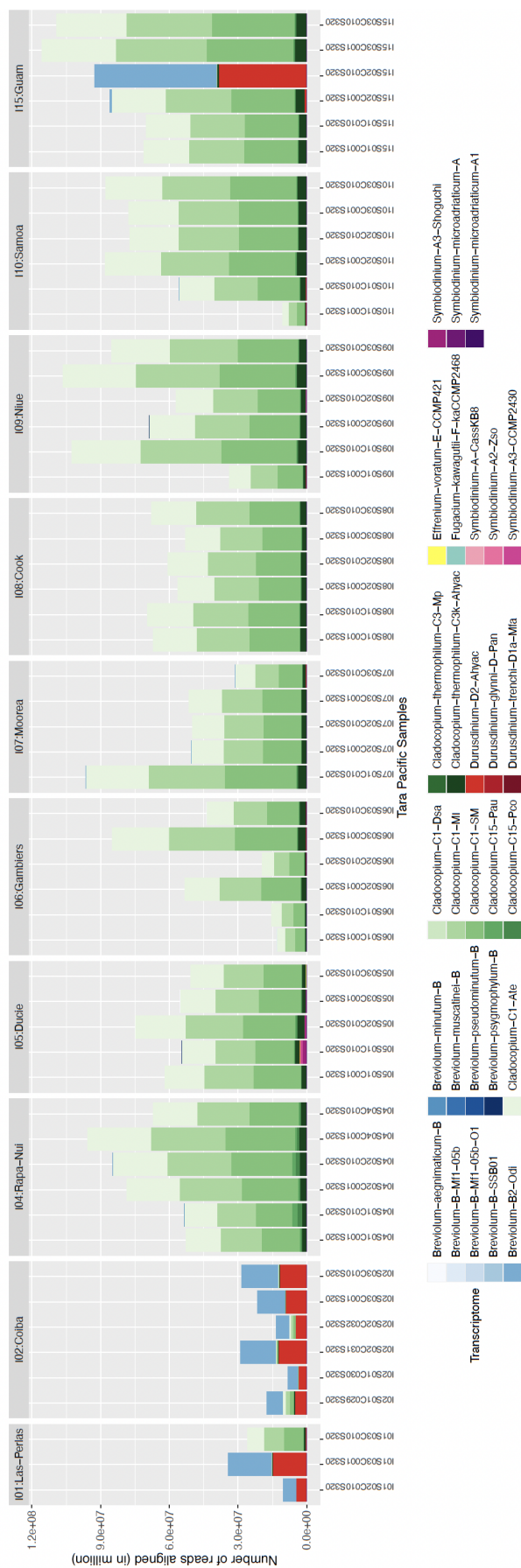


Figure IV.2: Barplots of the number of CSW reads aligned on Symbiodiniaceae reference transcriptomes.

Each color corresponds to a specific transcriptome. Similar colors are transcriptomes of the same Symbiodiniaceae genus.

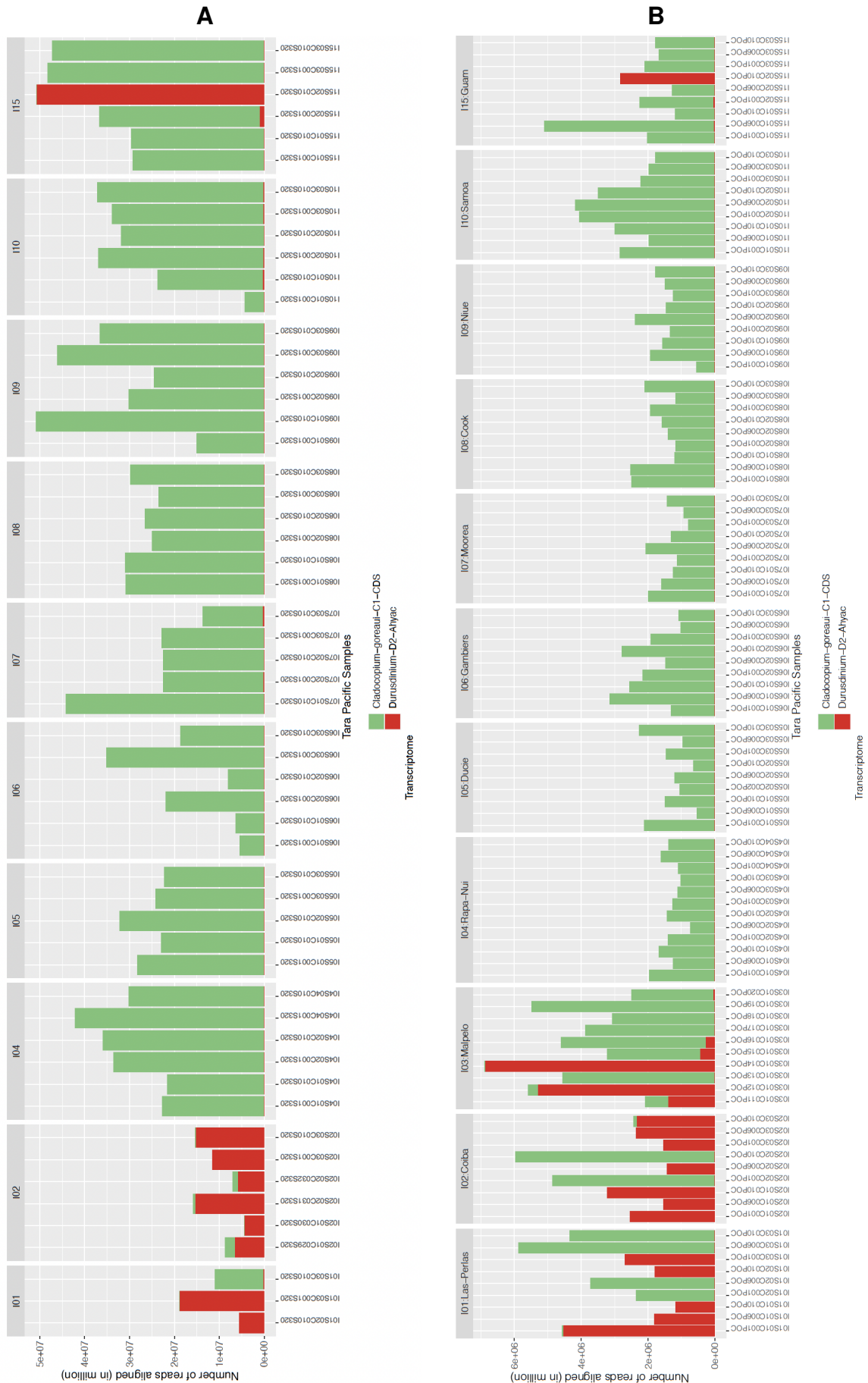


Figure IV.3: Barplots of the number of CSW (A) and dual-transcriptomic (B) reads aligned on specific Symbiodiniaceae references.

%Each color corresponds to a specific transcriptome or genome.

### IV.2.1.2 Potential coral contamination in coral surrounding water samples

To be sure that the corresponding Symbiodiniaceae between coral colonies and CSW samples are not due to Symbiodiniaceae living in coral polyp detached from the coral host, I studied the coral contamination on CSW samples. To this, CSW reads were aligned against *Pocillopora meandrina* reference genome (**Figure IV.4**).

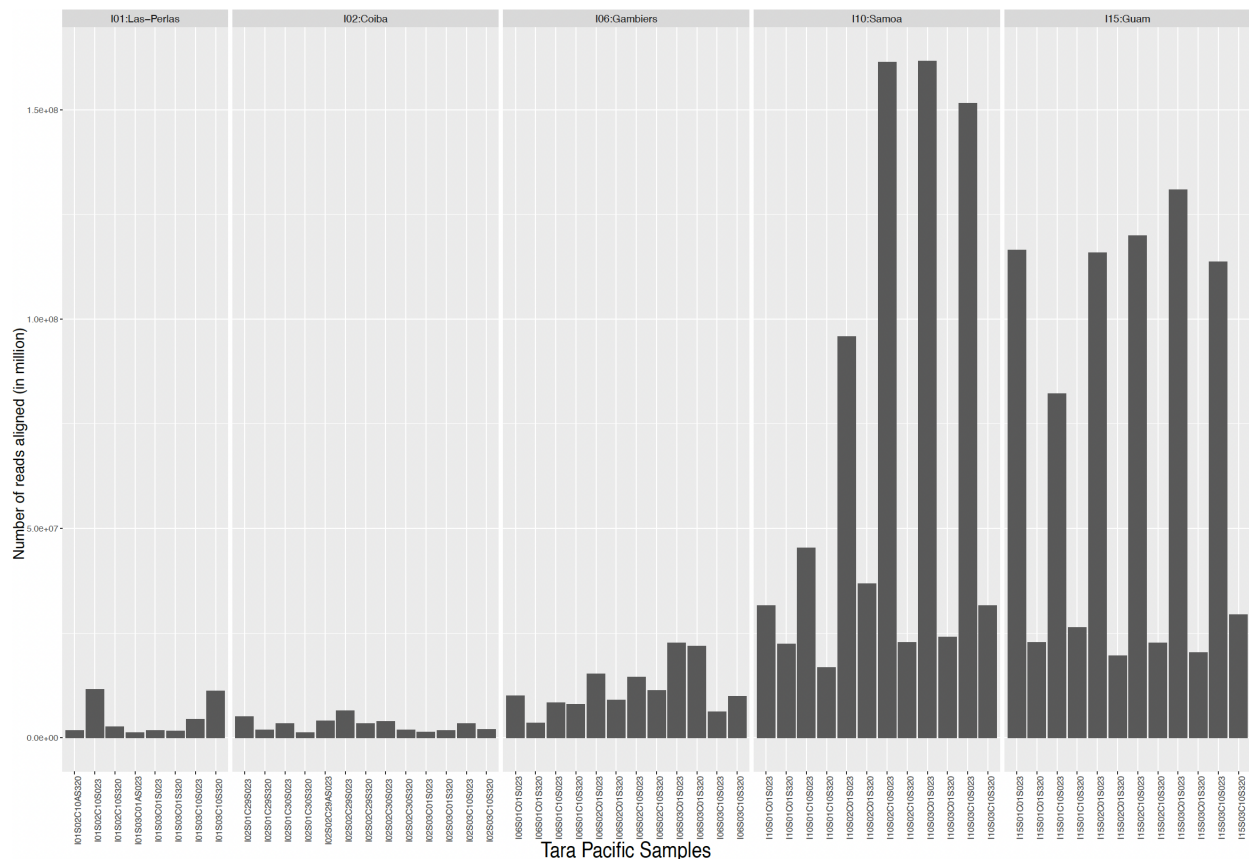


Figure IV.4: Barplots of the number of CSW reads aligned on *Pocillopora meandrina* reference genome.

About 0.1% of CSW reads are coral reads, because 20% of the coral sample reads belong to the symbiont. I estimated that only 0.02% of all reads are contaminant Symbiodiniaceae hospite. This number is negligible compared to the percentages of the CSW reads samples assigned to Symbiodiniaceae.



### **IV.2.1.3 Symbiodiniaceae diversity in the *Tara* Pacific surface samples using Symbiodiniaceae references**

In this analysis, I was interested in the diversity of Symbiodiniaceae living in surface water samples. Indeed, I was wondered if the Symbiodiniaceae living further from the coral colonies are the same as those living nearby.

To do this, I aligned the reads of the surface water samples against the Symbiodiniaceae reference transcriptomes according to the same protocol as in the previous analyses (**Chapter 2**).

The results showed a large diversity of Symbiodiniaceae associated to the 6 genera available based on these transcriptomic analyses (**Figure IV.5**). In fact, there is the 6 different clades available present in each surface sample with different proportions.

These results indicated, that the diversity of Symbiodiniaceae on coral reef surface is very large and should be essential to coral and Symbiodiniaceae adaptation in different environmental conditions.

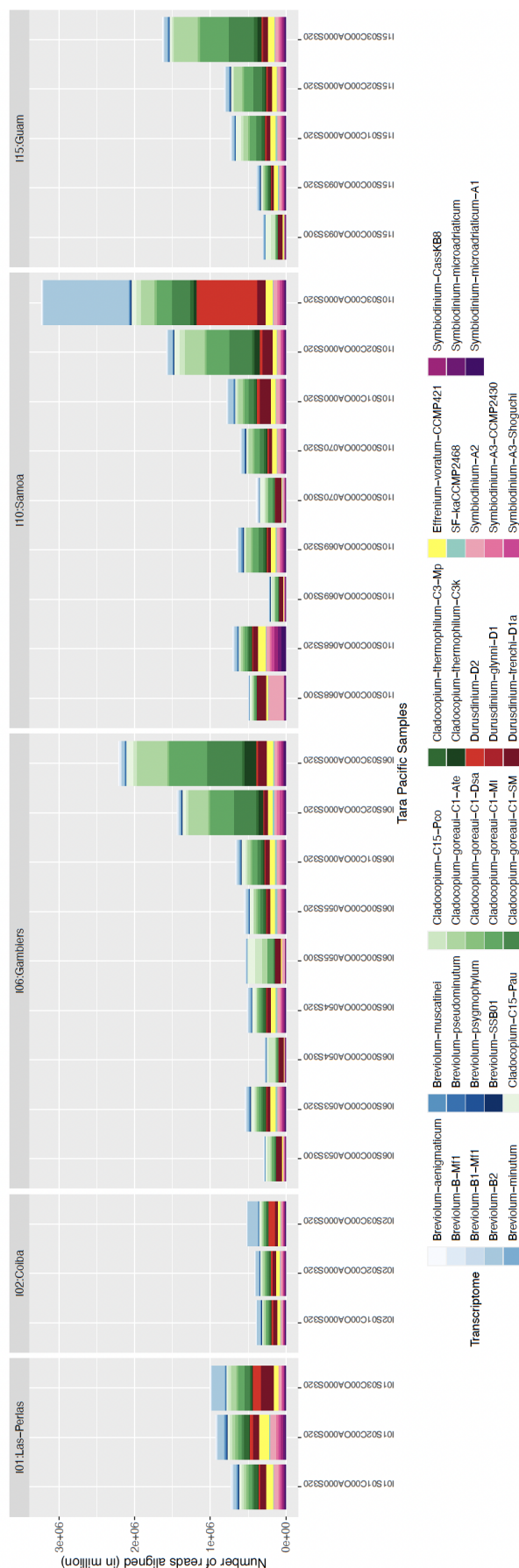


Figure IV.5: Barplots of the number of Surface samples reads aligned on Symbiodiniaceae reference transcriptomes.

Each color corresponds to a specific transcriptome.

#### IV.2.1.4 Population structure of *Cladocopium C1* in coral surrounding water samples

I performed a population structure of *Cladocopium C1* on the CSW samples using the protocols developed previously (**Chapter 2**).

SNPs were identified within *Cladocopium* transcripts using the GATK pipeline and using the protocol described above. A number of genes and a number of SNPs were obtained at each step (**Table IV.2**).

From these results I kept 494 SNPs for 193 genes which allow to build a dendrogram (**Figure IV.6**), with the Hclust function and the "Complete linkage" option.

Table IV.2: Number of remaining SNPs and number of genes covered by these SNPs in the different filtering steps of GATK pipeline for population structure of *Cladocopium C1* from CSW samples.

Tools	Programs	Number of genes	Number of SNPs
GATK - T GenotypeGVFs	Perform joint genotyping	41.330	394.238
Grep	Removed <i>Durusdinium D2</i> transcripts and samples	24.380	336.815
vcftools -min alleles 2 -max alleles 2	Keep only biallelic variance	24.355	306.745
bcftools view -i % qual $\geq$ 30	Filtered SNPs with quality score $\geq$ 30	24.328	301.504
Python program	Keep Biallelic SNPs SNPs Coverage $\geq$ 4 in all samples and frequency calculs	24.069	253.042

By looking at the dendrogram obtained and by comparing it with the different lineages obtained previously, I obtained very similar results. In fact, there are 5 lineages which are identical between our dual-transcriptomic and CSW samples. The lineage 1 is only present in Rapa Nui islands as obtained before. Then, the same lineages were found in the same samples except for 3 samples: I05S03C001, I06S03C010 and I07S02C010. There are only few SNPs detected in these samples that may explain the non attribution of these three samples to a specific lineage, in this analysis. For the other lineages, the results showed that the *Cladocopium C1* lineages are the same inside and outside the host.

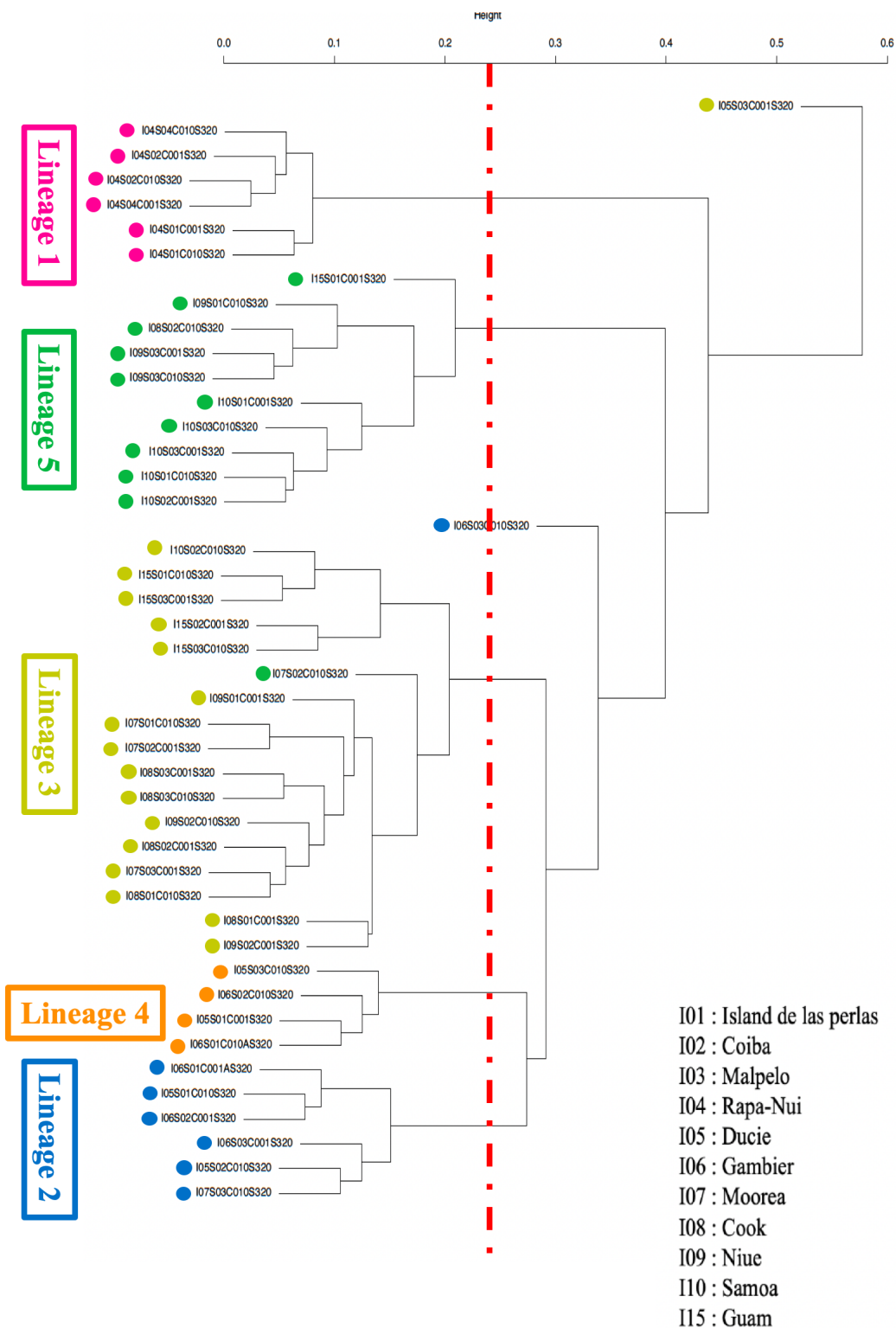


Figure IV.6: Hierarchical clustering of *Cladocopium C1* populations in CSW samples from metatranscriptomic reads.

Dendrogram representation of the hierarchical clustering of 494 SNPs detected by the mapping of metatranscriptomic reads on 193 genes of the *Cladocopium C1* reference genome. 55 *Cladocopium* samples were clustered into 5 genetic lineages based on a tree height cutoff threshold (red line) and using dual-transcriptomic results obtained in *Cladocopium C1* population structure.

