EPFL

Thèse n° 8039

Seascape genomics to empower coral reef conservation strategies

Présentée le 4 septembre 2020

à la Faculté de l'environnement naturel, architectural et construit
Laboratoire de systèmes d'information géographique
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

Oliver Michele SELMONI

Acceptée sur proposition du jury

Prof. D. Licina, président du jury
Dr S. Joost, V. Berteaux-Lecellier, directeurs de thèse
Prof. C. Voolstra, rapporteur
Dr L. Fumagalli, rapporteur
Prof. A. Meibom, rapporteur

École polytechnique fédérale de Lausanne

2020

Remerciements

L'histoire de cette thèse a commencé vers la fin de 2015. J'étais en train de terminer mon travail de master dans lequel j'avais passé plusieurs semaines renfermé dans les chambres froides de l'Université de Lausanne (on suivait l'éclosion des œufs de poisson). Je n'avais pas spécialement envie de prolonger cette activité sur les années à suivre, et je commençais à chercher des nouvelles pistes pour la suite.

C'est à ce moment que je suis tombé, un peu par hasard, sur la landscape genomics. Je ne connaissais pas ce sujet, mais j'y ai vu la promesse de pouvoir faire de la « bio-informatique à ciel ouvert », et cela m'a attiré. J'ai donc contacté Stéphane Joost, responsable de la landscape genomics au LASIG et on s'est donné rendez-vous. De façon assez inattendue, Stéphane m'a offert la possibilité de commencer un doctorat au LASIG. Le choix du sujet était à définir, et le début fixé pour Juin 2016. J'ai ainsi accepté.

Pendant les mois qui ont précédé le début de la thèse, je suis parti faire du volontariat dans un centre qui travaille sur les récifs coralliens de Koh Phangan (Thaïlande). Honnêtement, à ce moment précis, j'avais une idée très vague de ce que pouvait être un récif corallien ou un corail. Pendant ces deux mois, je me suis plongé (métaphoriquement et littéralement) dans le monde des coraux. Je suis rentré de Thaïlande avec une idée bien précise du possible sujet de ma recherche. Pour cela, je remercie l'équipe de CORESEA, et en particulier Eike, qui m'a fait découvrir les merveilles du récif, ainsi que son importance.

Au début de la thèse, j'ai donc proposé à Stéphane d'orienter ma recherche vers les récifs coralliens et de passer de la landscape à la seascape genomics. Là encore, de façon un peu inattendue, Stéphane m'a répondu un « oui » sans réserve. Je dois donc te dire un énorme merci, Stéphane, pour ta confiance envers moi depuis les débuts et envers ce sujet « inhabituel ». Merci également pour m'avoir soutenu, encadré et encouragé pendant la réalisation de ce projet.

Le LASIG m'a donné la possibilité de démarrer cette recherche sur un sujet qui me passionne beaucoup. Mais même au-delà de l'aspect scientifique, le LASIG a été un cadre idéal pour effectuer cette thèse. Pour cela je remercie d'abord François Golay, directeur du LASIG, qui m'a également soutenu et qui encourage quotidiennement son équipe à ne pas stresser et à s'accorder des pauses bénéfiques (cf. cafés à Sat). Je remercie aussi les collègues avec lesquels j'ai partagé les bureaux, les cafés, les repas pendant quatre années : Anaïs, Annie (merci pour rafraichir mon anglais), David, Elia, Estelle (merci de m'avoir encouragé à plusieurs reprises), Heydi, Jessie, Noemi, Marc, Marie-Christine, Matthew, Romain, Solange et Thibaud.

Mettre en place un projet sur les coraux c'est compliqué, surtout depuis la Suisse. La pénurie des récifs coralliens dans le Lac Léman nous obligeait à chercher un cas d'étude à l'étranger. Cela n'était pas évident, et initialement le projet sur les coraux avançait avec très peu de certitudes. Tout ça a changé grâce à Laurent Vigliola, chercheur à l'IRD de Nouméa, en