## IV.2.2 Gene expression variations of *Cladocopium C1* between CSW and coral samples

In this part, I was interested in the genetic and the environmental adaptation of *Cladocopium C1* in surrounding water samples compared to dual-transcriptomic samples.

To do this, the bioinformatic methods developed previously (**Chapter 3**) were applied to the datasets. Then, lists of differentially expressed genes were thus obtained in the various tests applied.

### IV.2.2.1 Expression level of *Cladocopium C1* genes in coral surrounding water samples

First, I determined the different expression profiles according to the sampling islands and sites on CSW samples (**Figure IV.7**).

The first axis PC1 explains 13.67 % variance and separate the samples containing *Cladocopium* from those containing *Durusdinium*. This result is expected because they are two different genera.

The PC2 axis explains 11.06 % of the variance, it appears that the samples from Rapa Nui island are distinct from all the other samples. Once again, this island shows a different expression profile compared to the other islands. This is an expected result, because the population of *Cladocopium C1* associated with this island is very different from the other. Moreover, samples seem to be order by islands. There is the Central Pacific (Ducie, Gambier, Moorea, Cook and Niue islands) grouped together and then the West Pacific (Samoa and Guam islands) in an other group. It appears that this axis separate samples using genetic and environmental conditions.

To conclude on these results, it appears that the determining factor which can explain our different datasets is the environment materialized by the different islands of our different samples but also the genetic determined by Symbiodiniaceae genera and lineages obtained.

In an other analyse, I calculated Pearson correlations on gene expression levels between each pair of CSW samples (**Figures IV.8**).

These methods add a supplemental information on the expression profiles of the dataset. In fact, Pearson coefficients intra-sites and inter-sites showed similar coefficients for the different groups determined above.

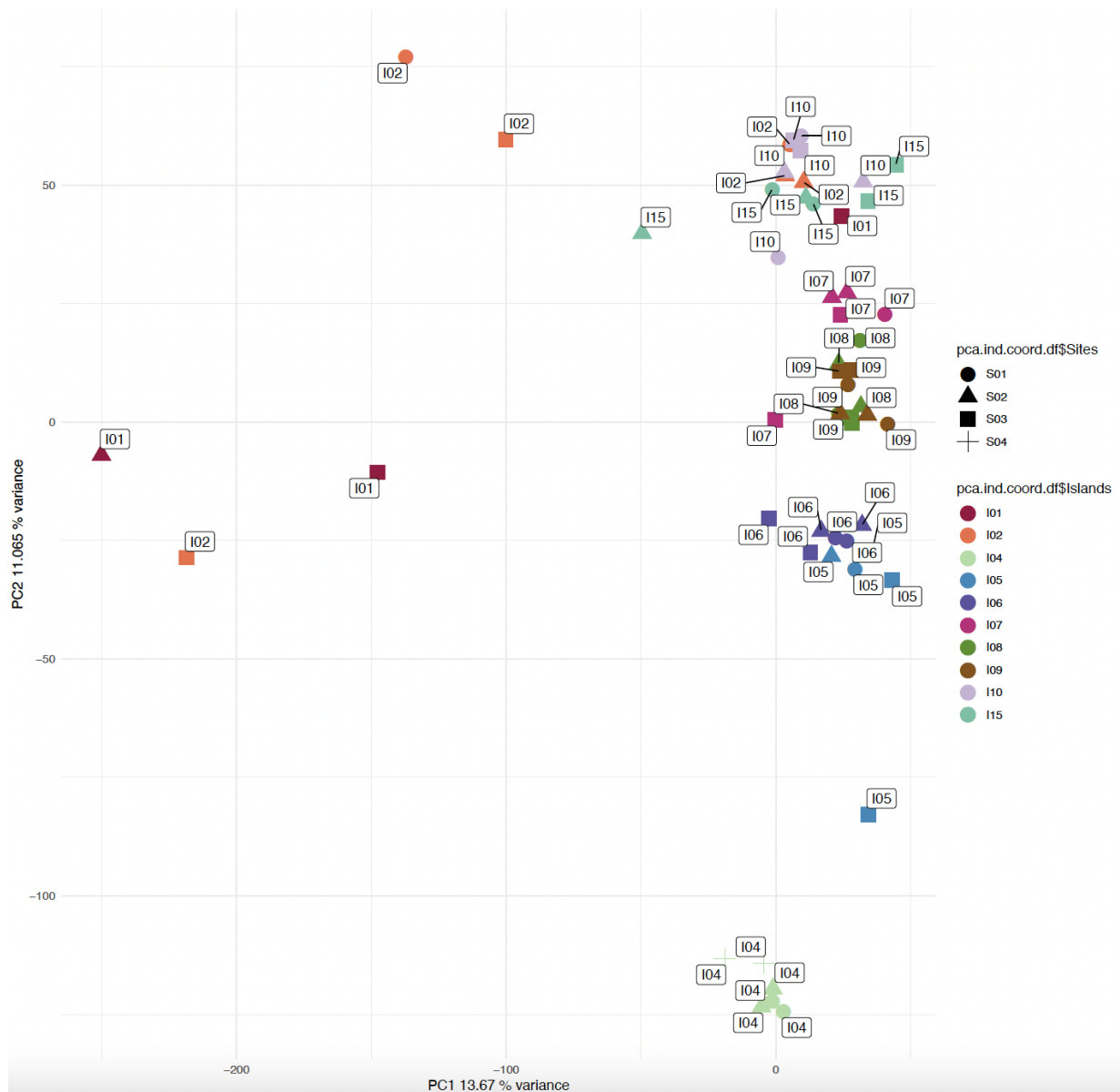


Figure IV.7: PCA of gene expression normalized in TPM of each CSW samples. Each color indicates an island; Forms indicate sites; I01: Island de las perlas, I02: Coiba, I04: Rapa Nui, I05: Ducie, I06: Gambier, I07: Moorea, I08: Cook, I09: Niue, I10: Samoa and I15: Guam.

From these results, it seems that samples from the same site can be considered as biological replicates, because their correlation coefficient is very close to 1 (higher than 0.7). Observations of the Pearson correlations between each island show similar results. There are correlations close to 1 between samples from the same island and these results are similar to those observed in dual-transcriptomic samples.

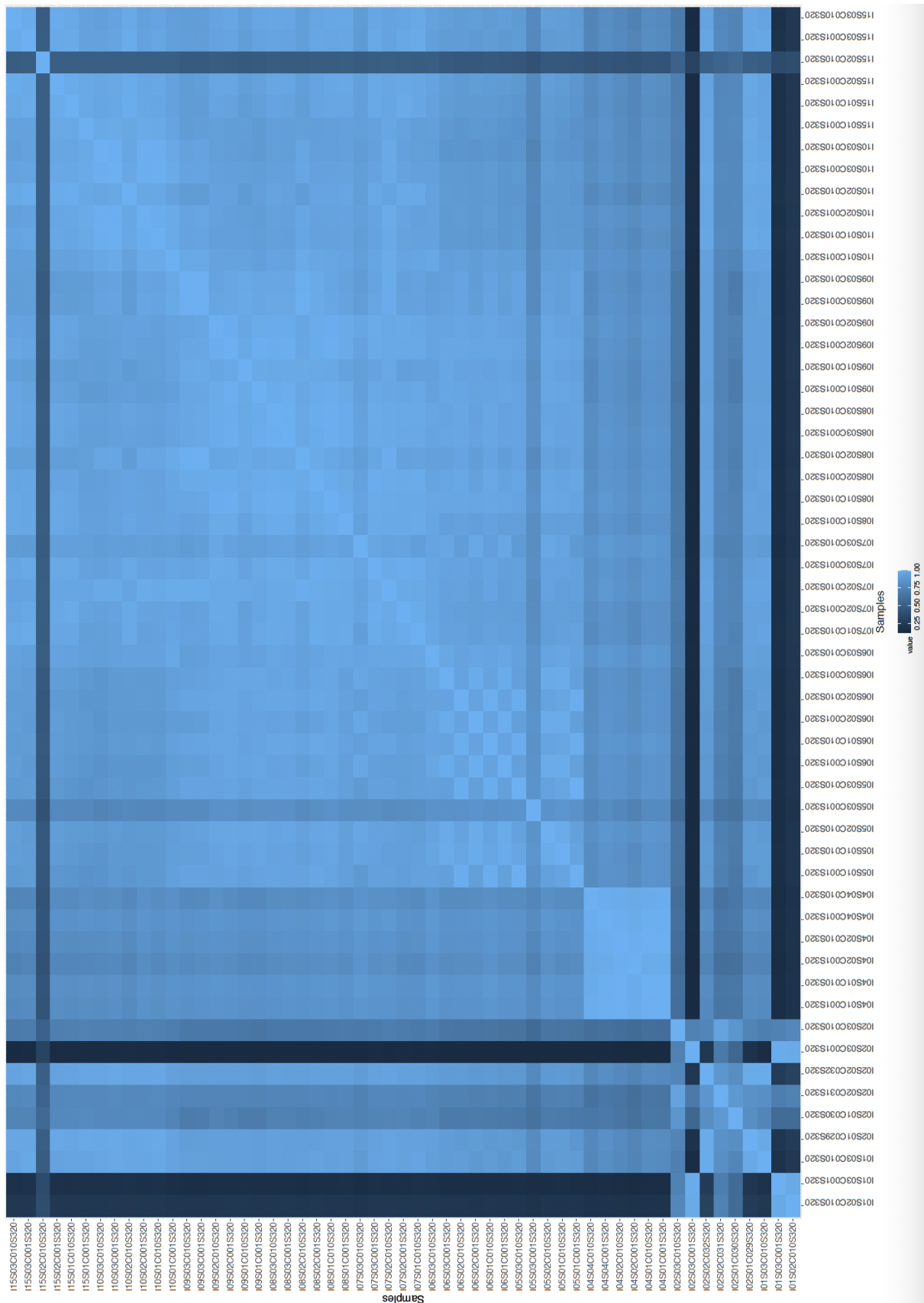


Figure IV.8: Pearson correlations on gene expression normalized in TPM of *Cladocopium C1* in CSW samples.

I01: Island de las perlas, I02: Coiba, I04: Rapa Nui, I05: Ducie, I06: Gambier, I07: Moorea, I08: Cook, I09: Niue, I10: Samoa and I15: Guam.

### IV.2.2.2 Differentially expressed genes of *Cladocopium C1* between coral and water samples

Two different clades of Symbiodiniaceae were detected in the CSW samples. Metatranscriptomic reads were aligned on *Cladocopium C1* CDS and/or *Durusdinium D2* transcriptome.

The C1 genome CDS contains a total of 35.913 CDS and a total of 25.413 CDS were aligned at 98% identity on the CSW samples. From these results, the goal was to detect differentially expressed genes between the different islands for *Cladocopium C1*, after removed samples with *Durusdinium D2*.

I obtained a list of differentially expressed genes (**Figure IV.9**) and I observed a very high number of DEG between the island of Rapa Nui compared to the other islands. These results again confirms a different expression profile for the samples belonging to this island.

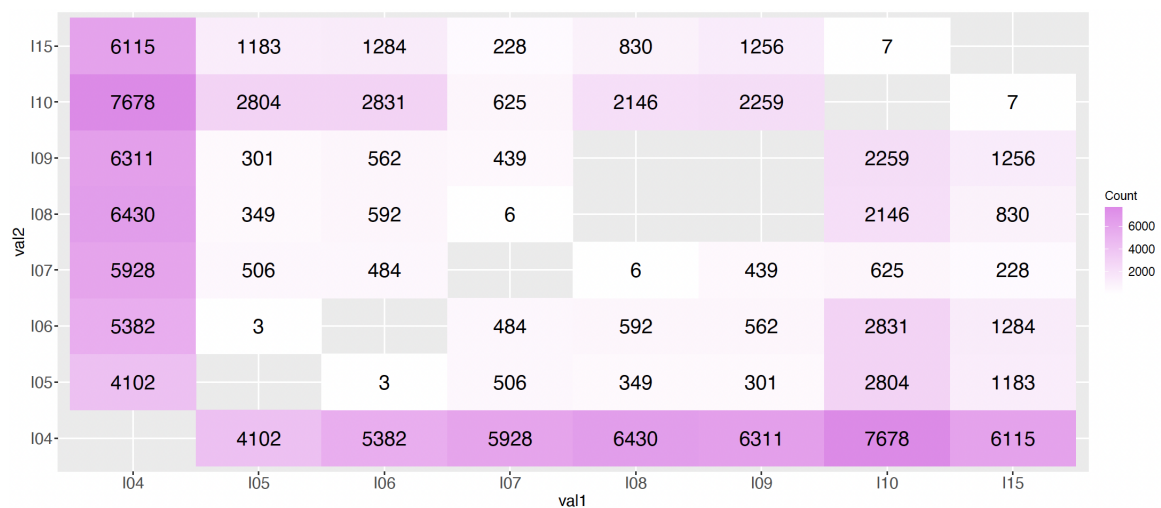


Figure IV.9: Number of differentially expressed genes in *Cladocopium C1*, from the CSW samples, using island as biological replicates.

I01: Island de las perlas, I02: Coiba, I04: Rapa Nui, I05: Ducie, I06: Gambier, I07: Moorea, I08: Cook, I09: Niue, I10: Samoa and I15: Guam.

Moreover, comparisons between Gambier and Ducie islands showed only 3 DEG between the two islands. The same proximity is observed between Moorea and Aitutaki islands (7 DEG) and Samoa and Guam islands (7 DEG).

Because there is the same number of biological replicates in these analyses, it appears that these low number observed in these comparisons, are not due to a bias of the method. It appears a very high proximity between these pairs of islands, linked to environmental conditions.



In an other analysis, the environmental adaptation of *Cladocopium C1* by taking in account the islands and the environment of the samples (DualT versus CSW) was done. I performed the differential expression analyses using these different conditions taking in account the colonies of one island and one environment as biological replicates (**Figure IV.10**).

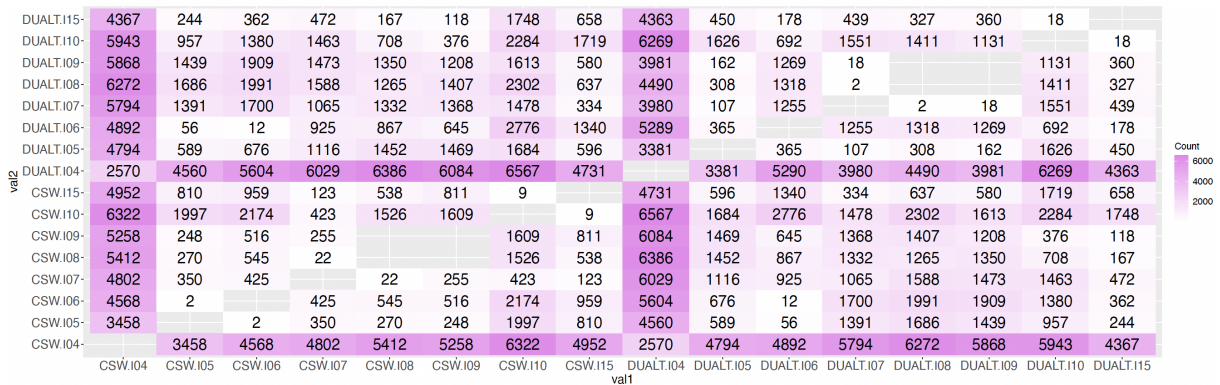


Figure IV.10: Number of differentially expressed genes in *Cladocopium C1*, using samples from the same island and environment as biological replicates. I01: Island de las perlas, I02: Coiba, I04: Rapa Nui, I05: Ducie, I06: Gambier, I07: Moorea, I08: Cook, I09: Niue, I10: Samoa and I15: Guam.

It can be seen that there is a very high number of DEG between CSW and DualT samples between Rapa Nui and the other islands (more than 3.458 DEG with the other conditions). Moreover looking at the DEG between CSW and DualT from Rapa Nui there is a very high number of DEG (2.570 DEG).

Observations between other patterns by comparing CSW vs DualT showed different patterns. Indeed, in Gambier island there is only 12 DEG between CSW and DualT conditions. In other cases there are between 589 and 2284 DEG comparing CSW and DualT samples in the different islands. These numbers of DEG may be related to the fact that the *Cladocopium C1* lineages was not taking in account.

In a last analysis, the environmental and the genetic adaptation of *Cladocopium C1* by taking in account the islands, the environments and the genetic of the samples was done between these different conditions (**Figure IV.11**).

It can be seen that there is a very low number of DEG between CSW and dual-transcriptomic samples in the first comparison (A). There are 37 DEG explained in Ducie island between two different environments. Then, there are 12 DEG in Gambier island between two different con-

ditions. These results showed a very high proximity in expression profiles between CSW and dual-transcriptomic samples.

By looking at the *Cladocopium* lineage 3 in SVD3, similar results were observed between Niue samples (43 DEG) and Guam samples (411 DEG). In comparison, samples of Aitutaki showed a very high number of DEG (1.056). These results that may be the Symbiodiniaceae living outside the host is under different environmental conditions in this island.

A low number of DEG is observed in lineage 3 and SVD5 host in Moorea island (168 DEG). Then, in lineage 4 and SVD4 there are similar results than observed before (low number of DEG between Ducie and Gambier islands).

To finish, in lineage 5 and SVD2 host, a low number of DEG is observed between two kind of samples in Aitutaki (53 DEG) whereas it is a high number in Samoa island (2189 DEG)

Observations between these different patterns by comparing CSW vs DualT showed that there are in some islands different profiles to understand the Symbiodiniaceae adaptation. Indeed, in Samoa, Ducie and Aitutaki islands, there is high number of DEG between CSW and DualT conditions, whereas in the other cases, there are between 589 and 2284 DEG comparing CSW and DualT samples in the different islands.

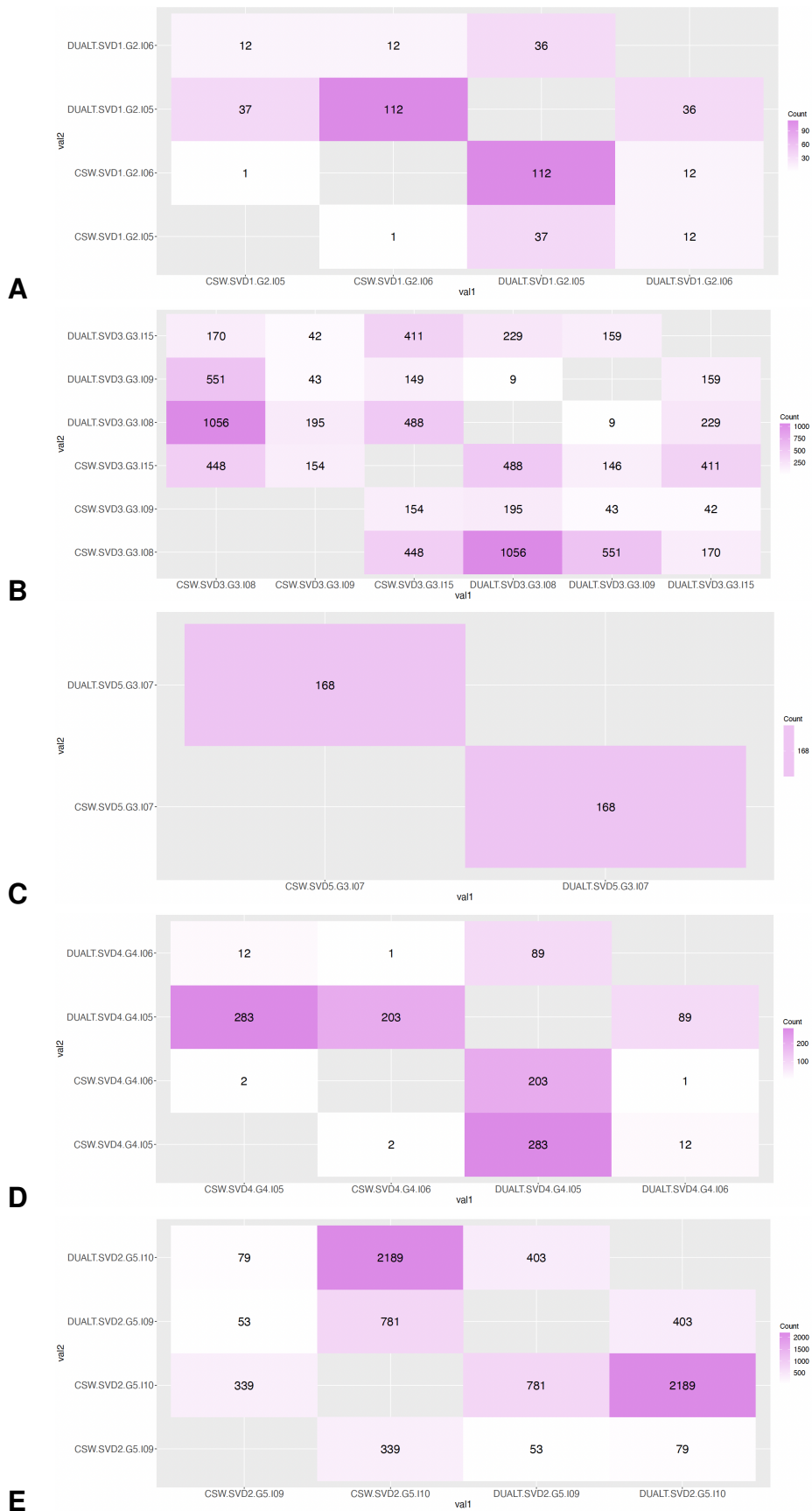


Figure IV.11: Number of differentially expressed genes of *Cladocopium C1* taking in account the genetic and the environment, in and out of *Pocillopora sp.* colonies.