Nouvelle-Calédonie. On ne se connaissait pas personnellement avec Laurent, et sa recherche n'était pas centrée sur les coraux. Cependant, il a fait l'effort de nous mettre en contact avec des chercheurs intéressés à une collaboration. Pour cela, je te remercie beaucoup, Laurent.
C'est comme ça que nous avons rencontré Véronique Berteaux-Lecellier, ma co-directrice de thèse, et Gaël Lecellier, mon co-co-directeur de thèse. Les deux sont chercheurs spécialisés dans le domaine moléculaire des invertébrés aquatiques de Nouvelle-Calédonie. Leurs compétences allaient parfaitement s'intégrer (et quelque fois même s'affronter) avec la seascape genomics. Véronique et Gaël se sont tout de suite engagés dans cette collaboration et dans la co-direction de cette thèse, et à partir de ce moment le projet sur les coraux a décollé. Merci beaucoup, Véro et Gaël, pour vous être investis autant dans ce projet et pour m'avoir supervisé malgré la distance et les fuseaux horaires.
La deuxième année de thèse a été pleine d'évènements. D'abord, l'ICRI (l'initiative international pour les récifs coralliens) et les Nations Unis ont décidé de soutenir financièrement notre recherche (merci à eux), qui devient dès lors le projet « SABLE ». Suisse et Nouvelle-Calédonie se sont rencontrées par visioconférence pendant plusieurs mois, afin d'organiser la récolte d'échantillons des coraux. En février 2018, je suis parti pour rejoindre la campagne de terrain du projet SABLE en Nouvelle-Calédonie, pour une durée de trois mois. Le premier impact avec la Nouvelle-Calédonie a été un peu traumatique : il faisait chaud, j'avais 24h de voyage sur le dos, le mal de mer basculait mon esprit et Véro et Gaël m'ont tout de suite mis la tête sous l'eau (merci pour ça aussi…). Après, j'ai pu trouver mes repères et pour cela je remercie l'IRD de Nouméa (en particulier le labo ENTROPIE) qui m'a accueilli pendant cette période, et aussi les doctorants « calédoniens » avec lesquels j'ai passé des très beaux moments.
Pendant la campagne de terrain, j'ai eu la chance de pouvoir découvrir la beauté du territoire de la Nouvelle-Calédonie dans ses multiples facettes : le lagon, les récifs, les plages, les montagnes, les rivières, les forêts, les prairies. La « bio-informatique à ciel ouvert » était devenue une réalité. L'échantillonnage a été un véritable challenge dans son organisation et aussi dans sa mise en œuvre. Si on a réussi à accomplir cette mission c'est grâce à l'équipe qui m'a accompagné pendant ces expéditions. Merci donc à Miguel Clarque, capitaine du bateau, à Gerard Mou-Tham et Christophe Peignon, plongeurs de l'équipe, et à Joseph Baly, technicien de la team. Merci pour m'avoir supporté et encouragé pendant les missions, merci pour votre patience et pour avoir été méticuleux du premier au dernier échantillon. Mais surtout merci pour les beaux moments de voyage, les casse-croûtes, les poissons qui sautaient dans le bateau, les découvertes culinaires (Miguel, merci pour la roussette !) et les siestes bien méritées.
Cette campagne nous a également mis en contact avec la population de la Nouvelle-Calédonie, souvent avec des personnes vivant dans des endroits reculés. J'aimerai donc remercier les familles et les communautés qui nous ont accueilli pendant nos expéditions, pour l'hospitalité qui a toujours été très généreuse. Merci également aux clans qui nous ont

permis de conduire cette recherche sur leurs récifs. L'attachement qu'ils ont montré envers leur terre a responsabilisé énormément notre recherche.
Les activités qui ont suivi la campagne de terrain ont été également passionnantes, mais d'une façon différente : Il y a eu le travail au labo, puis les analyses à l'ordinateur et enfin la rédaction des différents articles. Je remercie ici tous ceux qui ont contribué à ces différentes taches, en particulier ceux qui m'ont aidé avec les analyses au labo et les co-auteurs des articles.
L'histoire de cette thèse arrive donc à sa conclusion, mais il reste encore quelque personne à remercier. D'abord ma famille, et en particulier mes parents, Marcela et Graziano, qui m'ont permis d'étudier et qui ne m'ont jamais fait manquer leur soutien, même quand ce n'était pas très clair « qu'est-ce que j'allais faire avec ces coraux ». Grazie mille, maintenant vous n'avez plus qu'à lire les 200 prochaines pages pour découvrir cela. Merci également à mes amis, surtout à ceux qui ne se sont pas montrés particulièrement intéressé à ma thèse. Je le dis sans plaisanterie : merci, car ça fait du bien de penser à d'autres choses importantes, comme le foot, le tennis, les vacances, les soupers, les activités d'hiver.
Merci, enfin, à Morgane. Quand on s'est rencontré ma thèse se trouvait face à une impasse. Avec ta rencontre, presque par magie, tout s'est mis en bonne voie. Alors merci, car je crois que tu m'as porté chance, ainsi qu'un soutien constant et beaucoup de bonheur.

Summary

The future of coral reefs is under threat since anomalous heat waves are causing the death of reef building corals around the world. This phenomenon, known as “coral bleaching”, has already resulted in widespread loss of coral over the past 30 years. Without corals, the entire reef ecosystem is expected to collapse, threatening the survival of up to one third of marine wildlife. Despite the catastrophic perspectives, a glimmer of hope is brought by corals that persist at reefs exposed to recurrent heat waves. Evolutionary adaptation might underpin these observations. Characterizing the adaptive potential of corals is therefore essential, as conservation efforts could be oriented towards coral breeds resistant to future thermal conditions. To date, however, the characterization of heat adaptation in corals shows substantial gaps at different scales. At the molecular level, more needs to be discovered about the cellular pathways implicated. At a population level, there are no well-established frameworks to detect heat-tolerant corals for use in defining conservation priorities.

In this research, we employed a seascape genomics approach to contribute to filling these gaps. This approach combines the environmental characterization of the seascape with a genomic analysis of populations, with the goal to identify genetic variants underpinning an adaptive role. Seascape genomics therefore constitutes an exploratory approach that portrays the adaptive potential across a population, and this facilitates the transposition of results to a conservation perspective. Furthermore, the use of genomic analyses also provides the bases to formulate hypotheses from the molecular side of adaptation.