(A) In lineage 2 host SVD1 (B) In lineage 3 host SVD3 (C) In lineage 3 host SVD5 (B) In lineage 4 host SVD4 (B) In lineage 5 host SVD2; I01: Island de las perlas, I02: Coiba, I04: Rapa Nui, I05: Ducie, I06: Gambier, I07: Moorea, I08: Cook, I09: Niue, I10: Samoa and I15: Guam.

## IV.3 Discussion and Conclusions

### IV.3.1 bioinformatics methods to understand the acclimation of Symbiodiniaceae hospite versus free-living.

Thanks to new sequencing technologies, it is now possible to sequence a multitude of samples at a lower cost.

As part of the *Tara* Pacific expedition, a multitude of samples were sequenced, to better understand the adaptation and the genetic of coral reefs at the oceanic scale. Because many samples were sequenced during this expedition, it is possible to answer global questions about the adaptation of the coral to environmental changes in the Pacific Ocean.

Finally, thanks to the multitude of genomic and transcriptomic references in Symbiodiniaceae and thanks to bioinformatics methods, it is now possible to perform high throughput data analyses on several hundred samples.

### IV.3.2 Symbiodiniaceae diversity hospite versus free-living

A comparative study of Symbiodiniaceae in *Pocillopora sp.* colonies versus their presence in the immediate environment was performed.

This study has shown that the analysis of the diversity in Symbiodiniaceae using the reference transcriptomes makes it possible to identify the species found in the surrounding water.

Moreover, it was observed that the diversity in Symbiodiniaceae from a transcriptomic point of view is identical in the seawater and in the coral samples. Thus, it is therefore possible to say that the Symbiodiniaceae living in and around the coral have identical genetic diversity.

By comparing these results with the population structure obtained on the dual-transcriptomic samples, very similar results were obtained. In fact, the lineages in the surrounding water samples are identical to the hospite Symbiodiniaceae. This result suggests that a reservoir of the coral symbiont is always present around the coral colony.

This identical diversity present near to the coral colony may be essential for the recovery after a bleaching event.

Comparison of surface water samples further from the coral colony indicates a more complex diversity. Indeed, several clades of Symbiodiniaceae are present in surface water samples.

In comparison with the surrounding water samples, it appears that the diversity in Symbiodiniaceae near the coral colony is defined by its host. This means that the coral host permanently lives with an associated symbiont close to it. These results can therefore claim that it is an adaptation of the host organism and the associated symbiont within the framework of the coral symbiosis. Indeed, it can be assumed that the symbionts living near the coral are partially exchanged and reintegrated permanently with these results obtained. This mechanism supposed an adaptation of the coral and the symbiont to their environment and therefore to their respective way of life.

### **IV.3.3 *Cladocopium C1* adaptation in seawater versus coral host**

In these analyses, it was observed that the gene expression levels of *Cladocopium C1* in the surrounding water is driven by the genetic lineages of Symbiodiniaceae and by the sampled islands. These results are consistent with the previous analyses on dual-transcriptomic samples.

In addition, the Pearson correlations associated with each sample pair showed similar intra-site and inter-site coefficients.

Looking at the differentially expressed genes between the coral surrounding water samples, I observed that the island of Rapa Nui seems different from the other islands even in the surrounding water samples.

In another analysis, looking at the environment between the dual-transcriptomic and CSW samples, I got very similar results. Again Rapa Nui Island seems very different from other islands. This additional analysis, however, showed that between some DualT samples compared to CSW samples, there are few DEG. This means that under certain environmental conditions *Cladocopium C1* living in and outside the host coral does not be affected by different environmental conditions.

To finish on these analyses, the differential expression patterns taking in account genetics and the environment show very different results depending on the conditions used. Indeed, for some islands (Samoa, Ducie and Aitutaki islands), there is a strong difference in expression between the surrounding water samples and the dual-transcriptomic samples. This means that for these islands there is a disturbance in the expression of Symbiodiniaceae.

Other expression profiles showed opposite results with very few transcripts differentially expressed between surrounding water samples and dual-transcriptomic samples.

The gene functions of these differentially expressed genes are in progress, and they will be presented in a future second article as first author after the thesis.



## Chapter V

# Discussion and Conclusions

### V.1 Symbiodiniaceae diversity study in the *Tara Pacific* project

The results obtained on Symbiodiniaceae diversity study in the *Tara Pacific* project, showed that there is a close relationship between the coral host and its associated symbiont.

Indeed, for the 3 corals studied from the *Tara Pacific* expedition, we observed a diversity in Symbiodiniaceae which balances between the fidelity and the flexibility of this symbiosis with the host.

For the coral *Porites*, a very high fidelity of the symbiotic microalgae *Cladocopium C15* was observed, through the samples from the 4 islands distributed in the Pacific Ocean. These results therefore indicate that the symbiont *Cladocopium C15* is extremely faithful to its coral host under different environmental conditions.

For the coral *Millepora*, we observed a flexibility of the coral symbiosis, because the results showed the presence of two different microalgae, belonging to two different genera *Symbiodinium A3* and *Cladocopium C66*. These results therefore indicate that for different marine environments and for this coral *Millepora*, there is a flexibility of the coral symbiosis allowing this change in the population of symbiotic microalgae, and thus allowing a better adaptation of the coral and of the symbiont under different environmental conditions.

For the last coral *Pocillopora*, the results showed that coral symbiosis depends on the flexibility and fidelity of the symbiotic microalgae. Indeed, the results on the diversity for this coral showed an exchange between two different species belonging to two different genera which



are *Cladocopium C1* and *Durusdinium D2*, following thermal stress in the marine environment. However, the study of the genetic lineages of *Cladocopium C1* and its comparison with the genetic lineages of the host *Pocillopora*, showed that two identical genetic lineages of *Cladocopium C1* can occur in two different hosts.

These results observed for the three different coral symbioses made it possible to measure the importance of the flexibility and fidelity of the symbiosis between the coral host and the microalgae.

However, to carry out this type of analysis it is necessary to have a large number of samples to observe these interactions.

These results on the diversity of Symbiodiniaceae in the context of the *Tara Pacific* expedition, showed that it is also necessary to have a good taxonomic resolution to understand host/symbiont interactions. Indeed, a structure of the populations of *Cladocopium C1* obtained from the transcriptomes has been produced, thanks to the use of metabarcoding data and thanks to the use of metatranscriptomic data.

The use of the metatranscriptomic data has shown a more accurate way of determining the population structure of microalgae, than the use of marker genes to identify symbionts. However, the use of this type of data remains a higher cost.

## **V.2 Adaptation of Symbiodiniaceae in symbiosis across the Pacific Ocean**

As part of the studies on the adaptation of Symbiodiniaceae in symbiosis across the Pacific Ocean, most of the observed expression variations are explained by the different lineages of Symbiodiniaceae.

Indeed, for the coral *Pocillopora* and the microalgae *Cladocopium C1*, the most distant lineage (L1) has an expression profile very divergent from the other expression profiles of the

other four genetic lineages.

The analysis of the expression profiles of the 4 other lineages of *Cladocopium C1*, also showed slightly different expression profiles. These genetic lineages appear to be environmentally determined through the use of surface temperature analyzes and by obtaining genetic lineages from the coral host.

The genes differentially expressed between the lineages are mainly involved in the metabolisms of photosynthesis, nutrient transport and cytoskeleton/motility. These metabolic pathways are essential to the lifestyle of the symbiotic microalgae. These results therefore suggest an adaptation of the genes involved in these metabolisms for the survival of the symbiotic microalgae.

The results also showed that there was little environment-related transcriptomic regulation in *Cladocopium C1*. Indeed, the Symbiodiniaceae are not in direct contact with the environment within the framework of the coral symbiosis. In addition, translational regulation is important in dinoflagellates, which may explain the results obtained.

### **V.3 Adaptation of Symbiodiniaceae in surrounding water across the Pacific Ocean**

The study on the adaptation of Symbiodiniaceae in surrounding water across the Pacific Ocean, showed that Symbiodiniaceae in the water around *Pocillopora* colonies possess identical genetic lineages inside and outside the coral.

Indeed, the results showed that the 5 genetic lineages are found almost identically in and outside the coral. These results suggest that there are probably always microalgae being expelled and others re-entering the host. This mechanism of expulsion and recruitment of microalgae therefore appears to be essential to coral in the event of bleaching events.

Finally, the diversity of Symbiodiniaceae near the coral colony therefore suggests that it is probably essential to the resilience of coral organisms.

## V.4 Prospects

The results obtained within the framework of this thesis are to be continued. Indeed, studies of the genetic lineages of microalgae are to be carried out in *Millepora* and *Porites* corals, in order to better understand the flexibility and fidelity of coral symbiosis. These results would therefore make it possible to better understand the modes of adaptation of symbiotic coral microalgae in different environments.

In addition, for these two organisms, the study of these two kind of symbiosis would make it possible to better understand the variations in the level of expression of the genes of microalgae.

The thesis focused on the study of samples from the first 11 islands of the *Tara* Pacific expedition. The analyzes must be continued with all of the other samples from the Pacific Ocean. These results would therefore make it possible to have a greater diversity of marine environments to better understand the adaptation of microalgae with its coral host. Moreover, the study of these samples on this scale and with this large number of samples would make it possible to have more robust statistical analyses.

To continue within the framework of these studies, the analysis of the acclimatization of Symbiodiniaceae outside the coral is to be completed in order to understand the adaptation of microalgae in two different lifestyles.

These analyzes are also open to study the importance of other organisms of the coral holobiont, such as viruses, bacteria and other eukaryotes, in order to better understand how the coral entity works with its microbiome.

# BIBLIOGRAPHY

- Alexa A. and Rahnenfuhrer J. 2021. topGO: Enrichment Analysis for Gene Ontology *Bioconductor, R package version 2.44.0*. <https://bioconductor.org/packages/release/bioc/html/topGO.html>.
- Anders S. and Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology*. <https://doi.org/10.1186/gb-2010-11-10-r106>.
- Aranda M., Li Y., Liew Y. *et al.* 2016. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci Rep* 6, 39734. <https://doi.org/10.1038/srep39734>.
- Armstrong Eric J, Lê-Hoang Julie, Carradec Quentin, Aury Jean-Marc, Noel Benjamin, Poulain Julie, Belser Caroline, Da Silva Corrine, Wincker Patrick, Tara Pacific Consortium. Submitted. Transcriptomic plasticity and symbiont shuffling underpin *Pocillopora* acclimatization across heat-stress regimes in the Pacific Ocean. *bioRxiv* 2021.11.12.468330. <https://doi.org/10.1101/2021.11.12.468330>
- Ashburner M., Ball C., Blake J. *et al.* 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25(1):25-29. <https://doi.org/10.1038/75556>.
- Baumgarten S., Bayer T., Aranda M. *et al.* 2013. Integrating microRNA and mRNA expression profiling in *Symbiodinium microadriaticum*, a dinoflagellate symbiont of reef-building corals. *BMC Genomics* 14. <https://doi.org/10.1186/1471-2164-14-704>
- Bayer T., Aranda M., Sunagawa S., Yum L.K., Desalvo M.K., Lindquist E., Coffroth M.A., Voolstra C.R., Medina M. 2012. *Symbiodinium* Transcriptomes Genome Insights into the Dinoflagellate Symbionts of Reef-Building Corals. *PLoS One.* 2012;7(4):e35269. doi: 10.1371/journal.pone.0035269
- Bellantuono A.J., Hoegh-Guldberg O., Rodriguez-Lanetty M. 2012. Resistance to thermal stress in corals without changes in symbiont composition. *Biological Sciences*. doi: 10.1098/rspb.2011.1780.

- Bikerland C. *et al.* 1996. Life and Death of Coral Reefs. *Geological Magazine*, 134(4), 575-582. doi:10.1017/S0016756897237386.
- Bouchard J.N., Yamasaki H. 2008. Heat Stress Stimulates Nitric Oxide Production in *Symbiodinium microadriaticum* A possible linkage between Nitric Oxide and the Coral Bleaching Phenomenon. *Plant Cell Physiology*. doi: 10.1093/pcp/pcn037.
- Cary G.A., Cameron R.A., Hinman V.F. 2018. EchinoBase: Tools for Echinoderm Genome Analyses. *Methods Mol Biol.*, 1757:349-369. doi:10.1007/978-1-4939-7737-6\_12, 2018.
- Cevasco G.A. *et al.* 1997. Biographical Dictionary of American and Canadian Naturalists and Environmentalists. *Greenwood Press - ISBN : 978-0-313-23047-9.*
- Coffroth M.A., Santos S.R. 2005. Genetic Diversity of Symbiotic Dinoflagellates in the *genus Symbiodinium*. *Protist*. doi:10.1016/j.protis.2005.02.004.
- Coffroth M.A. and Shearer T.L. 2008. Barcoding corals: limited by interspecific divergence, not intraspecific variation. *Molecular Ecology Resources (2008) 8*, 247–255. doi: 10.1111/j.1471-8286.2007.01996.x.
- Daly *et al.*, 2007. The phylum Cnidaria: A review of phylogenetic patterns and diversity 300 years after Linnaeus *Zootaxa*, 1668(1), 127–182.. doi:10.11646/zootaxa.1668.1.11
- Danecek P., Auton A., Abecasis G., *et al.* 2011. The variant call format and VCFtools. *Bioinformatics 27*:2156–2158. doi:10.1093/bioinformatics/btr330.
- Darwin C. R., 1842. The structure and distribution of coral reefs. Being the first part of the geology of the voyage of the Beagle, under the command of Capt. *Fitzroy, R.N. during the years 1832 to 1836. London: Smith Elder and Co.*
- Decelle *et al.*, 2018, Worldwide Occurrence and Activity of the Reef-Building Coral Symbiont *Symbiodinium* in the Open Ocean *Current Biology 28*, 3625–3633 November 19, 2018 - Elsevier Ltd.. doi.org/10.1016/j.cub.2018.09.024.
- Douglas A.E. *et al.* 2003. Coral Bleaching How and Why. *Marine Pollution Bulletin*. [https://doi.org/10.1016/S0025-326X\(03\)00037-7](https://doi.org/10.1016/S0025-326X(03)00037-7).

- Dubinsky Z., *et al.* 1990. Coral Reefs. *Elsevier, Amsterdam*. ISBN 10: 0444873929 / ISBN 13: 9780444873927
- Fabricius K.E. 2005. Effects of terrestrial runoff on the ecology of corals and coral reefs review and synthesis. *Marine Pollution Bulletin*. <https://doi.org/10.1016/j.marpolbul.2004.11.028>.
- Fang H. *et al.* 2014. dcGOR: an R package for analysing ontologies and protein domain annotations. *PLoS Comput Biol*. 2014;10(10):e1003929. doi:10.1371/journal.pcbi.1003929.
- Ferrier-Pagès *et al.* 2003. Effect of natural zooplankton feeding on the tissue and skeletal growth of the scleractinian coral *Stylophora pistillata*. *Springer*, 22(3), 229–240. doi:10.1007/s00338-003-0312-7.
- Fitt W.K. and Trench R.K. 1983. Endocytosis of the symbiotic dinoflagellate *Symbiodinium microadriaticum* Freudenthal by endodermal cells of the scyphistomae of *Cassiopeia xamachana* and resistance of the algae to host digestion. *Journal of Cell Science*. PMID: 6141174
- Forsman Z.H., Ritson-Williams R., Tisthammer K. *et al.* 2020. Host-symbiont coevolution, cryptic structure, and bleaching susceptibility, in a coral species complex (Scleractinia; Poritidae). *Sci Rep* 10, 16995. <https://doi.org/10.1038/s41598-020-73501-6>
- Fournier A. *et al.* 2013. The story of symbiosis with zooxanthellae or how they enable their host to thrive in a nutrient poor environment. *University of Lyon*.
- Gardner S.G. *et al.* 2017. A multi-trait systems approach reveals a response cascade to bleaching in corals. *BMC Biology*. <https://doi.org/10.1186/s12915-017-0459-2>.
- Gierz S.L. *et al.* 2017. Transcriptomic Analysis of Thermally Stressed *Symbiodinium* Reveals Differential Expression of Stress and Metabolism Genes. *Frontiers in Plant Science*. doi: 10.3389/fpls.2017.00271.
- Gonzalez-Pech R.A. *et al.* 2017. Signatures of adaptation and symbiosis in genomes and transcriptomes of *Symbiodinium*. *Scientific Reports*. doi: 10.1038/s41598-017-15029-w.

- Granados-Cifuentes *et al.* 2013. Genetic diversity of free-living *Symbiodinium* in the Caribbean: the importance of habitats and seasons. *Coral Reefs* volume 34, pages 927–939 (2015). DOI: 10.1007/s00338-015-1291-1.
- Hawkins J. *et al.* 2004. Effects of Artisanal Fishing on Caribbean Coral Reefs. *Conservation Biology*. Volume 18, Issue 1. Pages 215-226 <https://doi.org/10.1111/j.1523-1739.2004.00328.x>.
- Hoegh-Guldberg O. *et al.* 2006. Climate change coral bleaching and the future of the world's coral reefs. *Marine and Freshwater Research* 50(8) 839 - 866 <https://doi.org/10.1071/MF99078>.
- Hoegh-Guldberg O. *et al.* 2007. Coral Reefs Under Rapid Climate Change and Ocean Acidification. *Science*. doi: 10.1126/science.1152509.
- Hoffman G.E., Schadt E.E. 2016. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* 17, 483. <https://doi.org/10.1186/s12859-016-1323-z>.
- Howells E.J. *et al.* 2013. Spatial and temporal genetic structure of *Symbiodinium* populations within a common reef-building coral on the Great Barrier Reef. *Molecular Ecology*. doi: 10.1111/mec.12342.
- Huang D. *et al.* 2015. The future of evolutionary diversity in reef corals. *Biological Sciences*. doi: 10.1098/rstb.2014.0010.
- Hughes T.P. *et al.* 2003. Climate Change Human Impacts and the Resilience of Coral Reefs. *Science*. doi: 10.1126/science.1085046.
- Hughes T.P. *et al.* 2017. Global Warming recurrent mass bleaching corals. *Science*. doi: 10.1126/science.1085046.
- Hughes T.P. *et al.* 2018. Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science* Vol 359, Issue 6371 DOI: 10.1126/science.aan8048.
- Hume B. *et al.* 2012. Corals from the Persian Arabian Gulf as models for thermotolerant reef-builders Prevalence of clade C3 *Symbiodinium* host fluorescence and *ex situ* temperature tolerance. *Marine Pollution Bulletin*. doi: 10.1016/j.marpolbul.2012.11.032.