The first two articles of this thesis prepare the ground for the application of seascape genomics to corals. The first focuses on the optimization of sampling strategies. By using computer simulations, we defined the practical guidelines to organize a sampling strategy securing sufficient statistical power. The second is the proof-of-concept for the application of seascape genomics to corals. We retrieved a pre-existing genomic dataset on corals from Japan, and used it to characterize adaptive potential against heat stress and to compute conservation priorities accordingly.

The third and the fourth articles of this thesis are based on new data collected in the frame of a project dedicated to the study of adaptation in three coral species from New Caledonia (Southwestern Pacific). We organized the sampling strategy using the guidelines of the first article, and processed data using the framework described in the second. The result is a dataset tailored to the needs of seascape genomics, and this enabled to refine our methods. In the third article we focus on the molecular side of adaptation and disclose cellular pathways potentially involved in heat stress adaptation. The fourth article emphasizes the role of adaptive potential under a conservation perspective. We employed pre-existing field survey data to show that reefs predicted with higher adaptive potential suffered reduced coral loss after heat stress.

Together, this thesis highlights the value of the seascape genomics approach for characterizing the adaptive potential of corals and for supporting reef conservation strategies. More broadly, this work shows that building bridges from academic research to tangible conservation actions is not only possible but also essential to face this crisis.

Keywords: Seascape genomics, coral reef, coral bleaching, local adaptation, climate change, marine conservation, heat adaptation, sampling strategy

Résumé

Le futur des récifs coralliens est menacé au niveau mondial et les vagues anormales de chaleur qui se succèdent en sont essentiellement responsables. La disparition des coraux, base structurale de l'architecture récifale, menace la survie de près d'un tiers des espèces marines. Malgré les perspectives catastrophiques, une lueur d'espoir est apportée par la présence de coraux qui se perpétuent dans des récifs exposés à des vagues de chaleur récurrentes. Des éléments liés à l'adaptation évolutive pourraient expliquer cette résistance. Caractériser le potentiel adaptatif des coraux est primordial, les efforts de conservation pouvant alors être orientés vers les coraux résistants aux futures conditions thermiques. Cependant, la caractérisation de l'adaptation des coraux à la chaleur présente des difficultés à différentes échelles. Au niveau moléculaire, les mécanismes cellulaires impliqués restent encore insuffisamment connus. A l'échelle de la population, il n'existe pas de démarche bien définie pour détecter les coraux adaptés et les intégrer dans les stratégies de conservation.

Dans cette thèse, nous utilisons une approche appelée *seascape genomics* (génomique de l'environnement sous-marin) pour contribuer à combler ces lacunes. Cette approche, située à l'intersection des analyses environnementale et génétique, permet de caractériser le potentiel adaptatif d'individus constituant une population. De plus, les analyses génomiques permettent également de formuler des hypothèses concernant les aspects moléculaires de l'adaptation.

Les deux premiers articles de cette thèse préparent le terrain pour l'application de la seascape genomics aux coraux. Le premier article est consacré aux stratégies d'échantillonnage. En utilisant des simulations, nous avons défini des lignes directrices à suivre dans le but d'organiser une stratégie d'échantillonnage assurant suffisamment de puissance statistique. Le deuxième article démontre la faisabilité de l'application de la *seascape genomics* aux coraux. Nous avons utilisé des données génétiques préexistantes sur des coraux du Japon pour évaluer leur potentiel d'adaptation face au stress thermique et ainsi proposé des priorités de conservation en conséquence.

Le troisième et le quatrième articles sont basés sur de nouvelles données, collectées dans le cadre d'un projet dédié à l'étude de l'adaptation de trois espèces de coraux de la Nouvelle-Calédonie (sud-ouest du Pacifique). La stratégie d'échantillonnage a été déterminée sur la base des lignes directrices définies dans le premier article et les analyses effectuées en utilisant les méthodes décrites dans le deuxième. Le troisième article est centré sur les aspects moléculaires et révèle des gènes candidats de l'adaptation au stress thermique. Le quatrième article traite les implications du potentiel adaptatif pour la conservation des coraux. En utilisant les données publiques de suivis de terrain, nous avons montré que la perte de la couverture corallienne est moindre chez les récifs que nous avons prédits comme étant à haut potentiel adaptatif.

Cette thèse démontre donc l'utilité de caractériser le potentiel adaptatif des coraux par la *seascape genomics* pour soutenir les stratégies de conservation récifale. Plus largement, ce travail montre que construire des ponts entre recherche académique et conservation est non seulement possible mais également indispensable pour affronter la crise du blanchiment.