- Hume B.C.C. *et al.* 2019. SymPortal: A novel analytical framework and platform for coral algal symbiont next-generation sequencing ITS2 profiling.. *Mol Ecol Resour.* 2019;19:1063–1080. <https://doi.org/10.1111/1755-0998.13004>.
- Iglesias-Prieto R. *et al.* 1992. Photosynthetic response to elevated temperature in the symbiotic dinoflagellate *Symbiodinium microadriaticum* in culture. *PNAS*. PMID: 11607337.
- Johnston E.C. *et al.* 2018. A simple molecular technique for distinguishing *species* reveals frequent misidentification of Hawaiian corals in the genus *Pocillopora*. *PeerJ*. doi: 10.7717/peerj.4355.
- Johnson Lisa, Alexander Harriet, Brown C. Titus. 2016. Marine Microbial Eukaryotic Transcriptome Sequencing Project, re-assemblies. *figshare. Dataset*. <https://doi.org/10.6084/m9.figshare.3840153.v3>.
- Kanehisa M. *et al.* 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27-30. doi:10.1093/nar/28.1.27
- Kayal E. *et al.* 2013. Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evol Biol* 13, 5. <https://doi.org/10.1186/1471-2148-13-5>.
- Keeling P.J. *et al.* 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) Illuminating the Functional Diversity. *Plos Biology*. <https://doi.org/10.1371/journal.pbio.1001889>.
- Kneeland J. *et al.* 2013. Lipid biomarkers in *Symbiodinium* dinoflagellates: new indicators of thermal stress. *Coral Reefs* 32, 923–934. doi: 10.1007/s00338-013-1076-3.
- Knowlton N. *et al.* 2015. Exploring Coral Reefs Using the Tools of Molecular Genetics *Springer Science Business Media Dordrecht*. DOI 10.1007/978-94-017-7249-5\_6.
- Kumagai N.H. *et al.* 2018. High-resolution modeling of thermal thresholds and environmental influences on coral bleaching for local and regional reef management. *PeerJ* 2018;6:e4382. doi:10.7717/peerj.4382.

- Ladner J. *et al.* 2012. Protein evolution in two co-occurring types of *Symbiodinium*: an exploration into the genetic basis of thermal tolerance in *Symbiodinium* clade D. *BMC Evolutionary Biology*. doi: 10.1186/1471-2148-12-217.
- LaJeunesse T.C. *et al.* 2004. "Species" Radiations of Symbiotic Dinoflagellates in the Atlantic and Indo-Pacific Since the Miocene-Pliocene Transition. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msi042.
- LaJeunesse T.C. *et al.* 2009. Do introduced endosymbiotic dinoflagellates 'take' to new hosts?. *Biological Invasions*. doi 10.1007/s10530-008-9311-5.
- LaJeunesse T.C. *et al.* 2018. Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts, *Current Biology*.  
<https://doi.org/10.1016/j.cub.2018.07.008>.
- Langfelder P., Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. <https://doi.org/10.1186/1471-2105-9-559>.
- Lawrence S. *et al.* 2017. Exploratory analysis of *Symbiodinium* transcriptomes reveals potential latent infection by large dsDNA viruses. *Environmental microbiology*. doi:10.1111/1462-2920.13782.
- Letunic *et al.* 2008. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci*. 33(3):101-3. <https://doi.org/10.1016/j.tibs.2008.01.001>.
- Levin R.A. *et al.* 2017. Engineering Strategies to Decode and Enhance the Genome of Coral. *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2017.01220>.
- Li H. and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324.
- Lin S. *et al.* 2015. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science*. doi: 10.1126/science.aad0408.
- Liu H. *et al.* 2018. *Symbiodinium* genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Commun Biol* 1, 95. <https://doi.org/10.1038/s42003-018-0098-3>.

- Macrander J.C. *et al.* 2017. Transcriptome sequencing and characterization of *Symbiodinium muscatinei* and *Elliptochloris marina*, symbionts found within the aggregating sea anemone *Anthopleura elegantissima*. *Marine Genomics*. <http://dx.doi.org/10.1016/j.margen.2017.08.010>.
- Mihailovich M. *et al.* 2010. Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays*. doi 10.1002/bies.200900122.
- Moberg F. *et al.* 1999. Ecological goods and services of coral reef ecosystems. *Ecological Economics*. [https://doi.org/10.1016/S0921-8009\(99\)00009-9](https://doi.org/10.1016/S0921-8009(99)00009-9).
- Mollica N.R. *et al.* 2018. OA affects coral growth by reducing density. *Proceedings of the National Academy of Sciences Feb 2018, 115 (8) 1754-1759*. DOI: 10.1073/pnas.1712806115.
- Muscatine L. *et al.* 1977. Reef Corals Mutualistic Symbioses Adapted to Nutrient-Poor Environments. *Oxford Journal*. <https://doi.org/10.2307/1297526>.
- Parkinson J. *et al.* 2016. Gene Expression Variation Resolves *species* and Individual Strains among Coral-Associated Dinoflagellates within the *genus Symbiodinium*. *Genome Biology Evolution (GBE)*. doi:10.1093/gbe/evw019.
- Paxton C.W. *et al.* 2013. Stress and death of cnidarian host cells play a role in cnidarian bleaching. *Journal of Experimental Biology*. doi: 10.1242/jeb.087858.
- Paz-García D.A. *et al.* 2015. Switch between Morphospecies of Pocillopora Corals. *NATURAL HISTORY NOTE The American Naturalist, Volume 186, Number 3*. <https://doi-org.insb.bib.cnrs.fr/10.1086/682363>.
- Phinney J.T. *et al.* 2006. Coral reefs and changing seawater carbonate chemistry. *Coral Reefs and Climate Change: Science and Management Volume 61, 10.1029/CE061(), 73–110*. doi:10.1029/61CE06 .
- Porter J.W. *et al.* 2001. Reef ecosystems: threats to their biodiversity. S.A. Levin (Ed.), *Encyclopedia of Biodiversity, Academic Press, San Diego (2001), pp. 73-95*. <https://doi.org/10.1016/B0-12-226865-2/00229-7>.

- Quigley K.M. *et al.* 2017. Heritability of the *Symbiodinium* community in vertically and horizontally transmitting broadcast spawning corals. *Scientific Reports*. doi: 10.1038/s41598-017-08179-4.
- Riddle D. 2016. An Update on *Symbiodinium* species and their hosts. *Smarter Reefs*.
- Robbins S.J. *et al.* 2019. A genomic view of the reef-building coral *Porites lutea* and its microbial symbionts. *Nat Microbiol* 4, 2090–2100 (2019).. <https://doi.org/10.1038/s41564-019-0532-4>.
- Rodríguez L., López C., Casado-Amezua P. *et al.* 2019. Genetic relationships of the hydrocoral *Millepora alcicornis* and its symbionts within and between locations across the Atlantic. *Coral Reefs* 38, 255–268. <https://doi.org/10.1007/s00338-019-01772-1>.
- Rosic N. *et al.* 2012. Phylogenetic analysis of genes involved in mycosporine-like amino acid biosynthesis in symbiotic dinoflagellates. *The ISME Journal*. doi: 10.1007/s00253-012-3925-3.
- Rowan R. *et al.* 1992. Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates zooxanthellae. *PNAS*. <https://doi.org/10.1073/pnas.89.8.3639>.
- Sadler L.A. *et al.* 1992. The nucleotide sequence of the small subunit ribosomal RNA gene from *Symbiodinium pilosum* a symbiotic dinoflagellate. *Genetics*. doi: 10.1007/BF00351703. PMID: 1525867.
- Safaie A. *et al.* 2018. High frequency temperature variability reduces the risk of coral bleaching *Nat Commun* 9, 1671. <https://doi.org/10.1038/s41467-018-04074-2>.
- Shinzato C. *et al.* 2014. A Snapshot of a Coral Holobiont A Transcriptome Assembly of the Scleratinian Coral, *Porites*, Captures a Wide Variety. *Plos One*. <https://doi.org/10.1371/journal.pone.0085182>.
- Shoguchi E. *et al.* 2013. Draft Assembly of the *Symbiodinium minutum* Nuclear Genome Reveals Dinoflagellate Gene Structure. *Current Biology*. <http://dx.doi.org/10.1016/j.cub.2013.05.062>.

- Shoguchi E. *et al.* 2018. Two divergent *Symbiodinium* genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. *BMC Genomics* 19, 458 (2018). <https://doi.org/10.1186/s12864-018-4857-9>.
- Shoguchi E. *et al.* 2020. A New Dinoflagellate Genome Illuminates a Conserved Gene Cluster Involved in Sunscreen Biosynthesis. *Genome Biology and Evolution*, Volume 13, Issue 2, February 2021. <https://doi.org/10.1093/gbe/evaa235>.
- Silverstein R.N. *et al.* 2014. Change in algal symbiont communities after bleaching not prior heat exposuer increases heat tolerance of reef corals. *Global Change Biology*. <https://doi.org/10.1111/gcb.12706>.
- Slavov C. *et al.* 2016. Super-quenching state protects *Symbiodinium* from thermal stress Implications for coral bleaching. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. doi: 10.1016/j.bbabi.2016.02.002.
- Smith E.C. *et al.* 2017. Genetic structure of coral-*Symbiodinium* symbioses on the worlds warmest reefs. *Plos One*. <https://doi.org/10.1371/journal.pone.0180169>.
- Spalding M. *et al.* 2017. Mapping the global value and distribution of coral reef tourism *Marine Policy*, Volume 82, August 2017, Pages 104-113. <https://doi.org/10.1016/j.marpol.2017.05.014>.
- Stothard P. *et al.* 2000. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28:1102-1104. DOI: 10.2144/00286ir01
- Supek F. *et al.* 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *Plos One*. <https://doi.org/10.1371/journal.pone.0021800>.
- Takama O. *et al.* 2018. Molecular Phylogeny Demonstrates the Need for Taxonomic Reconsideration of *species* Diversity of the Hydrocoral *genus Millepora Cnidaria Hydrozoa* in the Pacific. *Zoologie Sciences*. doi: 10.2108/zs170142.

- Tchernov D. *et al.* 2004. Membrane lipids of symbiotic algae are diagnostic of sensitivity to thermal bleaching in corals. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13531–13535.. doi: 10.1073/pnas.0402907101.
- Thompson J.R. *et al.* 2015. Microbes in the coral holobiont: partners through evolution, development, and ecological interactions. *Frontiers in Cellular and Infection Microbiology*, 4. doi:10.3389/fcimb.2014.00176.
- Weis V.M. 2008. Cellular mechanisms of Cnidarian bleaching stress causes the collapse of symbiosis. *Journal of Experimental Biology*. doi: 10.1242/jeb.009597.
- Wilkinson C.R. 1999. Ecological and Socioeconomic Impacts of 1998 Coral Mortality in the Indian Ocean: An ENSO Impact and a Warning of Future Change?.
- Wilkinson S.P. 2015. Intra-genomic variation in symbiotic dinoflagellates: recent divergence or recombination between lineages? *BMC Evol Biol* 15, 46 <https://doi.org/10.1186/s12862-015-0325-1>.
- Van der Auwera GA, O'Connor BD *et al.* 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. (1st Edition). O'Reilly Media, Inc
- Venn A.A. *et al.* 2006. The impact of coral bleaching on the pigment profile of the symbiotic alga *Symbiodinium*. *Plant Cell Environment*. DOI: 10.1111/j.1365-3040.2006.001587.x.
- Voolstra *et al.* 2015. The ReFuGe 2020 Consortium—using “omics” approaches to explore the adaptability and resilience of coral holobionts to environmental change. *Frontiers in Marine Science*, Volume 2, Pages 68. <https://www.frontiersin.org/article/10.3389/fmars.2015.00068>.
- Xiang T *et al.* 2015. *Symbiodinium* transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. *The plant journal*. doi: 10.1111/tpj.12789.
- Yorifuji M. *et al.* 2017. Shift of symbiont communities in *Acropora tenuis* juveniles under heat stress. *PeerJ*. doi 10.7717/peerj.4055.

Zaneveld J.R. *et al.* 2016. Overfishing and nutrient pollution interact with temperature to disrupt coral reefs down to microbial scales. *Nature*. <https://www.nature.com/articles/ncomms11833>.

Zdobnov E.M. *et al.* 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro *Bioinformatics*, Volume 17, Issue 9, September 2001, Pages 847–848,. <https://doi.org/10.1093/bioinformatics/17.9.847>.

# APPENDICES



## APPENDIX I : Résumé de la thèse en français

# Adaptation des micro-algues symbiotiques du corail aux changements environnementaux dans l'Océan Pacifique

## Introduction

Le terme corail est utilisé pour décrire des animaux marins que l'on retrouve dans l'embranchement des Cnidaires. Cet embranchement regroupe une multitude d'organismes marins divers et variés tels que les méduses, les coraux, les anémones etc. Aujourd'hui le terme corail est un terme générique qui désigne les organismes marins de cet embranchement, associés à un mode de vie sur support solide et contenant un squelette calcaire.

Chez les Cnidaires, on retrouve les organismes coralliens dans deux classes, qui sont les Anthozoaires et les Hydrozoaires. Dans la classe des Anthozoaires on retrouve les coraux appelés "coraux durs", dans les ordres des Hexacoralliaires et des Octocoralliaires. Pour la classe des Hydrozoaires, ces organismes coralliens sont plus communément appelés "coraux mous".

Le corail est un organisme généralement fixé sur un support solide, constitué de polypes, qui forment une colonie corallienne. Ces polypes lui permettent de se nourrir du plancton qui est nécessaire pour sa croissance. Les coraux possèdent deux modes de reproduction, ils peuvent se reproduire de manière sexuée ou asexuée. Ces organismes vivent dans des environnements marins avec des conditions environnementales de croissance et de reproduction différentes, en fonction des régions géographiques regardées (La température est optimale entre 22°C et 26°C dans l'Océan Pacifique, et au dessus de 30°C dans la Mer Rouge, par exemples). Ces polypes sont également utilisés pour identifier les espèces coralliennes.

Il existe une très grande diversité de coraux en terme de morphologies et d'espèces dans les deux classes coralliennes. On distingue une très grande richesse de forme et de structure entre les espèces d'un même genre (le genre *Millepora* par exemple). Cependant, pour certains genres la distinction des espèces reste complexe d'un point de vue morphologique,

dûe à une ressemblance importante des individus de différentes espèces au sein d'un même genre (Le genre *Pocillopora* par exemple). Ainsi, l'utilisation de gènes marqueurs tels que le gène COI<sup>1</sup> et l'utilisation des méthodes de séquençages à haut débit, permettent d'identifier ces espèces.

Les récifs coralliens sont un écosystème marin, représenté à travers différentes régions océaniques, et plus particulièrement près des régions côtières. On les retrouve dans les mers et les océans, et plus particulièrement, il y a une très grande concentration de ces organismes dans la région appelée "Triangle du corail" dans l'Océan Indo-Pacifique. Ces récifs coralliens présentent diverses formes (récifs frangeants, atolls, récifs barrières et bancs récifales) et sont à la base de la chaîne alimentaire marine. En effet, ils servent de refuge à un très grand nombre d'espèces marines, telles que les poissons, les crustacés, le zooplancton et le phytoplancton, c'est pourquoi ils constituent un réservoir de biodiversité, de l'environnement marin.

Les récifs coralliens sont également un atout majeur pour les populations qui vivent à proximité grâce à l'abondance des poissons qu'ils abritent. Ils sont également d'une importance capitale pour l'économie de ces régions, notamment par la protection des côtes face aux événements climatiques et pour le tourisme.

Les coraux vivent en symbiose obligatoire avec avec des organismes, qui sont des micro-algues photosynthétiques de la famille des Symbiodiniaceae.

Cette symbiose obligatoire est essentielle pour les deux organismes. En effet, la micro-algue photosynthétique produit grâce à la photosynthèse des nutriments qui sont essentiels à la vie du corail, dont 80 à 95% sont utilisés par l'organisme hôte. De plus la micro-algue produit des Micosporines like Amino Acids (MAAs) qui permettent une protection de l'hôte contre les radiations U.V. En échange, le corail fournit un environnement très concentré en nutriments qui sont nécessaires à la micro-algue, ainsi qu'une protection dûe à l'endosymbiose.

Aujourd'hui, les récifs coralliens sont menacés par des changements environnementaux globaux et locaux qui dégradent ces écosystèmes. Parmi ces changements environnementaux on trouve l'augmentation de la température des océans, des pollutions locales et une

---

<sup>1</sup>Cytochrome Oxidase I

augmentation de l'acidification des océans. Ces perturbations de l'environnement marin fragilisent la symbiose obligatoire entre le corail et sa micro-algue, pouvant entraîner une rupture symbiotique (blanchissement des coraux), puis la mort du corail.

Ce mécanisme de rupture symbiotique, suite à des changements dans l'environnement, induisent l'expulsion de la micro-algue photosynthétique de l'organisme corallien.

Les Symbiodiniaceae sont des micro-algues unicellulaires eucaryotes, appartenant aux dinoflagellés qui sont des protistes vivants dans l'environnement marin. Cette micro-algue est connue pour sa symbiose obligatoire avec l'organisme corallien. Elle peut vivre en symbiose ou en version libre, dans l'environnement marin.

La famille des Symbiodiniaceae regroupe une très grande diversité, classée en 9 genres. L'utilisation des séquences ITS2<sup>2</sup>, qui sont des séquences de 300 paires de bases, entre les ARN ribosomiaux 5.8S et 28S permettent de distinguer les différentes espèces de micro-algues. On retrouve ces dinoflagellés dans différents hôtes, appartenant à différents embranchements tels que les mollusques, les coraux ou les foraminifères par exemples.

Il existe des micro-algues spécifiques à un hôte et d'autres plus généralistes que l'on retrouve dans plusieurs hôtes différents. Une étude récente a permis d'actualiser les 9 genres de la famille des Symbiodiniaceae, en attribuant des noms aux genres les plus communs. Parmi ces genres, il y a principalement les genres *Symbiodinium*, *Cladocopium* et *Durusdinium* qui sont étudiés. Il existe une disparité en terme d'espèces entre les genres des Symbiodiniaceae. Par exemple, il y a très peu d'espèces répertoriées pour le genre *Durusdinium* alors qu'il y a plus de 50 espèces répertoriées pour le genre *Cladocopium*.

Au sein des différents genres de Symbiodiniaceae, il y a une spécificité des genres qui permettent une adaptation environnementale avec un hôte corallien donné. En effet, le genre *Durusdinium* est plus thermo-tolérant contrairement aux autres genres, ce qui permet une adaptation environnementale de son organisme hôte, après une augmentation de la température dans l'environnement marin par exemple. Ces espèces de Symbiodiniaceae sont également essentielles et fidèles à leur hôte, permettant chacune en fonction de la symbiose regardée une meilleure adaptation et une meilleure acclimatation à l'environnement.

---

<sup>2</sup>Internal Transcribed Spacer 2

Grâce aux nouvelles technologies de séquençage, il est aujourd'hui possible d'étudier de manière précise la diversité de ces micro-algues photosynthétiques. Il existe un très grand nombre de références transcriptomiques et génomiques, associées à cette micro-algue (plus de 30 références transcriptomiques et 8 références génomiques).

Dans le contexte des évènements de blanchissement des coraux en masse, la plupart des études génomiques et transcriptomiques tentent de comprendre les mécanismes permettant la stabilité de la symbiose corallienne dans des conditions de stress, afin de comprendre comment les récifs coralliens peuvent s'adapter dans ce contexte environnemental. Parmi ces études on trouve des recherches sur la régulation des gènes de Symbiodiniaceae, des études sur l'adaptation de la micro-algue dans des conditions de stress (augmentation de la température, de l'acidification, de la luminosité etc.). Ces études se portent également sur l'adaptation géographique des micro-algues dans différents environnements marins (Océan Pacifique et Mer Rouge par exemples).

Dans ce contexte, l'expédition *Tara Pacific* se propose d'étudier l'écosystème corallien à l'échelle d'un océan, en réalisant un vaste plan d'échantillonnage sur 32 îles réparties dans l'Océan Pacifique. Au cours de cette expédition (2016-2018), trois genres de coraux ont été sélectionnés et étudiés : *Pocillopora*, *Millepora* et *Porites*, afin d'avoir une vision globale de ces espèces à l'échelle océanique. Pour ce faire, des méthodes de séquençage haut débit en métagénomique, métatranscriptomique et métabarcodes ont été appliquées sur chaque échantillon de corail.

Dans le cadre de cette expédition et afin d'avoir une vision globale de l'écosystème corallien, des échantillons dual-transcriptomiques, dual-génomiques, des eaux environnantes, des eaux de surface et des sédiments, ont été prélevés, ainsi que les paramètres environnementaux tels que la température, la salinité, le pH etc.

Avec les données *Tara Pacific* des 11 premières îles, les objectifs de la thèse ont été de réaliser une étude transcriptomique globale de la micro-algue photosynthétique, en utilisant des outils bio-informatiques, afin de mieux comprendre sa diversité et son adaptation à l'échelle d'un océan.

Dans une première analyse, cette thèse tente de mieux comprendre la diversité des Symbiodiniaceae dans des échantillons de coraux à travers l'Océan Pacifique, pour les trois genres *Pocillopora*, *Millepora* et *Porites* qui ont été prélevés.