Mots-clés: Génomique de l' environnement sous-marin, récif corallien, blanchissement du corail, adaptation locale, changement climatique, conservation de la mer, adaptation à la chaleur, stratégie d'échantillonnage

Table of contents

Remerciements *1*

Summary *4*

Résumé *5*

Table of contents *6*

List of figures **10**

List of tables **10**

1. Introduction *12*

1.1. The coral reef crisis **12**

1.1.1. The importance of coral reefs 12

1.1.2. What are corals? 12

1.1.3. Coral bleaching 14

1.1.4. The worldwide decline of coral reef 16

1.1.5. Conservation strategies 17

1.1.6. Evolutionary adaptation to the rescue? 18

1.2. Seascape genomics **19**

1.2.1. Local adaptation: from land to sea 19

1.2.2. Remote sensing of the seascape 19

1.2.3. Genomics of marine organisms 21

1.2.4. Genotype-environment association analysis 23

1.2.5. Strengths of seascape genomics 24

1.2.6. Weaknesses of seascape genomics 25

1.3. Aim of the thesis **26**

1.4. Layout of the thesis **27**

1.4.1. Preparatory studies 27

1.4.2. The SABLE project 28

2. Article A - Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach *33*

2.1. Abstract **34**

2.2. Introduction **35**

2.3. Material and Methods **37**

2.3.1. Environmental data 39

2.3.2. Computation of genotypes 39

2.3.3. Evolutionary scenarios and parametrization 40

2.3.4. Sampling design 43

2.3.5. Landscape genomics analysis 45

2.3.6. Simulations and evaluation of the performance 46

2.4. Results **47**

2.4.1. Parameters of simulations 47

2.4.2. True Positive Rate....48
2.4.3. False Discovery Rate....49
2.4.4. Positive Predictive Value....50

2.5. Discussion....51
2.5.1. How many samples are required to detect any adaptive signal?....51
2.5.2. How many sampling sites?....53
2.5.3. Where to sample?....54
2.5.4. Limitation....54
2.5.5. Conclusions....55

2.6. Acknowledgements....56

3. Article B - Seascape genomics as a new tool to empower coral reef conservation strategies: An example on north-western Pacific Acropora digitifera....57

3.1. Abstract....58

3.2. Introduction....59

3.3. Material and Methods....61
3.3.1. Genomic dataset....63
3.3.2. Environmental data....64
3.3.3. Seascape genomics....65
3.3.4. Annotation of seascape genomics results....65
3.3.5. Probability of presence of heat stress adapted genotypes....66
3.3.6. Sea current data....66
3.3.7. Connectivity analysis....67
3.3.8. Connectivity predictions....68
3.3.9. Evaluation of the adaptive potential against heat stress....69

3.4. Results....69
3.4.1. Seascape genomics....69
3.4.2. Probability of presence of heat stress adapted genotypes....71
3.4.3. Connectivity modelling....72
3.4.4. Evaluation of the adaptive potential....73

3.5. Discussion....74
3.6.2. Adaptation to heat stress....74
3.6.3. Connectivity patterns....74
3.6.4. Heat stress adaptive potential in the 2016 bleaching event....75
3.6.5. Limitations and future directions....77
3.6.6. Application in conservation....78
3.6.7. Conclusions....79

3.6. Acknowledgements....79

3.7. Data Archiving Statement....79

4. Article C - Seascape genomics reveals candidate molecular targets of heat stress adaptation in three coral species....80

4.1. Abstract....81

4.2. Introduction....82

4.3. Material and Methods....85
4.3.1. Environmental data....85
4.3.2. Sampling....86
4.3.3. Species identification....87

4.3.4. Screening and SNP genotyping 87
4.3.5. SNP filtering 87
4.3.6. Neutral genetic structure analysis 88
4.3.7. Seascape genomics 89
4.3.8. Annotation analysis of heat stress associated SNPs 89

4.4. Results 90
4.4.1. Neutral genetic structure 91
4.4.2. Local adaptation 92
4.4.3. Functional annotations of heat stress associated SNPs 93

4.5. Discussion 95
4.5.1. Different types of heat stress adaptation 95
4.5.2. Candidate molecular targets for heat stress adaptation 95
4.5.3. Limitations and future directions 98
4.5.4. Conclusions 100

4.6. Acknowledgments 100

5. Article D - Coral cover surveys corroborate predictions on reef adaptive potential to thermal stress 101

5.1. Abstract 102

5.2. Introduction 103

5.3. Results 105
5.3.1. Heat stress and probability of adaptation 105
5.3.2. Reef connectivity and genetic correlation between corals 107
5.3.3. Reef connectivity indices 107
5.3.4. Coral cover analysis 109

5.4. Discussion 110
5.4.1. Local divergences in conservation indices 110
5.4.2. Predictions on adaptive potential match coral cover 111
5.4.3. Limitations and future directions 113
5.4.4. Conclusions 114

5.5. Methods 115
5.5.1. Remote sensing of sea surface temperature 115
5.5.2. Seascape connectivity graph 115
5.5.3. Connectivity indices 116
5.5.4. SNPs dataset 116
5.5.5. Probability of heat stress adaptation 117
5.5.6. Reef connectivity and genetic correlations between corals 117
5.5.7. Coral cover data 118
5.5.8. Environmental characterization of survey sites 118
5.5.9. Analysis of coral cover change 119