Ensuite, dans une autre analyse, la thèse cherche à comprendre et à étudier les gènes permettant une adaptation génétique et environnementale des Symbiodiniaceae dans ces différents coraux, en regardant les voies métaboliques associées aux gènes différentiellement exprimés sous plusieurs conditions.

Pour finir, une dernière analyse de la thèse est portée sur l'adaptation environnementale des micro-algues dans les échantillons des eaux environnantes et des échantillons de coraux. Le but de cette analyse est de déterminer la diversité des Symbiodiniaceae dedans et en dehors des colonies coralliennes, et de comprendre les gènes permettant leur adaptation à deux environnements différents.

## **Étude de la diversité des Symbiodiniaceae dans le cadre de l'expédition *Tara Pacific***

Dans une première analyse, la thèse a cherché à répondre à la question : Quelle est la diversité des micro-algues symbiotiques du corail dans trois coraux différents à travers l'Océan Pacifique ?

Pour ce faire, une première analyse des références génomiques et transcriptomiques de Symbiodiniaceae a été faite. Ces analyses ont permis de rassembler et de créer une base de données avec ces références, pour l'étude des Symbiodiniaceae dans les échantillons de l'expédition *Tara Pacific*. 28 références transcriptomiques de 6 genres différents et 6 références génomiques pour 4 genres différents ont été utilisés. Une recherche de la contamination potentielle pour chacun des transcrits et pour chacune des séquences codantes (CDS), a été réalisée en utilisant la base de données Uniref90 et une base de données créée au Genoscope regroupant des bases de données d'organismes marins (ReefGenomics, MMETSP, Refuge2020 etc.).

Les résultats de cette étude ont permis d'éliminer certaines références potentiellement contaminées et d'attribuer une affiliation taxonomique à chaque transcrit et CDS de chaque

référence de Symbiodiniaceae.

Ensuite, en utilisant des méthodes bio-informatiques et ces références mise à disposition par la communauté scientifique, il a été possible de déterminer la diversité des Symbiodiniaceae dans les échantillons dual-transcriptomiques de l'expédition *Tara Pacific*.

Pour le corail *Millepora*, les résultats ont montré la présence de deux clades différents, avec deux espèces différentes *Symbiodinium A3* et *Cladocopium C66*, dépendantes de la localisation géographique. En effet, dans le Pacifique Ouest *Cladocopium C66* domine, tandis que dans les autres îles du Pacifique Central et Est *Symbiodinium A3* est présent. Ces résultats montrent une adaptation environnementale de la symbiose corallienne pour le corail *Millepora* en fonction de sa localisation géographique.

Dans les échantillons de *Porites*, les résultats montrent une symbiose avec les micro-algues de l'espèce *Cladocopium C15* dans les quatre îles réparties à travers l'Océan Pacifique. Ces résultats montrent donc une très grande fidélité entre l'hôte et le symbiote nécessaire à leur mode de vie dans ces environnements différents.

Pour finir, les résultats ont montré que le corail *Pocillopora* possède deux genres de Symbiodiniaceae associés aux clades *Cladocopium* et *Durusdinium* en fonction des conditions environnementales. En effet, la présence du clade *Durusdinium* dans certains échantillons du Pacifique Est et un échantillon du Pacifique Ouest, indique un changement environnemental récent dans ces régions. Grâce à la littérature, ces résultats montrent qu'il y a eu une adaptation symbiotique du corail *Pocillopora* suite à un changement dans ces régions. En effet, le changement entre les deux clades *Cladocopium* pour *Durusdinium* indique une adaptation environnementale, car ce dernier clade, est plus thermo-tolérant contrairement aux micro-algues du genre *Cladocopium*.

Une analyse spécifique sur les 11 premières îles de l'expédition a été portée sur *Cladocopium C1*, dans les échantillons du corail *Pocillopora*. Pour réaliser cette étude, l'analyse de la structure des populations des micro-algues *Cladocopium C1* a été faite.

Les résultats ont permis de distinguer 5 lignées génétiques différentes dans ces échantillons. Ces lignées ont été obtenues en utilisant une approche nouvelle, à savoir la structure de population à partir des échantillons transcriptomiques.

La comparaison de ces lignées génétiques avec la structure de population du corail hôte et l'utilisation des profils ITS2, ont montré une très grande spécificité hôte-symbionte nécessaire à l'adaptation environnementale de ces organismes dans des conditions différentes. Ces résultats ont également permis d'utiliser une nouvelle approche pour définir les populations de Symbiodiniaceae en utilisant des échantillons transcriptomiques.

## **Étude de l'adaptation des Symbiodiniaceae en symbiose à travers l'Océan Pacifique**

Par la suite, cette thèse s'est portée sur l'étude des gènes permettant l'adaptation transcriptomique de ces micro-algues photosynthétiques dans les échantillons des trois coraux de l'expédition *Tara* Pacifique. Le but de ces études sur les expressions différentielles, est d'identifier les gènes permettant une adaptation génétique et environnementale de la micro-algue en symbiose, et par la suite de mieux comprendre les voies métaboliques portées par ces gènes différenciellement exprimés.

Pour ce faire, l'utilisation de méthodes bio-informatiques sur les échantillons dual-transcriptomiques ainsi que les résultats obtenus précédemment ont permis ces analyses.

Dans un premier temps, l'observation des profils d'expression de chacune des micro-algues obtenues dans les échantillons de coraux, a été faite en utilisant des outils de visualisation tels que la PCA. Puis une étude des corrélations de Pearson entre chaque paire d'échantillons pour chaque corail différent a été faite.

Les résultats de visualisation des profils d'expression ont permis de montrer que deux composantes majeures permettent d'expliquer nos jeux de données transcriptomiques. Ces paramètres sont la génétique des Symbiodiniaceae, qui permet l'explication de certains échantillons comme par exemple chez le corail *Pocillopora* ou deux genres différents sont présents dans les échantillons. Les analyses plus spécifiques des populations de *Cladocopium C1* viennent appuyer ces observations. En effet, les profils sont distincts entre les lignées génétiques de *Cladocopium C1* obtenues grâce à la structure de population.

Il y a également la distinction en utilisant le paramètre environnement, comme l'île qui permet d'obtenir des profils d'expression différents pour ces micro-algues. Ces résultats sont observés pour *Porites*, *Pocillopora* et *Millepora* où une distinction entre les différentes régions du Pacifique est observée.

L'utilisation des corrélations de Pearson a permis de montrer que l'environnement est déterminant pour expliquer les profils d'expression entre les échantillons. En effet, les résultats ont montré que les échantillons entre site d'une même île et entre même site sont plus fortement corrélés que les échantillons entre des îles différentes chez les micro-algues pour les trois organismes coralliens.

L'utilisation des outils classiques d'expression différentielle (package DESeq2 sous R) et les différentes définitions des différentes conditions ont permis d'obtenir des listes de gènes différentiellement exprimés entre plusieurs conditions.

Dans un premier temps, l'étude des gènes différentiellement exprimés entre les îles et entre les sites sans prendre en compte la génétique de l'hôte, ont permis de montrer que pour les trois coraux, l'utilisation des sites comme condition, ne permet pas d'obtenir les gènes impliqués dans l'adaptation environnementale du symbionte. Ces résultats ont cependant permis de montrer que lorsque l'on utilise les îles comme condition, des similarités et des dissimilarités sont observées entre les paires d'îles pour chaque corail. De ce fait, ces résultats ont donc montré qu'il était nécessaire de prendre en compte la génétique du symbionte pour comprendre l'expression de *Cladocopium C1* dans *Pocillopora*, *Symbiodinium A3* dans *Millepora* et *Cladocopium C15* dans *Porites*.

Les analyses d'expressions différentielles en prenant en compte seulement la génétique du symbionte *Cladocopium C1* dans les échantillons de *Pocillopora*, n'ont pas permis d'obtenir des résultats concluants.

Ainsi, afin de mieux comprendre l'adaptation des micro-algues symbiotiques du corail dans les échantillons dual-transcriptomiques, il est devenu nécessaire d'utiliser les paramètres tels que la génétique et l'environnement, sur les données dual-transcriptomiques. Cette étude spécifique a été faite sur *Cladocopium C1* dans les échantillons de *Pocillopora*.

Les résultats montrent que l'adaptation génétique de la micro-algue est déterminante pour l'adaptation du corail hôte dans différents environnements de l'Océan Pacifique.



Pour finir, dans une dernière analyse, les voies métaboliques impliquées dans l'adaptation génétique et environnementale de *Cladocopium C1* dans *Pocillopora* et l'adaptation environnementale de *Symbiodinium A3* dans *Millepora* ont été réalisées.

Les résultats ont montré que la photosynthèse et les transports de nutriments sont des voies métaboliques essentielles à l'adaptation environnementale de *Symbiodinium A3* dans les échantillons de *Millepora* à travers l'Océan Pacifique.

Les résultats sont similaires lorsque l'on regarde les gènes différenciellement exprimés de *Cladocopium C1* dans les échantillons de *Pocillopora*. Ces résultats ont également montré que la photosynthèse et les transports de nutriments sont essentiels à l'adaptation génétique et environnementale de *Cladocopium C1* dans les échantillons de *Pocillopora* à travers l'Océan Pacifique.

## **Étude de l'adaptation des Symbiodiniaceae dans l'eau environnante à travers l'Océan Pacifique**

Pour finir, cette thèse s'est consacrée à une dernière étude afin de comprendre l'adaptation des micro-algues symbiotiques du corail dans l'eau environnante à travers l'Océan Pacifique. Pour ce faire, deux questions se sont posées : La diversité en Symbiodiniaceae est-elle la même dans les échantillons de coraux et dans les échantillons d'eau environnante autour de la colonie corallienne ? Quels sont les gènes permettant l'adaptation des Symbiodiniaceae dans deux environnements différents, à savoir l'eau environnante et la vie en symbiose corallienne ?

Dans le cadre de l'expédition *Tara* Pacifique des échantillons d'eau environnante autour des colonies du corail *Pocillopora* ont été prélevés et séquencés en meta-barcode, meta-génomique et meta-transcriptomique.

Ainsi, dans un premier temps, une étude comparative sur la diversité des micro-algues photosynthétiques présentent dans l'hôte, avec celles présentent dans l'eau environnante a été conduite. Pour ce faire, les protocoles développés sur les échantillons de dual-transcriptomiques

ont été appliqués sur les échantillons d'eau environnante en utilisant les mêmes références génomiques et transcriptomiques précédentes.

Les résultats ont montré que la diversité en Symbiodiniaceae est identique entre les échantillons dual-transcriptomiques et les échantillons métatranscriptomiques d'eau environnante. En effet, l'utilisation de la structure de population dans les échantillons d'eau environnante a permis d'obtenir 5 lignées de *Cladocopium C1* dans les 11 îles à travers l'Océan Pacifique. En comparant ces lignées avec les lignées obtenues précédemment dans les échantillons dual-transcriptomiques, on obtient une correspondance identique entre la colonie corallienne et l'eau environnante.

Ces résultats permettent donc de montrer que la diversité en dehors et dans la colonie corallienne est identique. Il s'agit donc d'un mécanisme nécessaire à la symbiose corallienne, par l'expulsion et/ou le maintien constant d'une population de micro-algues spécifiques pour un hôte donné. Ces résultats sont montrés dans les 11 îles réparties dans l'Océan Pacifique.

Dans une deuxième analyse, l'étude de l'adaptation transcriptionnelle de la micro-algue dans deux environnements différents est étudiée. Le but de cette étude est de comprendre comment la micro-algue survie en dehors de son hôte, notamment lors d'évènements de blanchissement des coraux, et quels sont les gènes portés dans le cadre de cette adaptation.

Pour répondre à ces questions, des méthodes bio-informatiques appliquées aux échantillons dual-transcriptomiques ont été réalisées sur les échantillons d'eau environnante. Dans une première partie les profils d'expression à l'aide de l'outil PCA et des corrélations de Pearson ont été utilisés.

Les résultats ont montré que l'expression de *Cladocopium C1* peut être expliquée par la diversité génétique des lignées de *Cladocopium C1* et par les paramètres environnementaux (îles).

Ensuite, l'étude des gènes différentiellement exprimés a été menée en comparant les colonies d'une même île dans les deux environnements différents, avec ou sans prise en compte des lignées génétiques de *Cladocopium C1*.

Les résultats ont montré qu'il existe des gènes différentiellement exprimés entre des mêmes îles, pour deux environnements différents. En fonction de l'île regardée, il y a beaucoup ou très

peu de gènes différentiellement exprimés.

Dans le premier cas, où il y a beaucoup de gènes différentiellement exprimés entre les deux milieux d'une même île, on suppose qu'il y a des gènes impliqués dans l'adaptation de la micro-algue dans deux milieux différents. Ces gènes peuvent être liés à une adaptation environnementale ou une adaptation génétique de la micro-algue afin de mieux répondre à son mode de vie dans un environnement donné.

Dans le second cas, où il y a peu de gènes différentiellement exprimés entre les micro-algues d'une même île pour deux environnements différents, on suppose que les micro-algues en dehors et dans l'hôte possèdent des profils d'expression identiques, ce qui signifie que la micro-algue dans deux environnements différents possède la même adaptation génétique et environnementale.

L'étude des fonctions associées à ces gènes différentiellement exprimés dans ces deux environnements différents est à poursuivre. Ces études permettront de mieux comprendre les voies métaboliques impliquées dans l'adaptation environnementale de *Cladocopium C1* dans et en dehors de l'hôte corallien, à travers l'Océan Pacifique.

## APPENDIX II : Illumina sequencing

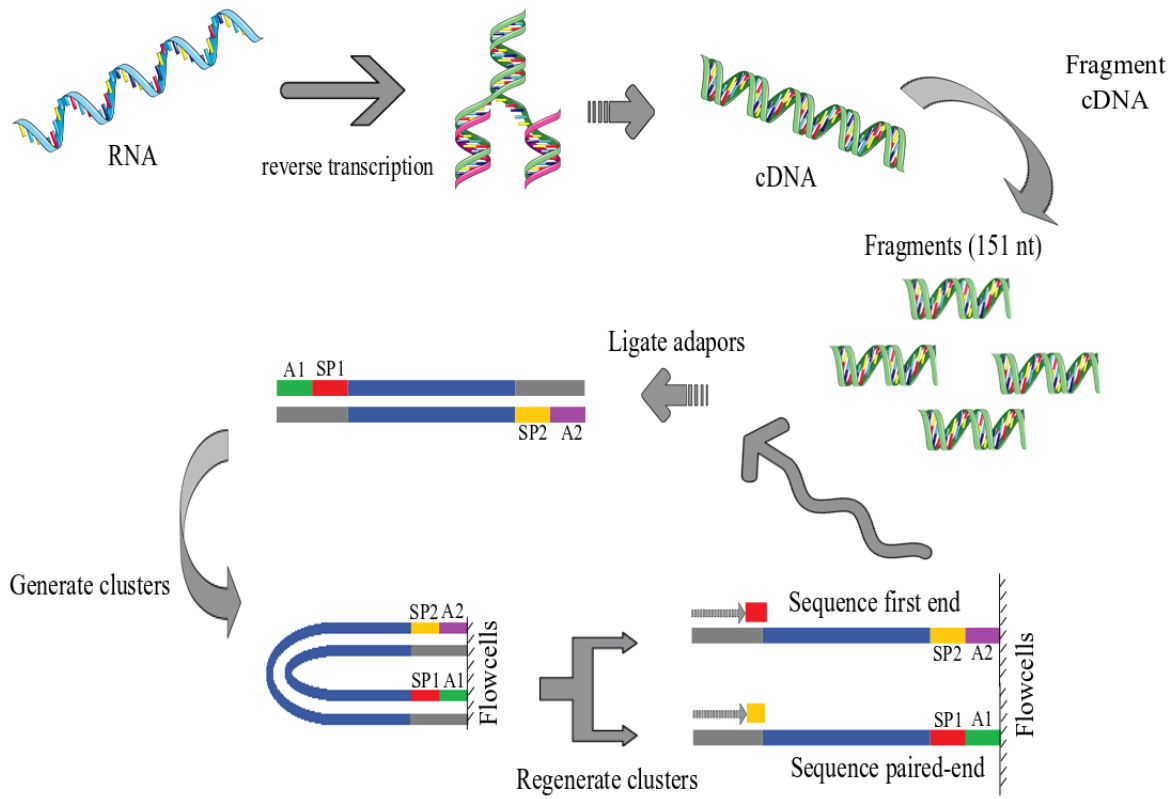


Figure 1: Paired-End sequencing protocol applied on polyA RNA sequences from coral samples of the *Tara Pacific* expedition.

**APPENDIX III : Detection of the contaminations in Symbiodiniaceae  
references**

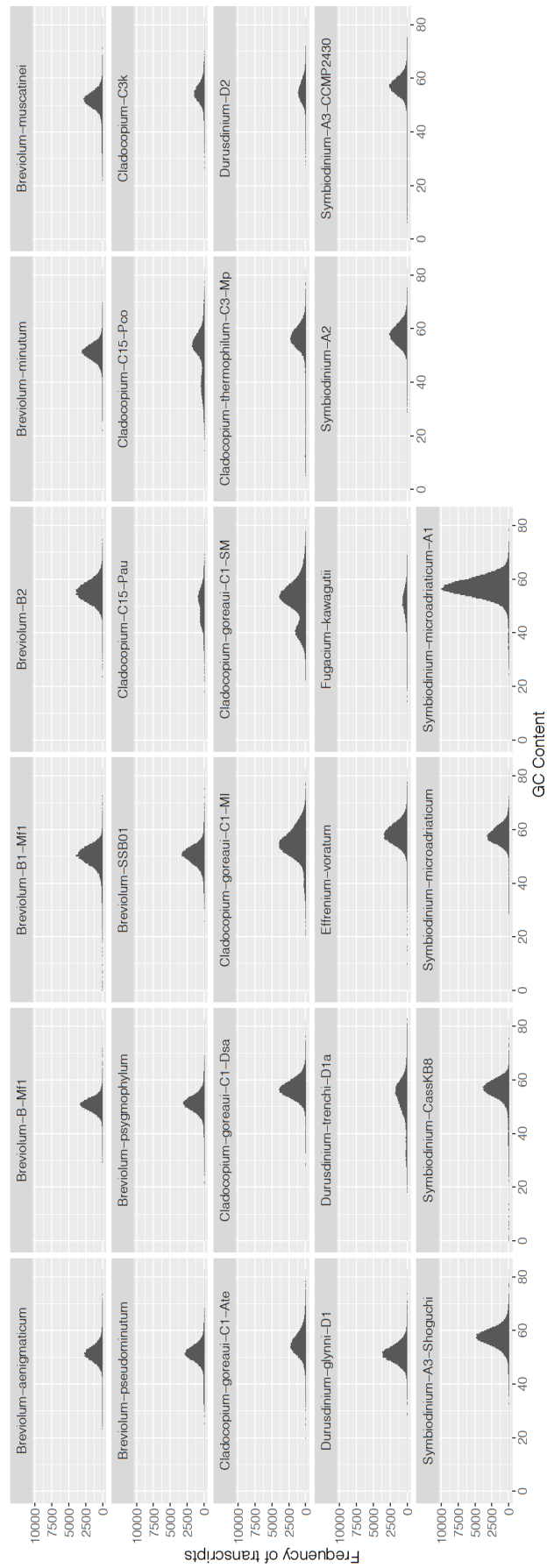


Figure 2: GC content of each Symbiodiniaceae reference transcriptomes selected and studied during the thesis.

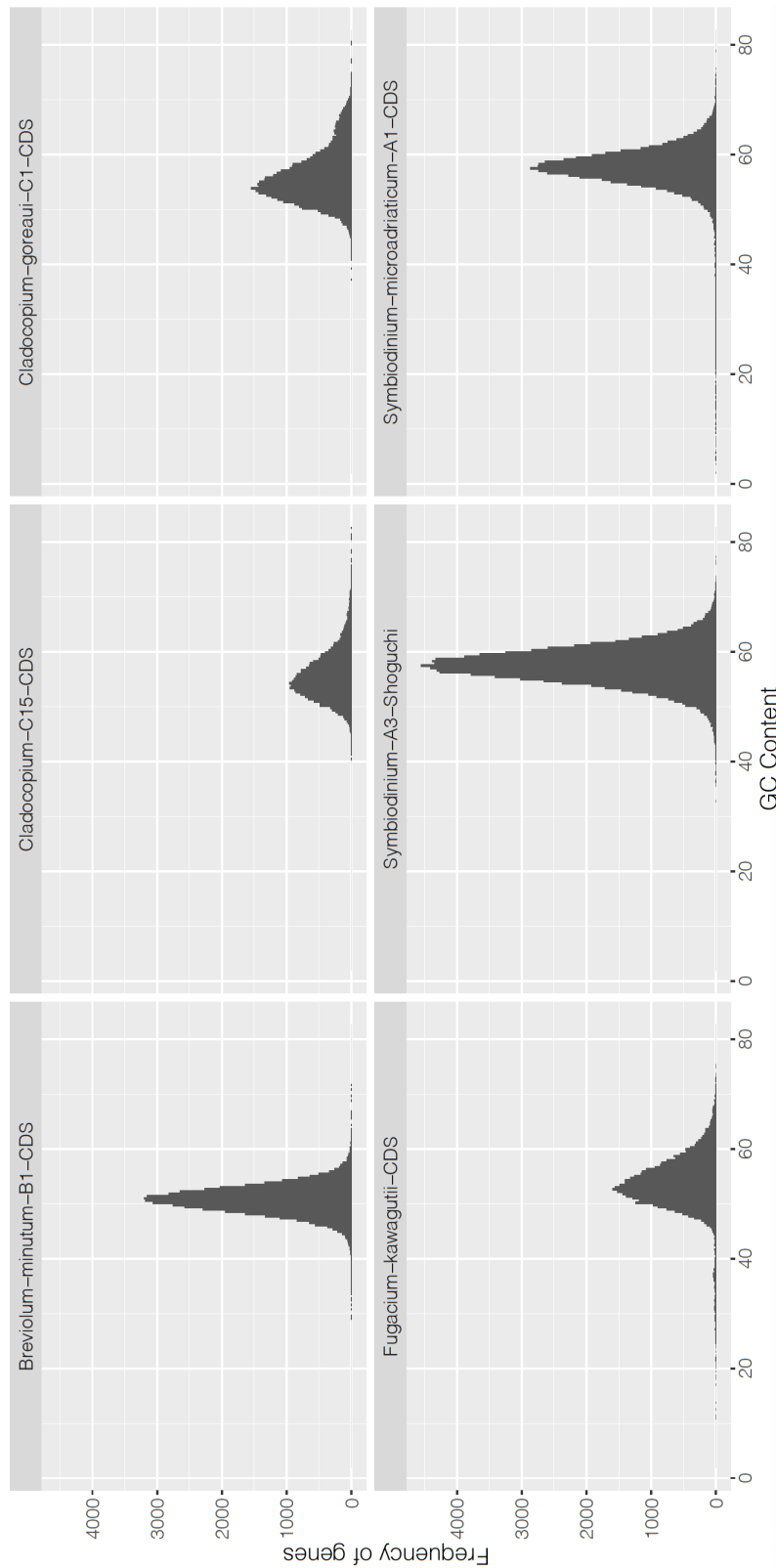


Figure 3: GC content of each Symbiodiniaceae reference genomes selected and studied during the thesis.

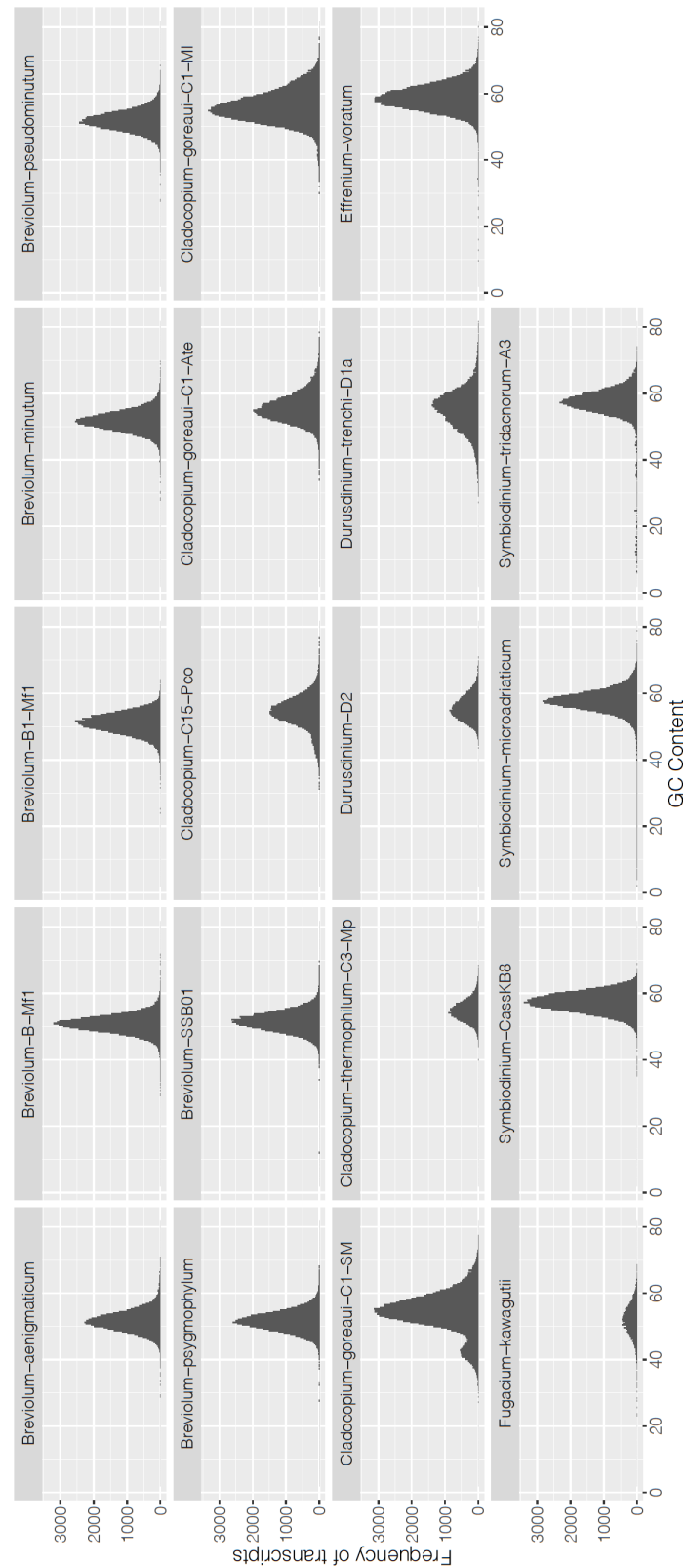


Figure 4: GC Content of 19 Symbiodiniaceae reference transcriptomes shared and studied with **Gonzales et al.** in 2017.



**APPENDIX IV : Estimating Symbiodiniaceae diversity in the *Tara* Pacific coral samples using all Symbiodiniaceae reference transcriptomes**

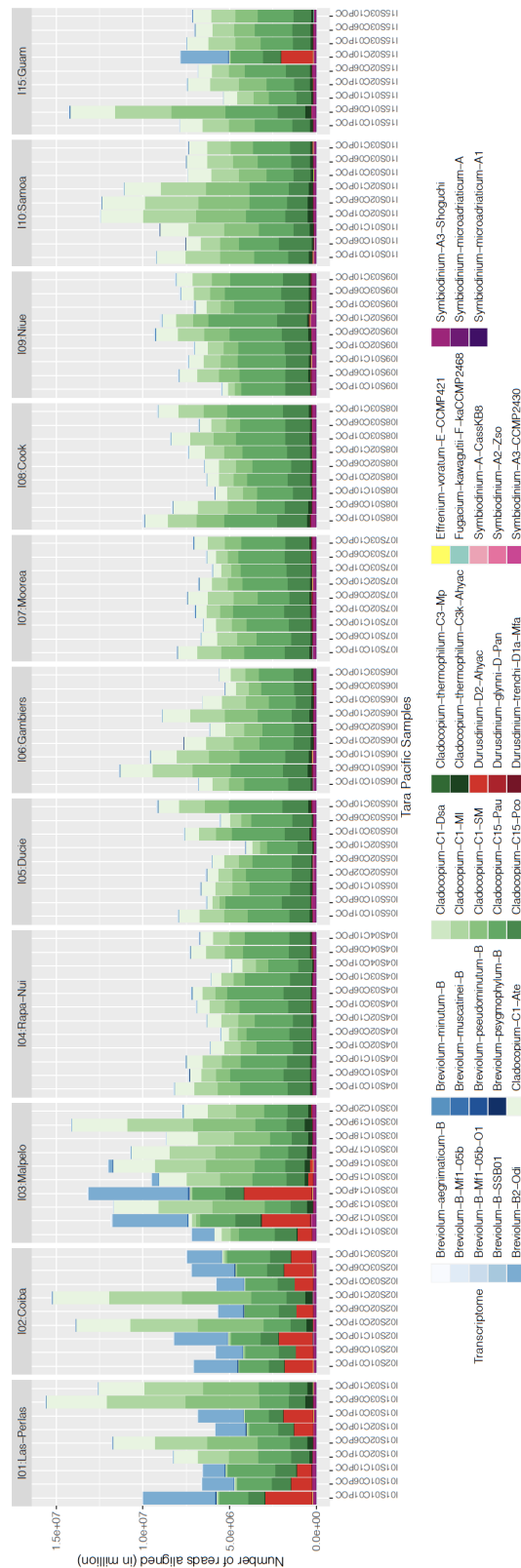


Figure 5: Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes filtered at 70% identity and 30% coverage.

Each color corresponds to a specific transcriptome. Similar colors are transcriptomes of the same Symbiodiniaceae genus in *Pocillopora sp.* samples.

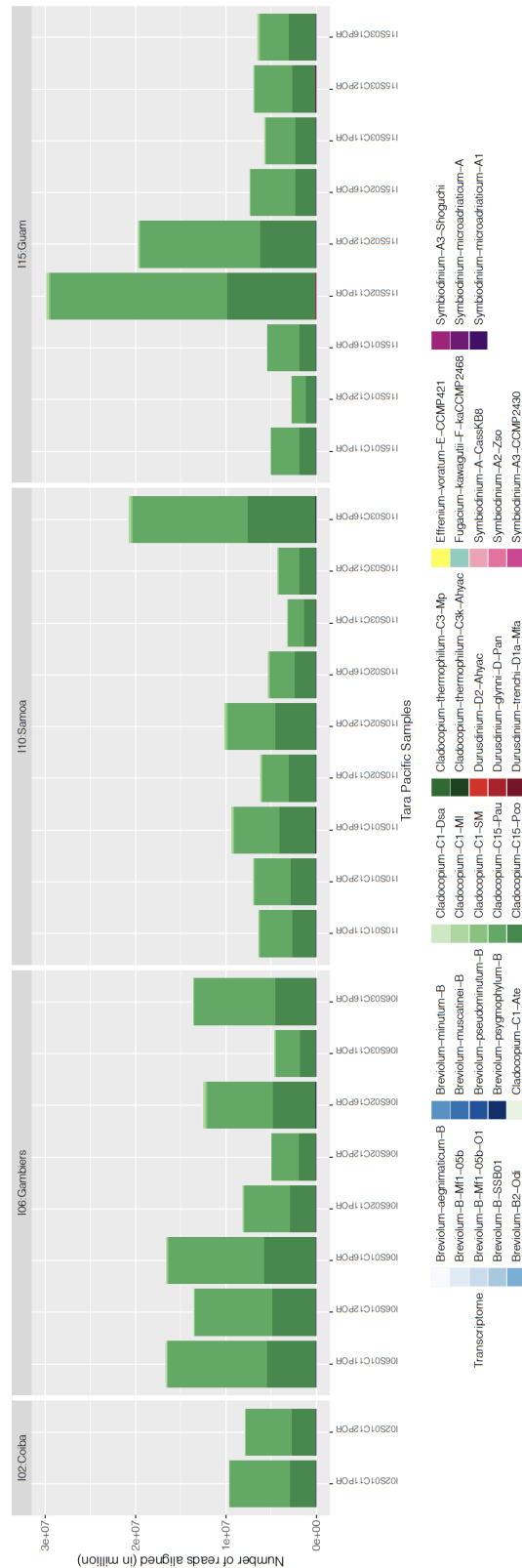


Figure 6: Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes filtered at 70% identity and 30% coverage.

Each color corresponds to a specific transcriptome. Similar colors are transcriptomes of the same Symbiodiniaceae genus in *Porites sp.* samples.

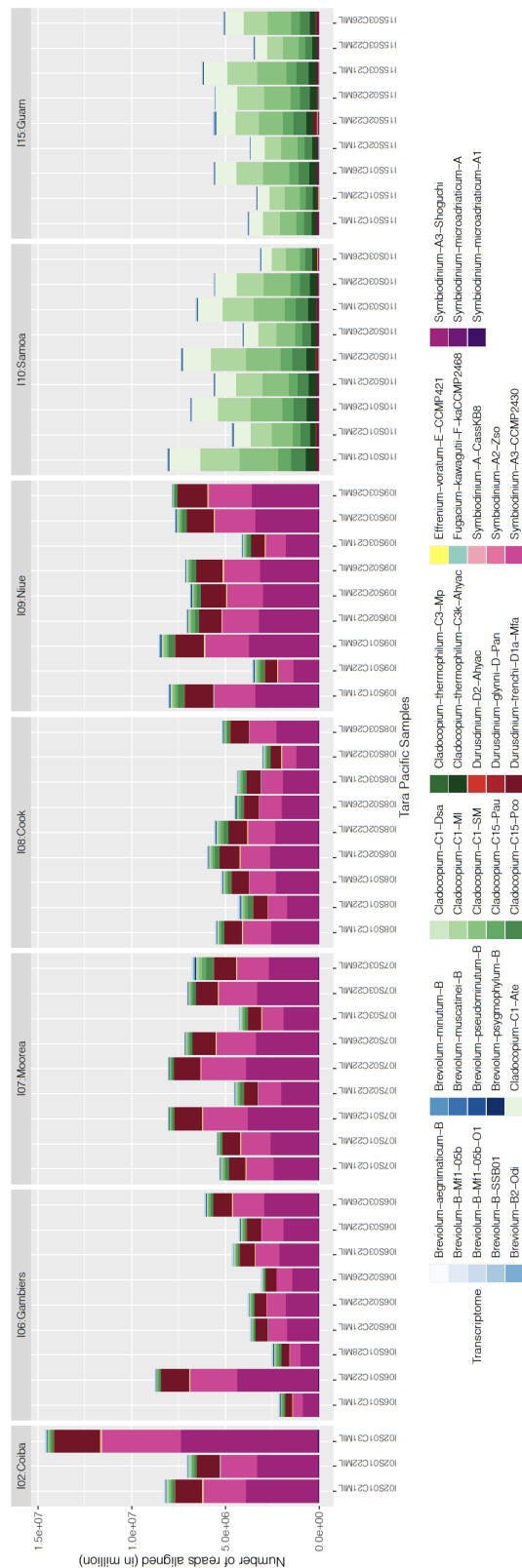


Figure 7: Barplots of the number of reads aligned on Symbiodiniaceae transcriptomes filtered at 70% identity and 30% coverage.

Each color corresponds to a specific transcriptome. Similar colors are transcriptomes of the same Symbiodiniaceae genus in *Millepora sp.* samples.

## APPENDIX V : Estimating Symbiodiniaceae diversity in the Tara Pacific coral samples using all Symbiodiniaceae reference transcriptomes

Table 1: Some calculations on *Millepora sp.* reads aligned on Symbiodiniaceae reference transcriptomes.

Count reads on each reference transcriptomes aligned, Mean of identity percent for transcripts aligned and Mean of coverage percent for transcripts aligned. Calculs were ordered by decreasing count reads.

Transcriptomes	Coral	Count Reads	Identity (%)	Coverage (%)
Symbiodinium-A3-Shoguchi	Millepora	99 606 660	99,64	53,75
Symbiodinium-A3-CCMP2430	Millepora	62 083 843	99,70	73,50
Durusdinium-trenchi-D1a-Mfa	Millepora	35 860 730	99,70	78,65
Cladocopium-C1-MI	Millepora	15 146 706	97,24	40,33
Cladocopium-C1-SM	Millepora	14 706 214	97,22	40,04
Cladocopium-C1-Ate	Millepora	12 810 528	97,35	45,75
Cladocopium-C15-Pco	Millepora	4 961 025	97,03	31,29
Cladocopium-thermophilum-C3k-Ahyac	Millepora	4 938 217	97,34	56,47
Cladocopium-C15-Pau	Millepora	4 704 919	97,12	18,23
Symbiodinium-A-CassKB8	Millepora	1 011 572	98,41	25,88
Symbiodinium-microadriaticum-A	Millepora	487 452	98,30	15,15
Effrenium-voratum-E-CCMP421	Millepora	379 134	98,47	22,03
Breviolum-B-Mf1-05b	Millepora	249 816	98,89	31,95
Symbiodinium-microadriaticum-A1	Millepora	196 862	97,97	16,73
Cladocopium-thermophilum-C3-Mp	Millepora	87 260	98,50	26,85
Breviolum-B2-Odi	Millepora	73 252	99,43	33,07
Cladocopium-C1-Dsa	Millepora	49 465	98,36	36,50
Durusdinium-D2-Ahyac	Millepora	35 253	99,49	22,01
Breviolum-minutum-B	Millepora	27 276	98,11	33,29
Breviolum-B-Mf1-05b-O1	Millepora	26 583	96,83	12,10
Breviolum-psygmophylum-B	Millepora	26 520	98,14	24,00
Breviolum-pseudominutum-B	Millepora	22 062	98,27	28,74
Symbiodinium-A2-Zso	Millepora	15 490	98,50	29,39
Breviolum-aegnimaticum-B	Millepora	13 324	97,89	20,51
Durusdinium-glynni-D-Pan	Millepora	12 939	98,08	33,41
Breviolum-B-SSB01	Millepora	11 420	97,65	20,75
Breviolum-muscatinei-B	Millepora	7 753	97,41	14,22
Fugacium-kawagutii-F-kaCCMP2468	Millepora	2 243	96,71	54,71

Table 2: Some calculations on *Porites sp.* reads aligned on Symbiodiniaceae reference transcriptomes.

Count reads on each reference transcriptomes aligned, Mean of identity percent for transcripts aligned and Mean of coverage percent for transcripts aligned. Calculs were ordered by decreasing count reads.

Transcriptomes	Coral	Count Reads	Identity (%)	Coverage (%)
Cladocopium-C15-Pau	Porites	140 301 867	99,20	58,10
Cladocopium-C15-Pco	Porites	88 697 193	99,31	75,04
Cladocopium-C1-MI	Porites	1 739 079	97,12	21,26
Cladocopium-C1-SM	Porites	1 641 876	97,12	21,52
Cladocopium-C1-Ate	Porites	984 806	97,42	23,28
Cladocopium-thermophilum-C3k-Ahyac	Porites	181 149	98,10	36,69
Symbiodinium-A3-Shoguchi	Porites	36 262	99,29	15,08
Breviolum-psygmophylum-B	Porites	15 939	98,70	21,36
Durusdinium-D2-Ahyac	Porites	12 148	99,24	24,61
Breviolum-pseudominutum-B	Porites	11 573	98,97	25,45
Breviolum-B-Mf1-05b	Porites	11 565	99,07	24,49
Breviolum-B2-Odi	Porites	7 221	99,35	29,92
Breviolum-minutum-B	Porites	7 135	98,83	18,46
Breviolum-aegnimaticum-B	Porites	6 840	98,78	23,36
Durusdinium-glynni-D-Pan	Porites	6 521	98,91	34,14
Symbiodinium-A3-CCMP2430	Porites	4 212	99,29	18,16
Durusdinium-trenchi-D1a-Mfa	Porites	4 204	98,89	21,83
Breviolum-B-SSB01	Porites	3 911	98,36	11,68
Breviolum-B-Mf1-05b-O1	Porites	2 463	97,40	10,53
Cladocopium-thermophilum-C3-Mp	Porites	2 331	98,59	22,04
Effrenium-voratum-E-CCMP421	Porites	1 342	97,97	12,69
Breviolum-muscatinei-B	Porites	1 120	98,12	10,26
Fugacium-kawagutii-F-kaCCMP2468	Porites	825	97,01	47,15
Cladocopium-C1-Dsa	Porites	774	97,68	24,31
Symbiodinium-A2-Zso	Porites	631	97,60	21,89
Symbiodinium-A-CassKB8	Porites	625	98,09	15,29
Symbiodinium-microadriaticum-A1	Porites	188	97,54	14,13
Symbiodinium-microadriaticum-A	Porites	150	97,59	9,59

Table 3: Some calculations on *Pocillopora sp.* reads aligned on Symbiodiniaceae reference transcriptomes.