5.6. Acknowledgments 119

6. General discussion 121

6.1. Answers to research questions 121

6.2. Perspectives for coral reef conservation 122

6.3. Perspectives in seascape genomics 125

6.4. General conclusion 126

7. Appendices ***128***

7.1. Supplementary material from article A **128**

7.2. Supplementary material from article B **134**

7.3. Supplementary material from article C **145**

7.4. Supplementary material from article D **151**

7.5. The ManaCo consortium **153**

8. Bibliography ***159***

9. Curriculum vitae ***170***

List of figures

Figure 1-1. Coral reefs, colonies and polyps. 13
Figure 1-2. Coral bleaching. 15
Figure 1-3. The decline of coral cover in the Indo-Pacific. 16
Figure 1-4. The *seascape* behind seascape genomics. 20
Figure 1-5. The *genomics* behind seascape genomics. 22
Figure 1-6. A Seascape genomics Approach to improve coral reefs conservation strategies against BLEaching (SABLE). 29
Figure 2-1. Workflow for each iteration of the simulative approach. 38
Figure 2-2. Comparison of genotypes simulated with CDPOP and our method. 42
Figure 2-3. The three sampling design approaches accounting for landscape characteristics. 44
Figure 2-4. Effects of sampling strategy on the landscape genomics simulations. 50
Figure 3-1. Study area. 61
Figure 3-2. Workflow between the steps of the approach. 62
Figure 3-3. Calculation of connectivity and adaptive potential indices. 68
Figure 3-4. Probability of carrying heat stress adapted genotypes ($PAheat$). 71
Figure 3-5. Connectivity indices. 72
Figure 3-6. Index of adaptive potential against heat stress ($APIheat$). 73
Figure 4-1. Study area and sampling sites. 84
Figure 4-2. Principal component analysis (PCA) of the genotype matrices for three species studied, *Acropora millepora*, *Pocillopora damicornis* and *Pocillopora acuta*. 91
Figure 4-3. Example of significant genotype-environment association. 92
Figure 5-1. Reef system of New Caledonia. 104
Figure 5-2. Bleaching alert frequency and probability of heat stress adaptation. 106
Figure 5-3. Seascape cost distance and genetic correlation between corals. 107
Figure 5-4. Connectivitiy indices. 108
Figure 5-5. Coral cover association analysis. 109
Figure 6-1. The Digital Reef Adaptive Potential EvAlUator (DRAPEAU). 124

List of tables

Table 2-1. Sampling design in landscape genomics studies 36
Table 2-2. Table of factors varying in the simulative approach. 47
Table 2-3. Effect sizes of element defining the sampling strategy on TPR, FDR and PPV. 48
Table 3-1. Significant genotype-environment associations (SGEA) 70
Table 3-2. Field report of the 2016 mass bleaching event 76
Table 4-1. Workflow of the analysis. 91
Table 4-2. Significant genotype-environment associations in *Acropora millepora* (*AM*), *Pocillopora damicornis* (*PD*) and *Pocillopora acuta* (*PA*) 93
Table 4-3. Functional annotations of heat stress associated SNPs. 94
Table 4-4. Candidate molecular targets for coral adaptation to heat stress. 99
Table 5-1. Implications for reef conservation in New Caledonia 112

1. Introduction

The first section of the introduction focuses on coral reefs and describes the importance of this ecosystem, including the reasons why it is under threat and the issues related to conservation strategies. The second section presents the field of seascape genomics and describes the method, as well as its strengths and weaknesses. The third section states the scientific aims of this thesis, with particular emphasis on the gaps in coral reef research and conservation that can be filled by seascape genomics. Finally, the fourth section details the structure of the thesis and highlights the content of the four scientific articles that follow.

1.1. The coral reef crisis

1.1.1. The importance of coral reefs

Coral reefs are estimated to cover less than 1% of the ocean floor, yet they host approximately one third of marine wildlife (Moberg & Folke, 1999). This is due to the fact that many marine species, across a range of taxonomical categories, depend on this habitat for spawning, breeding, feeding, hunting or nursing (Moberg & Folke, 1999; Roff et al., 2016; J. L. W. Ruppert, Travers, Smith, Fortin, & Meekan, 2013). It is therefore not surprising that coral reef health shapes the structure and abundance of fish communities, and can induce significant bottom-up changes on the planet's food chains and biogeochemical cycles (Roff et al., 2016; J. L. W. Ruppert et al., 2013). This has a direct impact on local fishing activities, which is noteworthy as coral reefs are present in over a hundred countries and are a source of protein for tens of millions of people (Costanza et al., 2014; Moberg & Folke, 1999). Moreover, coral reefs represent a crucial physical barrier to mitigate the impact of currents, waves and storms on the coastlines, with an economical value estimated of up to several billions of dollars (Cesar, Burke, & Pet-soede, 2003; Costanza et al., 2014). The tourism industry, which contributes conspicuously to the economy of tropical regions, is also hugely reliant on this ecosystem due to its aesthetic value (White, Vogt, & Arin, 2000). Additionally, reefs have an impact on human health, by de-toxifying human waste in water and by providing useful molecules suitable for the development of anti-cancer treatments in biomedical sciences (Imhoff, Labes, & Wiese, 2011; Mayer & Gustafson, 2006; Moberg & Folke, 1999).