Count reads on each reference transcriptomes aligned, Mean of identity percent for transcripts aligned and Mean of coverage percent for transcripts aligned. Calculs were ordered by decreasing count reads.

Transcriptomes	Coral	Count Reads	Identity (%)	Coverage (%)
Cladocopium-C1-MI	Pocillopora	126 608 644	99,69	60,46
Cladocopium-C1-SM	Pocillopora	121 937 132	99,69	59,14
Cladocopium-C1-Ate	Pocillopora	103 876 642	99,74	68,22
Breviolum-B2-Odi	Pocillopora	37 561 817	99,62	83,09
Durusdinium-D2-Ahyac	Pocillopora	25 528 925	99,57	78,71
Cladocopium-thermophilum-C3k-Ahyac	Pocillopora	13 272 856	99,70	54,20
Cladocopium-C15-Pco	Pocillopora	5 381 821	98,53	31,94
Symbiodinium-A3-Shoguchi	Pocillopora	5 172 188	99,59	14,34
Cladocopium-C15-Pau	Pocillopora	4 941 833	98,24	12,66
Breviolum-B-Mf1-05b	Pocillopora	335 117	99,06	27,04
Symbiodinium-A3-CCMP2430	Pocillopora	198 132	99,60	17,05
Durusdinium-trenchi-D1a-Mfa	Pocillopora	110 353	99,55	22,31
Breviolum-psygmophylum-B	Pocillopora	72 784	98,63	24,69
Breviolum-pseudominutum-B	Pocillopora	72 078	98,75	29,82
Breviolum-minutum-B	Pocillopora	48 427	98,66	21,62
Breviolum-aegnimaticum-B	Pocillopora	45 708	98,45	23,33
Durusdinium-glynni-D-Pan	Pocillopora	39 475	99,10	35,07
Breviolum-B-SSB01	Pocillopora	26 571	97,84	15,32
Breviolum-B-Mf1-05b-O1	Pocillopora	24 579	97,25	11,51
Cladocopium-C1-Dsa	Pocillopora	22 368	97,88	16,88
Breviolum-muscatinei-B	Pocillopora	13 663	97,37	10,29
Cladocopium-thermophilum-C3-Mp	Pocillopora	11 585	98,06	19,27
Symbiodinium-A-CassKB8	Pocillopora	9 884	98,54	18,42
Effrenium-voratum-E-CCMP421	Pocillopora	6 802	98,07	15,72
Symbiodinium-microadriaticum-A	Pocillopora	4 321	98,55	16,18
Symbiodinium-microadriaticum-A1	Pocillopora	2 877	97,96	14,23
Symbiodinium-A2-Zso	Pocillopora	1 494	97,61	17,81
Fugacium-kawagutii-F-kaCCMP2468	Pocillopora	514	97,22	40,00

## APPENDIX VI : Estimating Symbiodiniaceae diversity in the Tara Pacific coral samples using all Symbiodiniaceae reference transcriptomes

Table 4: Some calculs on dual-transcriptomic reads aligned on specific Symbiodiniaceae reference transcriptomes or genomes.

Count reads on each reference transcriptomes or genomes aligned, Mean of identity percent for transcripts or CDs aligned and Mean of coverage percent for transcripts or CDS aligned. Calculs were ordered by decreasing count reads.

Transcriptomes	Islands	Corals	Count Reads	Identity	Coverage
Symbiodinium-A3-Predicted-Transcripts-From-Predicted-Genes	I02:Coiba	Millepora	17 648 802	99,59	66,69
Symbiodinium-A3-Predicted-Transcripts-From-Predicted-Genes	I06:Gambiers	Millepora	23 448 861	99,69	58,10
Symbiodinium-A3-Predicted-Transcripts-From-Predicted-Genes	I07:Moorea	Millepora	34 640 487	99,73	63,59
Symbiodinium-A3-Predicted-Transcripts-From-Predicted-Genes	I08:Cook	Millepora	25 321 643	99,74	59,45
Symbiodinium-A3-Predicted-Transcripts-From-Predicted-Genes	I09:Niue	Millepora	34 973 134	99,73	61,21
Cladocopium-thermophilum-C3k-Ahyac	I10:Samoa	Millepora	4 971 199	98,63	56,50
Symbiodinium-A3-Predicted-Transcripts-From-Predicted-Genes	I10:Samoa	Millepora	7 003	99,55	12,02
Cladocopium-thermophilum-C3k-Ahyac	I15:Guam	Millepora	4 044 272	98,62	55,86
Symbiodinium-A3-Predicted-Transcripts-From-Predicted-Genes	I15:Guam	Millepora	10 696	99,61	12,31
Cladocopium-goreau-C1-CDS	I01:Las-Perlas	Pocillopora	16 315 576	99,63	70,48
Durusdinium-D2-Ahyac	I01:Las-Perlas	Pocillopora	12 017 759	99,69	85,48
Durusdinium-D2-Ahyac	I02:Coiba	Pocillopora	14 951 979	99,69	87,60
Cladocopium-goreau-C1-CDS	I02:Coiba	Pocillopora	10 942 577	99,63	61,58
Cladocopium-goreau-C1-CDS	I03:Malpelo	Pocillopora	27 582 080	99,76	72,72
Durusdinium-D2-Ahyac	I03:Malpelo	Pocillopora	14 279 736	99,83	76,62
Cladocopium-goreau-C1-CDS	I04:Rapa-Nui	Pocillopora	15 856 610	99,78	69,92
Durusdinium-D2-Ahyac	I04:Rapa-Nui	Pocillopora	56 401	99,79	22,02
Cladocopium-goreau-C1-CDS	I05:Ducie	Pocillopora	11 685 443	99,76	69,17
Durusdinium-D2-Ahyac	I05:Ducie	Pocillopora	25 473	99,79	21,69
Cladocopium-goreau-C1-CDS	I06:Gambiers	Pocillopora	17 395 397	99,70	75,03
Durusdinium-D2-Ahyac	I06:Gambiers	Pocillopora	25 063	98,77	52,15
Cladocopium-goreau-C1-CDS	I07:Moorea	Pocillopora	12 459 765	99,76	72,05
Durusdinium-D2-Ahyac	I07:Moorea	Pocillopora	29 503	99,79	21,92
Cladocopium-goreau-C1-CDS	I08:Cook	Pocillopora	15 551 718	99,77	73,48
Durusdinium-D2-Ahyac	I08:Cook	Pocillopora	18 130	98,73	52,70
Cladocopium-goreau-C1-CDS	I09:Niue	Pocillopora	13 727 938	99,79	70,76
Durusdinium-D2-Ahyac	I09:Niue	Pocillopora	16 925	98,74	55,73
Cladocopium-goreau-C1-CDS	I10:Samoa	Pocillopora	25 478 530	99,70	78,19
Durusdinium-D2-Ahyac	I10:Samoa	Pocillopora	35 984	99,37	32,98
Cladocopium-goreau-C1-CDS	I15:Guam	Pocillopora	17 338 502	99,70	75,37
Durusdinium-D2-Ahyac	I15:Guam	Pocillopora	2 893 967	99,76	69,02
Cladocopium-C15-CDS	I02:Coiba	Porites	4 15 179	99,44	62,08
Cladocopium-C15-CDS	I06:Gambiers	Porites	4 124 860	99,70	71,95
Cladocopium-C15-CDS	I10:Samoa	Porites	5 871 009	99,39	72,84
Cladocopium-C15-CDS	I15:Guam	Porites	3 668 988	99,42	65,48



## APPENDIX VII : Revealing sub-variants of the ITS2 sequences in *Cladocopium C1*

Table 5: Number of variations and type of mutations in different *Cladocopium* ITS2 sequences

Clade1	Clade2	ID-Seq1	ID-Seq2	Length-Seq1	Length-Seq2	Mutation	Number-Of-Nucleotid-Different
C	C	C1	C116	283	283	Substitution	6
C	C	C1	C15	283	283	Substitution	5
C	C	C1	C1b	283	282	Indel	1
C	C	C1	C1d	283	283	Substitution	3
C	C	C1	C1m	283	283	Substitution	2
C	C	C1	C3	283	283	Substitution	1
C	C	C1	C3	283	234	Indel	49
C	C	C1	C42 (type 1)	283	283	Substitution	3
C	C	C1	C42 (type 2)	283	283	Substitution	2
C	C	C116	C1	283	283	Substitution	6
C	C	C116	C15	283	283	Substitution	1
C	C	C116	C1b	283	282	Indel	1
C	C	C116	C1d	283	283	Substitution	9
C	C	C116	C1m	283	283	Substitution	8
C	C	C116	C3	283	283	Substitution	5
C	C	C116	C3	283	234	Indel	49
C	C	C116	C42 (type 1)	283	283	Substitution	9
C	C	C116	C42 (type 2)	283	283	Substitution	8
C	C	C15	C1	283	283	Substitution	5
C	C	C15	C116	283	283	Substitution	1
C	C	C15	C1b	283	282	Indel	1
C	C	C15	C1d	283	283	Substitution	8
C	C	C15	C1m	283	283	Substitution	7
C	C	C15	C3	283	283	Substitution	4
C	C	C15	C3	283	234	Indel	49
C	C	C15	C42 (type 1)	283	283	Substitution	8
C	C	C15	C42 (type 2)	283	283	Substitution	7
C	C	C1b	C1	282	283	Indel	1
C	C	C1b	C116	282	283	Indel	1
C	C	C1b	C15	282	283	Indel	1
C	C	C1b	C1d	282	283	Indel	1
C	C	C1b	C1m	282	283	Indel	1
C	C	C1b	C3	282	283	Indel	1
C	C	C1b	C3	282	234	Indel	48
C	C	C1b	C42 (type 1)	282	283	Indel	1
C	C	C1b	C42 (type 2)	282	283	Indel	1
C	C	C1d	C1	283	283	Substitution	3
C	C	C1d	C116	283	283	Substitution	9
C	C	C1d	C15	283	283	Substitution	8
C	C	C1d	C1b	283	282	Indel	1
C	C	C1d	C1m	283	283	Substitution	5
C	C	C1d	C3	283	283	Substitution	4
C	C	C1d	C3	283	234	Indel	49
C	C	C1d	C42 (type 1)	283	283	Substitution	2
C	C	C1d	C42 (type 2)	283	283	Substitution	1
C	C	C1m	C1	283	283	Substitution	2
C	C	C1m	C116	283	283	Substitution	8
C	C	C1m	C15	283	283	Substitution	7
C	C	C1m	C1b	283	282	Indel	1
C	C	C1m	C1d	283	283	Substitution	5
C	C	C1m	C3	283	283	Substitution	3
C	C	C1m	C3	283	234	Indel	49
C	C	C1m	C42 (type 1)	283	283	Substitution	5
C	C	C1m	C42 (type 2)	283	283	Substitution	4

Clade1	Clade2	ID-Seq1	ID-Seq2	Length-Seq1	Length-Seq2	Mutation	Number-Of-Nucleotid-Different
C	C	C3	C1	283	283	Substitution	1
C	C	C3	C116	283	283	Substitution	5
C	C	C3	C15	283	283	Substitution	4
C	C	C3	C1b	283	282	Indel	1
C	C	C3	C1d	283	283	Substitution	4
C	C	C3	C1m	283	283	Substitution	3
C	C	C3	C3	283	234	Indel	49
C	C	C3	C42 (type 1)	283	283	Substitution	4
C	C	C3	C42 (type 2)	283	283	Substitution	3
C	C	C3	C1	234	283	Indel	49
C	C	C3	C116	234	283	Indel	49
C	C	C3	C15	234	283	Indel	49
C	C	C3	C1b	234	282	Indel	48
C	C	C3	C1d	234	283	Indel	49
C	C	C3	C1m	234	283	Indel	49
C	C	C3	C3	234	283	Indel	49
C	C	C3	C42 (type 1)	234	283	Indel	49
C	C	C3	C42 (type 2)	234	283	Indel	49
C	C	C42 (type 1)	C1	283	283	Substitution	3
C	C	C42 (type 1)	C116	283	283	Substitution	9
C	C	C42 (type 1)	C15	283	283	Substitution	8
C	C	C42 (type 1)	C1b	283	282	Indel	1
C	C	C42 (type 1)	C1d	283	283	Substitution	2
C	C	C42 (type 1)	C1m	283	283	Substitution	5
C	C	C42 (type 1)	C3	283	283	Substitution	4
C	C	C42 (type 1)	C3	283	234	Indel	49
C	C	C42 (type 1)	C42 (type 2)	283	283	Substitution	1
C	C	C42 (type 2)	C1	283	283	Substitution	2
C	C	C42 (type 2)	C116	283	283	Substitution	8
C	C	C42 (type 2)	C15	283	283	Substitution	7
C	C	C42 (type 2)	C1b	283	282	Indel	1
C	C	C42 (type 2)	C1d	283	283	Substitution	1
C	C	C42 (type 2)	C1m	283	283	Substitution	4
C	C	C42 (type 2)	C3	283	283	Substitution	3
C	C	C42 (type 2)	C3	283	234	Indel	49
C	C	C42 (type 2)	C42 (type 1)	283	283	Substitution	1

## APPENDIX VIII: Genomic vertical coverage to identify the population structure of *Cladocopium C1*

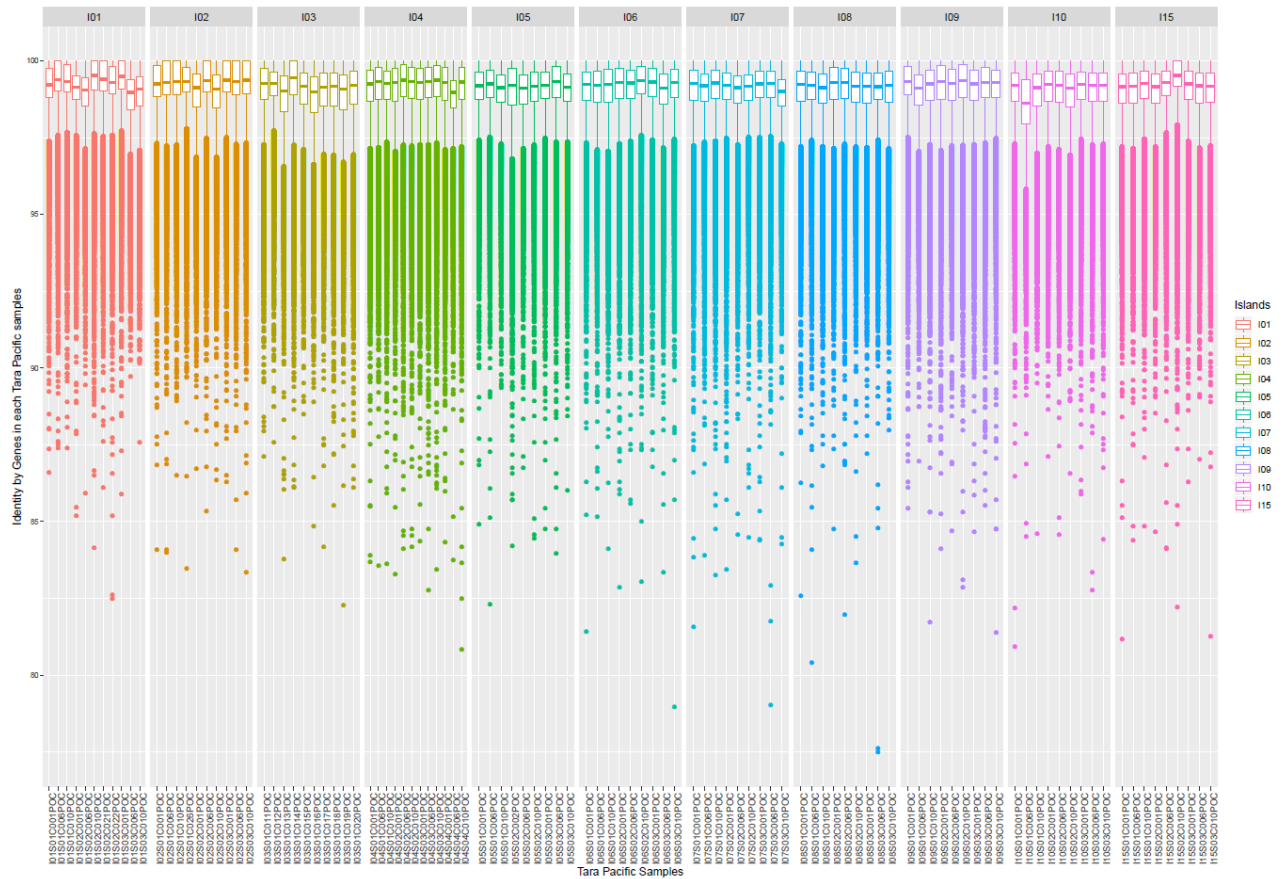


Figure 8: Boxplots of mean identity by contig from *Cladocopium C1* reference genome aligned on dual-genomic samples from *Pocillopora* samples on the first 11 islands.

## APPENDIX IX: Population structure of *Cladocopium C1* using meta-transcriptomic samples

Table 6: Comparison between ITS2 profiles and population structure of *Pocillopora sp.* and *Cladocopium C1*, obtained using dual-genomic and dual-transcriptomic respectively, in each sample

Sample	ITS2 profiles	Pocillopora SVD	Population C1
I01S01C001	D1	SVD4	NA
I01S01C006	D1	SVD4	NA
I01S01C010	D1	SVD4	NA
I01S02C001	NA	SVD4	Group5
I01S02C006	C42g	SVD4	Group5
I01S02C010	D1	SVD4	NA
I01S03C006	C42g	SVD4	Group5
I01S03C010	C42a	SVD4	Group5
I02S01C001	D1	SVD4	NA
I02S01C006	D1	SVD4	NA
I02S01C010	D1	SVD4	NA
I02S02C001	C42g	SVD4	Group5
I02S02C006	D1	SVD4	NA
I02S02C010	C42g	SVD4	Group5
I02S03C001	D1	SVD4	NA
I02S03C006	D1	SVD4	NA
I02S03C010	D1	SVD4	NA
I03S01C011	C42a	SVD4	NA
I03S01C012	D1	SVD4	NA
I03S01C013	C42a	SVD4	Group5
I03S01C014	D1	SVD4	NA
I03S01C015	C42a	SVD4	NA
I03S01C016	C42a	SVD4	NA
I03S01C017	C42a	SVD4	Group5
I03S01C018	C42a	SVD4	Group5
I03S01C019	C42a	SVD4	Group5
I03S01C020	C42a	SVD4	NA
I04S01C001	C1cf	SVD5	Group1
I04S01C006	C1cf	SVD5	Group1
I04S01C010	C1cf	SVD5	Group1
I04S02C001	C1cf	SVD5	Group1
I04S02C006	C1cf	SVD5	Group1
I04S02C010	C1cf	SVD5	Group1
I04S03C001	C1cf	SVD5	Group1
I04S03C006	C1cf	SVD5	Group1
I04S03C010	C1cf	SVD5	Group1
I04S04C001	C1cf	SVD5	Group1
I04S04C006	C1cf	SVD5	Group1
I04S04C010	C1cf	SVD5	Group1
I05S01C001	C42.2	SVD4	Group4
I05S01C006	C42.2	SVD4	Group4
I05S01C010	C42.2	SVD1	Group2
I05S02C002	C42.2	SVD1	NA
I05S02C006	C42.2	SVD1	Group2
I05S02C010	C42.2	SVD1	Group2
I05S03C001	C42.2	SVD5	Group3
I05S03C006	C42.2	SVD5	Group3
I05S03C010	C42.2	SVD4	Group4