1.1.2. What are corals?

Coral reefs are composed of an amazingly intricate network of different organisms, of which reef building scleractinian corals (*i.e.* corals with a hard skeleton) are the pivotal elements as

they physically shape this ecosystem (Fig. 1-1; Moberg & Folke, 1999; Roff et al., 2016). Scleractinian corals are marine invertebrates of the phylum Cnidaria, organized into tightly associated colonies of identical polyps that share a calcareous external skeleton (E. Ruppert, Fox, & Barnes, 2004). There are hundreds of scleractinian coral species, encompassing variegated morphologies of the colony, where the variety of these morphologies determines the type of reef structures (Cairns, 1999). Life strategies are also diversified between species, with differential growth rates and reproductive modes (Darling, Alvarez-Filip, Oliver, McClanahan, & Côté, 2012). For instance, one of the reproductive mode is broadcast spawning, in which gametes are mass-released once a year and the resulting larvae can be passively dragged by sea currents for up to hundreds of kilometres (Darling et al., 2012; Nishikawa, Katoh, & Sakai, 2003). In contrast, in corals reproducing using a brooding strategy the reproductive season is longer, fertilization is internal and larvae usually settle over shorter distances (Ayre & Hughes, 2000).

Despite numerous differences, the key to coral's evolutive success is shared between nearly all the species: the symbiotic association with unicellular brown algae from the family Symbiodiniaceae (LaJeunesse et al., 2018; Mydlarz, McGinty, & Harvell, 2010). These algae receive protection by living intracellularly in the coral cells and, in turn, provide the host with photosynthetic products that are essential for its sustainment (Davy, Allemand, & Weis, 2012). In fact, at least half of the metabolic requirements of a healthy coral are provided by its symbionts (Houlbrèque & Ferrier-Pagès, 2009; Palardy, Rodrigues, & Grottoli, 2008). In addition to Symbiodiniaceae, corals are also symbiotically associated with an external microbial community composed by bacteria, archaebacteria and viruses (Grabherr et al., 2011; Marhaver, Edwards, & Rohwer, 2008). This symbiotic assembly of coral and all of its symbionts is called the "coral holobiont".

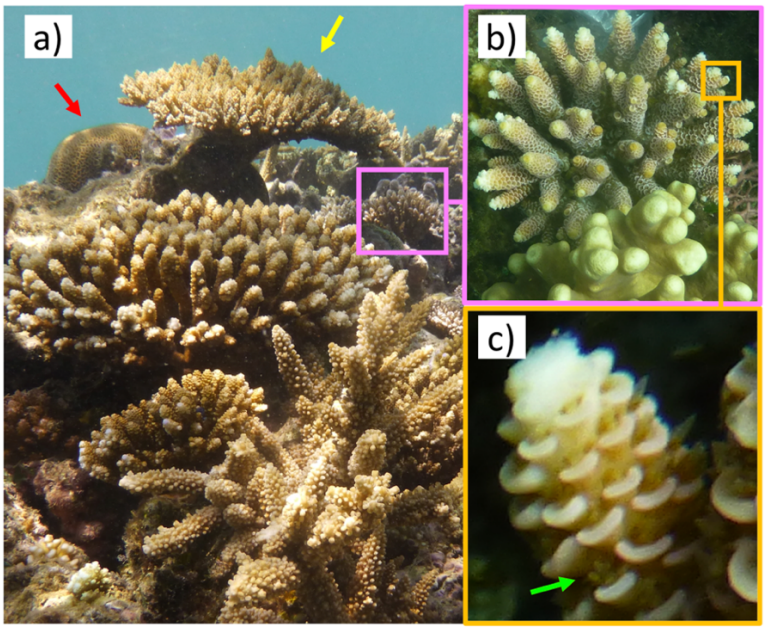


Figure 1-1. Coral reefs, colonies and polyps. Picture (a) shows a coral reef of New Caledonia. Multiple coral colonies of different species and with distinct morphologies are visible. For instance, the red arrow points a coral with a *massive* morphology, while the yellow one a colony with a *branching* morphology. Picture (b) shows a single colony with branching morphology. The zoom on a single branch (c) displays the presence of translucent polyps (*e.g.* green arrow) inside each of the calyces formed by the calcareous skeleton. [photo credits: Oliver Selmoni]

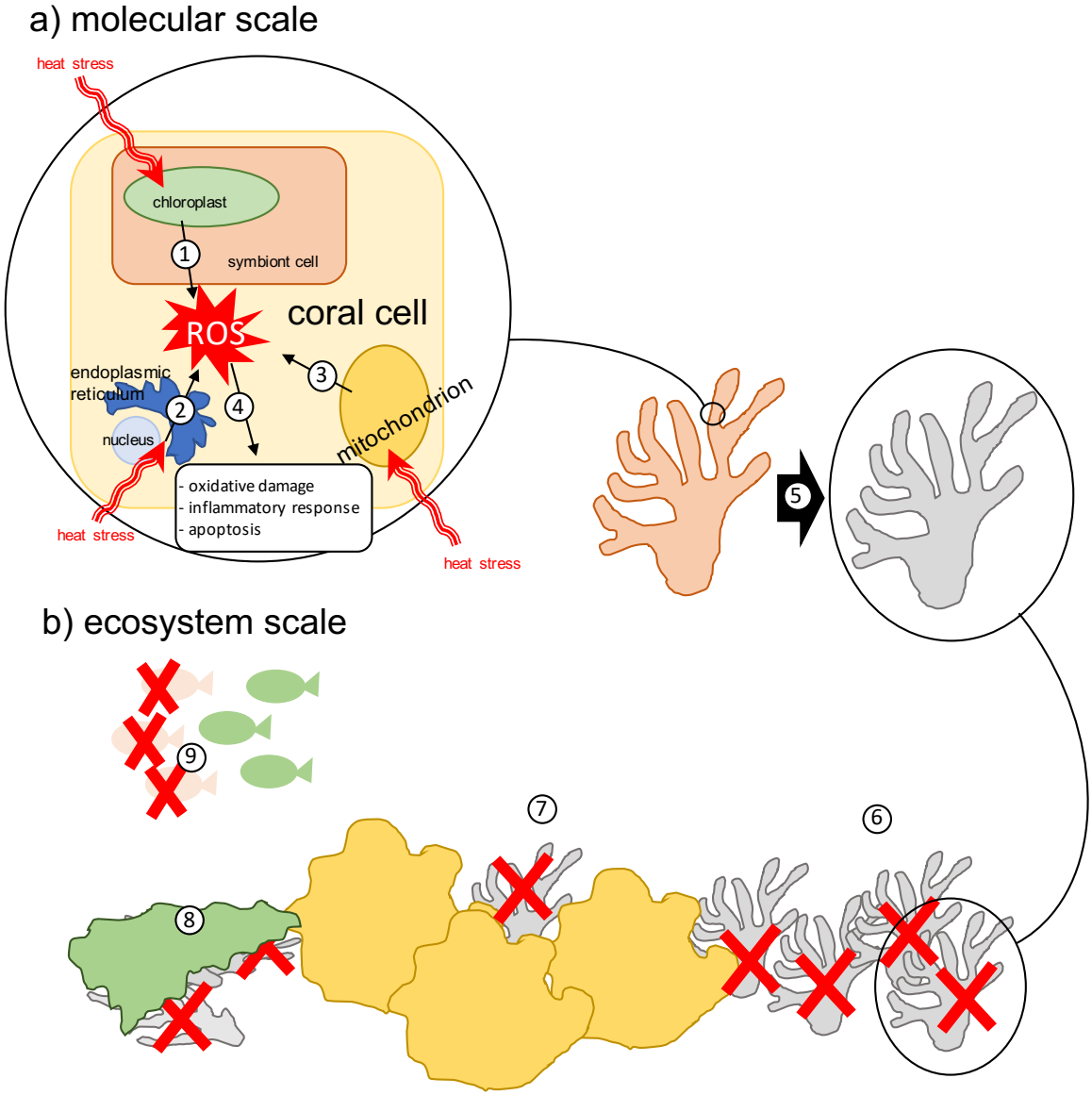
1.1.3. Coral bleaching

When corals are exposed to stressful conditions, such as an increase of water temperature, the symbiotic relationship with Symbiodiniaceae is broken: brown coloured symbionts leave the coral host cells causing the colony to lose colour, which is termed 'bleaching' (van Oppen & Lough, 2009). The bleaching state is reversible, but it cannot be endured over long periods. In fact, a bleached colony lacks of the metabolic input provided by symbiotic algae and necessary for the efficiency of physiological processes, such as defence mechanisms or immune system (Mydlarz et al., 2010). This is why a coral that is constantly in a bleached state will die and be covered by macro algae (Diaz-Pulido & McCook, 2002). The susceptibility to coral bleaching is variable between different coral species (Loya et al., 2001). For instance, coral species that have a branched morphology are more sensitive to heat stress, compared with corals with a more compact morphology (known as massive corals; Loya et al., 2001; Mydlarz et al., 2010). Furthermore, the type of symbionts hosted can also contribute to the degree of heat tolerance (Howells et al., 2020; Howells, Berkelmans, van Oppen, Willis, & Bay, 2013; Hume et al., 2013; Mydlarz et al., 2010; Sampayo et al., 2016).

The molecular mechanisms participating in the bleaching cascade have been extensively investigated over the last decades (Desalvo et al., 2008; Maor-Landaw & Levy, 2016; Nielsen, Petrou, & Gates, 2018; Ricaurte, Schizas, Ciborowski, & Boukli, 2016; Rosic, Pernice, Dove, Dunn, & Hoegh-Guldberg, 2011). According to one of the most supported theories, the crucial point of the bleaching cascade is the accumulation of toxic oxygen molecules (known as reactive oxygen species, ROS) inside the cytoplasm of coral cells (Fig. 1-2; Nielsen et al., 2018; Oakley et al., 2017). ROS are oxidants that can cause severe damage of the coral tissues, leading to the activation of inflammatory responses and programmed cell death pathways (Patel, Rinker, Peng, & Chilian, 2018). In response to heat stress, the coral cell elicits antioxidative mechanisms to contrast ROS accumulation (Kültz, 2005; Nielsen et al., 2018; Oakley et al., 2017; Voolstra et al., 2009, 2011). Another well-characterized cellular response against heat stress is the activation of heat shock proteins, also known as molecular chaperones (Desalvo, Sunagawa, Voolstra, & Medina, 2010; Desalvo et al., 2008; Fuller et al., 2019; Ishikawa, Wirz, Vranka, Nagata, & Bächinger, 2009; Oakley et al., 2017; Rosic et al., 2011; van Oppen & Lough, 2009). As heat stress interferes with the folding/unfolding of new proteins in the Endoplasmic Reticulum, the role of molecular chaperones is to stabilize these new proteins so that they can become functional.