Sample	ITS2 profiles	Pocillopora SVD	Population C1
I06S01C001	C42.2	SVD1	Group2
I06S01C006	C42.2	SVD1	Group2
I06S01C010	C42.2	SVD4	Group4
I06S02C001	C42.2	SVD1	Group2
I06S02C006	C42.2	SVD1	Group2
I06S02C010	C42.2	SVD4	Group4
I06S03C001	C42.2	SVD1	Group2
I06S03C006	C42.2	SVD1	Group2
I06S03C010	C42.2	SVD1	Group2
I07S01C001	C1d	SVD5	Group3
I07S01C006	C1d	SVD5	Group3
I07S01C010	C1d	SVD5	Group3
I07S02C001	C1d	SVD5	Group3
I07S02C006	C1d	SVD5	Group3
I07S02C010	C42g	SVD2	Group5
I07S03C001	NA	NA	Group3
I07S03C006	C1d	SVD5	Group3
I07S03C010	C42.2	SVD1	Group2
I08S01C001	C42.2	SVD3	Group3
I08S01C006	C42g	SVD2	Group5
I08S01C010	C42.2	SVD3	Group3
I08S02C001	C1d	SVD3	Group3
I08S02C006	C42g	SVD2	Group5
I08S02C010	C42g	SVD2	Group5
I08S03C001	C42.2	SVD3	Group3
I08S03C006	C42g	SVD2	Group5
I08S03C010	C42.2	SVD3	Group3
I09S01C001	C42.2	SVD3	Group3
I09S01C006	C42g	SVD2	Group5
I09S01C010	C42g	SVD2	Group5
I09S02C001	NA	NA	Group3
I09S02C006	C42.2	SVD3	Group3
I09S02C010	C42.2	SVD3	Group3
I09S03C001	C42g	SVD2	Group5
I09S03C006	C1d	SVD3	Group3
I09S03C010	C42g	NA	Group5
I10S01C001	C42g	SVD2	Group5
I10S01C006	C1d	SVD3	NA
I10S01C010	C42g	SVD2	Group5
I10S02C001	C42g	SVD2	Group5
I10S02C006	C42g	SVD2	Group5
I10S02C010	C1d	SVD3	Group3
I10S03C001	C42g	SVD2	Group5
I10S03C006	C42g	SVD2	Group5
I10S03C010	C42g	SVD2	Group5
I15S01C001	C42g	SVD2	Group5
I15S01C006	C1d	SVD3	Group3
I15S01C010	C1d	SVD3	Group3
I15S02C001	C1d	SVD3	Group3
I15S02C006	C42g	SVD2	Group5
I15S02C010	D1	SVD3	NA
I15S03C001	C1d	SVD3	Group3
I15S03C006	C42g	SVD2	Group5
I15S03C010	C1d	SVD3	Group3

## APPENDIX X: Differentially expressed functions of *Cladocopium C1* in *Pocillopora sp.* samples across the Pacific Ocean

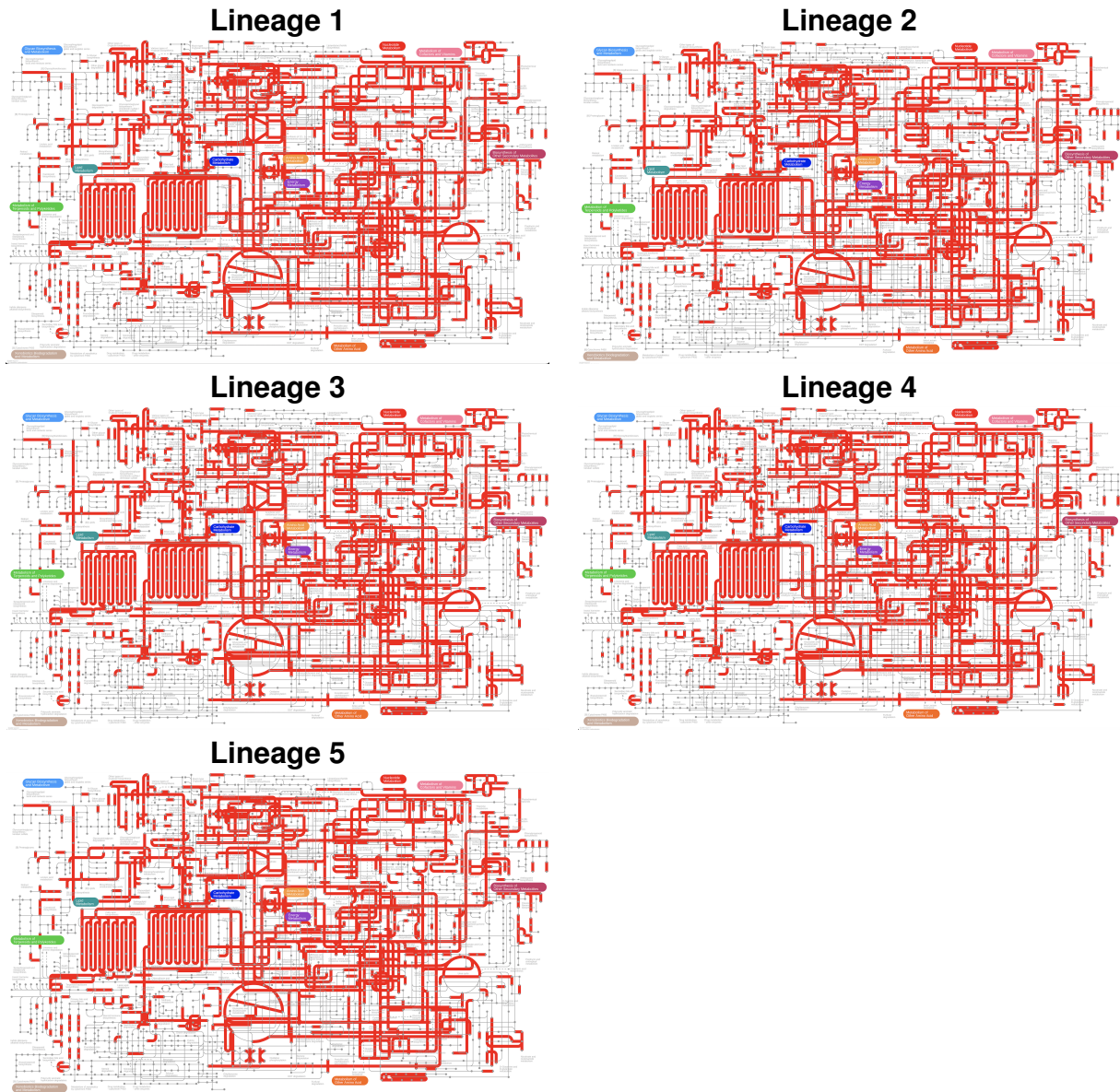


Figure 9: KEGG enriched in each *Cladocopium C1* lineages aligned on dual-transcriptomic samples, using iPath3 visualization.

## APPENDIX XI: Differentially expressed functions of *Cladocopium C1* in *Pocillopora sp.* samples across the Pacific Ocean

PFAM	x	k	m	n	pvalue	odds ratio	padj	PFAM definition	PFAM description
PF00520	64	2003	309	23951	7,42E-12	2,97E+00	5,11E-09	Ion transport protein	G2vsG5-Up
PF00520	64	2003	309	23951	7,42E-12	2,97E+00	5,11E-09	Ion transport protein	G5vsG2-Down
PF07727	9	4744	208	24052	1,87E-10	1,85E-01	2,13E-07	Reverse transcriptase (RNA-dependent DNA polymerase)	G1vsG2-Up
PF07727	9	4744	208	24052	1,87E-10	1,85E-01	2,13E-07	Reverse transcriptase (RNA-dependent DNA polymerase)	G2vsG1-Down
PF07727	15	5243	208	24052	2,33E-08	2,80E-01	2,81E-05	Reverse transcriptase (RNA-dependent DNA polymerase)	G1vsG5-Up
PF07727	15	5243	208	24052	2,33E-08	2,80E-01	2,81E-05	Reverse transcriptase (RNA-dependent DNA polymerase)	G5vsG1-Down
PF07727	13	4722	208	24052	8,44E-08	2,74E-01	9,72E-05	Reverse transcriptase (RNA-dependent DNA polymerase)	G1vsG3-Up
PF07727	13	4722	208	24052	8,44E-08	2,74E-01	9,72E-05	Reverse transcriptase (RNA-dependent DNA polymerase)	G3vsG1-Down
PF07727	12	4546	208	24052	8,79E-08	2,64E-01	0,00011184	Reverse transcriptase (RNA-dependent DNA polymerase)	G1vsG5-Down
PF07727	12	4546	208	24052	8,79E-08	2,64E-01	0,00011184	Reverse transcriptase (RNA-dependent DNA polymerase)	G5vsG1-Up
PF07727	7	3312	208	24052	7,73E-07	2,19E-01	0,00071768	Reverse transcriptase (RNA-dependent DNA polymerase)	G1vsG4-Up
PF07727	7	3312	208	24052	7,73E-07	2,19E-01	0,00071768	Reverse transcriptase (RNA-dependent DNA polymerase)	G4vsG1-Down
PF00504	20	2090	70	24190	1,11E-06	4,27E+00	0,00079641	Chlorophyll A-B binding protein	G2vsG5-Down
PF00504	20	2090	70	24190	1,11E-06	4,27E+00	0,00079641	Chlorophyll A-B binding protein	G5vsG2-Up
PF07727	2	2090	208	24052	2,58E-06	1,02E-01	0,0009277	Reverse transcriptase (RNA-dependent DNA polymerase)	G2vsG5-Down
PF07727	2	2090	208	24052	2,58E-06	1,02E-01	0,0009277	Reverse transcriptase (RNA-dependent DNA polymerase)	G5vsG2-Up
PF00504	25	3312	70	24190	2,75E-06	3,53E+00	0,00127571	Chlorophyll A-B binding protein	G1vsG4-Up
PF00504	25	3312	70	24190	2,75E-06	3,53E+00	0,00127571	Chlorophyll A-B binding protein	G4vsG1-Down
PF07690	35	3312	117	24143	4,40E-06	2,72E+00	0,00136181	Major Facilitator Superfamily	G1vsG4-Up
PF07690	35	3312	117	24143	4,40E-06	2,72E+00	0,00136181	Major Facilitator Superfamily	G4vsG1-Down
PF07727	4	2558	208	24052	2,26E-06	1,65E-01	0,0020183	Reverse transcriptase (RNA-dependent DNA polymerase)	G1vsG4-Down
PF07727	4	2558	208	24052	2,26E-06	1,65E-01	0,0020183	Reverse transcriptase (RNA-dependent DNA polymerase)	G4vsG1-Up
PF07727	11	3922	208	24052	2,55E-06	2,88E-01	0,00295066	Reverse transcriptase (RNA-dependent DNA polymerase)	G1vsG2-Down
PF07727	11	3922	208	24052	2,55E-06	2,88E-01	0,00295066	Reverse transcriptase (RNA-dependent DNA polymerase)	G2vsG1-Up
PF00504	14	1290	70	24190	1,58E-05	4,49E+00	0,00443979	Chlorophyll A-B binding protein	G4vsG5-Down
PF00230	7	1290	17	24243	1,44E-05	1,25E+01	0,00443979	Major intrinsic protein	G4vsG5-Down
PF00504	14	1290	70	24190	1,58E-05	4,49E+00	0,00443979	Chlorophyll A-B binding protein	G5vsG4-Up
PF00230	7	1290	17	24243	1,44E-05	1,25E+01	0,00443979	Major intrinsic protein	G5vsG4-Up
PF00520	28	889	309	23951	1,12E-05	2,67E+00	0,00455759	Ion transport protein	G2vsG4-Up
PF00520	28	889	309	23951	1,12E-05	2,67E+00	0,00455759	Ion transport protein	G4vsG2-Down
PF00027	18	2003	72	24188	1,61E-05	3,73E+00	0,00554573	Cyclic nucleotide-binding domain	G2vsG5-Up
PF00027	18	2003	72	24188	1,61E-05	3,73E+00	0,00554573	Cyclic nucleotide-binding domain	G5vsG2-Down
PF13418	6	889	18	24242	3,03E-05	1,32E+01	0,00617551	Galactose oxidase, central domain	G2vsG4-Up
PF13418	6	889	18	24242	3,03E-05	1,32E+01	0,00617551	Galactose oxidase, central domain	G4vsG2-Down
PF07690	18	1290	117	24143	4,56E-05	3,27E+00	0,00638856	Major Facilitator Superfamily	G4vsG5-Down
PF00450	5	1290	9	24251	4,44E-05	2,23E+01	0,00638856	Serine carboxypeptidase	G4vsG5-Down
PF07690	18	1290	117	24143	4,56E-05	3,27E+00	0,00638856	Major Facilitator Superfamily	G5vsG4-Up
PF00450	5	1290	9	24251	4,44E-05	2,23E+01	0,00638856	Serine carboxypeptidase	G5vsG4-Up
PF10371	5	3312	5	24255	4,73E-05	Inf	0,00654256	Domain of unknown function	G1vsG4-Up
PF00112	15	3312	37	24223	4,94E-05	4,33E+00	0,00654256	Papain family cysteine protease	G1vsG4-Up
PF01855	5	3312	5	24255	4,73E-05	Inf	0,00654256	Pyruvate flavodoxin/ferredoxin oxidoreductase, thiamine diP-bdg	G1vsG4-Up
PF01326	8	3312	12	24248	3,54E-05	1,27E+01	0,00654256	Pyruvate phosphate dikinase, AMP/ATP-binding domain	G1vsG4-Up
PF10371	5	3312	5	24255	4,73E-05	Inf	0,00654256	Domain of unknown function	G4vsG1-Down
PF00112	15	3312	37	24223	4,94E-05	4,33E+00	0,00654256	Papain family cysteine protease	G4vsG1-Down
PF01855	5	3312	5	24255	4,73E-05	Inf	0,00654256	Pyruvate flavodoxin/ferredoxin oxidoreductase, thiamine diP-bdg	G4vsG1-Down
PF01326	8	3312	12	24248	3,54E-05	1,27E+01	0,00654256	Pyruvate phosphate dikinase, AMP/ATP-binding domain	G4vsG1-Down
PF00520	48	1944	309	23951	1,09E-05	2,14E+00	0,00740558	Ion transport protein	G2vsG3-Up
PF00520	48	1944	309	23951	1,09E-05	2,14E+00	0,00740558	Ion transport protein	G3vsG2-Down
PF07690	46	5243	117	24143	1,36E-05	2,36E+00	0,00821668	Major Facilitator Superfamily	G1vsG5-Up
PF07690	46	5243	117	24143	1,36E-05	2,36E+00	0,00821668	Major Facilitator Superfamily	G5vsG1-Down

Table 7: PFAM enriched in DEG using genetic of *Cladocopium C1* and Fisher test exact

**Titre:** Adaptation des micro-algues symbiotiques du corail aux changements environnementaux dans l'océan Pacifique

**Mots clés:** Bio-informatique, métatranscriptomique, *Tara Pacific*, Adaptation, Corail, Symbiodiniaceae

**Résumé:** Les récifs coralliens sont un écosystème, qui regroupe une très grande biodiversité marine. Le corail est un animal marin vivant en symbiose obligatoire avec une micro-algue photosynthétique de la famille des Symbiodiniaceae. La famille des Symbiodiniaceae regroupe une très grande diversité classée en 9 genres. Les récifs coralliens sont également un atout majeur pour les populations qui vivent à proximité grâce à l'abondance des poissons qu'ils abritent. Ils sont également d'une importance capitale pour l'économie de ces régions, notamment par la protection des côtes et pour le tourisme. Aujourd'hui, les récifs coralliens sont menacés par des changements environnementaux globaux et locaux qui dégradent ces écosystèmes. Parmi ces changements environnementaux on trouve l'augmentation de la température des océans, des pollutions locales et une augmentation de l'acidification des océans. Ces perturbations de l'environnement marin fragilisent la symbiose obligatoire entre le corail et sa micro-algue, pouvant entraîner une rupture symbiotique (blanchissement des coraux), puis la mort du corail. Dans ce contexte, l'expédition *Tara Pacific* se propose d'étudier l'écosystème corallien à l'échelle d'un océan, en réalisant un vaste plan d'échantillonnage sur 32 îles réparties dans l'océan Pacifique. Au cours de cette expédition (2016-2018), trois genres de coraux ont été étudiés : *Pocillopora*, *Millepora* et *Porites*. Afin d'avoir une vision globale de ces espèces à l'échelle océanique, des méthodes de séquençage haut débit en métagénomique, métatranscriptomique et métabarcodage ont été appliquées sur chaque échantillon de corail. Avec les données *Tara Pacific* des 11 premières îles, les objectifs de la thèse ont été de réaliser une étude transcriptomique globale de la micro-algue photosynthétique, en utilisant

des outils bio-informatiques, afin de mieux comprendre sa diversité et son adaptation à l'échelle d'un océan. La thèse a permis de réaliser des analyses sur l'expression de la micro-algue photosynthétique dans les différentes espèces de coraux séquencés. Une analyse spécifique sur les 11 premières îles de l'expédition, a été portée sur *Cladocopium* et *Durusdinium*, qui sont les micro-algues symbiotiques du corail *Pocillopora*. Pour réaliser cette étude, l'analyse de la structure de populations des micro-algues *Cladocopium* a été faite, suivie par des études d'expression différentielle afin d'identifier les gènes permettant cette adaptation génétique et environnementale. Les résultats montrent que l'adaptation génétique de la micro-algue est déterminante pour l'adaptation du corail hôte dans différents environnements de l'Océan Pacifique. En effet, une grande fidélité symbiotique entre la micro-algue *Cladocopium* et l'espèce *Pocillopora* a été observée. Les voies métaboliques impliquées dans l'adaptation de *Cladocopium* à l'environnement ont été étudiées. Ces résultats ont été comparés avec l'organisme hôte afin de comprendre les mécanismes permettant l'adaptation de l'holobiont corallien *Pocillopora* dans l'océan Pacifique. Pour finir, une étude comparative des micro-algues photosynthétiques présentes dans l'hôte, avec celles présentes dans l'eau environnante a été conduite. Le but de cette étude est de comprendre comment la micro-algue survie en dehors de son hôte, notamment lors d'événements de blanchissement des coraux. Cette thèse a permis par des méthodes bio-informatiques originales de mieux comprendre la diversité et l'adaptation des micro-algues symbiotiques du corail *in situ* et à l'échelle d'un océan.

**Title:** Adaptation of symbiotic coral micro-algae to environmental changes in the Pacific Ocean

**Keywords:** Bio-informatic, metatranscriptomic, *Tara Pacific*, Adaptation, Coral, Symbiodiniaceae

**Abstract:** Coral reefs are an ecosystem which gather a very large marine biodiversity. The coral is a marine animal living in obligatory symbiosis with a photosynthetic micro-algae of the Symbiodiniaceae family. The Symbiodiniaceae family includes a very high diversity classified in 9 genus. Coral reefs are also a major asset for populations living nearby, thanks to the abundance of fish they shelter. They also have a capital importance for the economy of these regions, in particular through the coast protection and the tourism. Today, coral reefs are threatened by global and local environmental changes that degrade these ecosystems. Among these environmental changes, we find oceanic temperature increases, ocean acidification and local pollutions. These disturbances of the marine environment weaken the obligatory symbiosis between the coral and its micro-algae, which can lead to a symbiotic break (coral bleaching), and then the coral death. In this context, the *Tara Pacific* expedition proposes to study the coral ecosystem at ocean scale, carrying out a vast sampling plan on 32 islands distributed in the Pacific Ocean. During this expedition (2016-2018), three kinds of corals genera were studied: *Pocillopora*, *Millepora* and *Porites*. In order, to get a global view of these species at ocean scale, metagenomic, metatranscriptomic and metabarcoding high throughput sequencing methods were applied to each coral sample. With *Tara Pacific* datasets from the first 11 islands, the thesis objectives were to carry out a global transcriptomic study of the pho-

tosynthetic micro-algae, using bioinformatic tools, to better understand its diversity and its adaptation at ocean scale. The thesis made it possible to carry out analyses on the expression of the photosynthetic micro-algae in the different coral species sequenced. A specific analysis was carried out on *Cladocopium* and *Durusdinium*, which are symbiotic micro-algae of the *Pocillopora* coral. Population structure of *Cladocopium* micro-algae was done, followed by differential expression studies, to identify and study the genes allowing this genetic and environmental adaptation. The results indicate that the genetic adaptation of the micro-algae is crucial for the adaptation of the coral host, in different environments of the Pacific. Indeed, a high symbiotic fidelity between the micro-algae *Cladocopium* and the *Pocillopora* species has been observed. The metabolic pathways involved in the adaptation of *Cladocopium* to the environment have been investigated. These results were compared with the host organism in order to understand the mechanisms allowing the adaptation of the coral holobiont *Pocillopora* in the Pacific Ocean. Finally, a comparative study of the photosynthetic micro-algae present in the host, with those present in the surrounding water was done. The aim of this study is to understand how the micro-algae survives outside of its host, especially during coral bleaching events. This thesis made it possible, with original bioinformatic methods, to better understand the diversity and adaptation of coral symbiotic micro-algae *in situ* and at ocean scale.