Numerous cellular pathways participating in the bleaching cascade and heat stress resistance have already been identified. However, little is known about which of these molecules might be subjected to evolutionary processes that increase coral heat tolerance (van Oppen, Oliver, Putnam, & Gates, 2015).

Figure 1-2. Coral bleaching. Panel (a) shows the molecular pathways leading to coral bleaching according to "oxidative stress theory". Heat stress causes oxidative damage in the symbiont chloroplast, resulting in a leakage of reactive oxygen species (ROS) into the coral cell cytoplasm (1). Heat stress also enhances the endogenous production of ROS in the coral cell, through the protein folding machinery in the endoplasmic reticulum (2) and in the mitochondrion (3). Accumulated ROS cause oxidative damage to coral tissues and elicits inflammatory and apoptotic (programmed cell death) responses (4). The symbiosis between coral and symbionts is broken, resulting in coral bleaching (5). In panel (b), some of the consequences of coral bleaching at the scale of the reef ecosystem are shown. Persistent bleaching leads to the death of corals and therefore to a loss of coral cover (6). The coral community assemblage shifts towards coral morphs that are more resistant to bleaching (7). Macro algae overgrow on dead corals, destroying the reef architecture (8). Because of this algal bloom, reef communities shift towards herbivory (9). [Drawing: Oliver Selmoni, adapted from Desalvo et al., 2008; Oakley & Davy, 2018]



1.1.4. The worldwide decline of coral reef

Over the last two decades, mass coral bleaching events have increasingly been reported around the world in concomitance with anomalous heat waves (Donner, Rickbeil, & Heron, 2017). Of note, coral bleaching conditions are not driven by peaks of high sea temperature alone, but rather by peaks exceeding the seasonal maxima (Liu, Strong, & Skirving, 2003). Although such anomalous sea water oscillations occur naturally (*e.g.* those associated with the *El niño* phenomenon; Hughes et al., 2018), climate change is exacerbating their amplitude (B. Wang et al., 2019) and by 2050 bleaching conditions are expected to be persistent worldwide (Van Hooidonk, Maynard, & Planes, 2013). Furthermore, additional factors are synergistically contributing to coral loss, such as ocean acidification, water eutrophication, sedimentation and overfishing (Ateweberhan et al., 2013; Hughes et al., 2017; Maina, McClanahan, Venus, Ateweberhan, & Madin, 2011).

Global warming and the degradation of ocean conditions are causing the severe losses of coral cover worldwide (Fig. 1-3; Bruno & Selig, 2007; Wilkinson, 2008). The Australian Great Barrier Reef is one of the reef systems most heavily impacted, with local coral losses of up to 50% during the 2016 bleaching event (Hughes, Kerry, et al., 2018). In addition, coral bleaching is causing marked changes in the composition of the coral community, resulting in a loss of biodiversity (Hughes, Kerry, et al., 2018; Loya et al., 2001). These changes, together with the overgrowth of macro-algae on dead corals, trigger marked bottom-up transformations in reef ecosystems (De'ath, Fabricius, Sweatman, & Puotinen, 2012; Pratchett, McCowan, Maynard, & Heron, 2013; J. L. W. Ruppert et al., 2013). In the most extreme cases, coral loss may cause a decline in fish species richness of up to 60% (Pratchett, Thompson, Hoey, Cowman, & Wilson, 2018), where such negative effects can move up the food chain and strike top predators of the reefs (Roff et al., 2016; J. L. W. Ruppert et al., 2013).

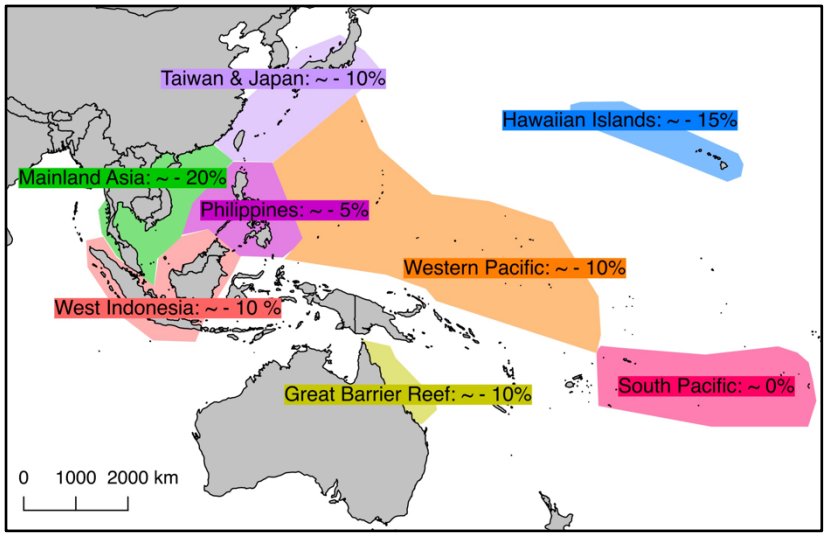


Figure 1-3. The decline of coral cover in the Indo-Pacific. For eight regions of the Indo-Pacific, the map shows the average change in coral cover reported in field survey records covering the period 1997-2004, in comparison with records covering the period 1968-1983. The figure is adapted from Bruno and Selig (2007), simplified here for illustrative purposes.

1.1.5. Conservation strategies

The decline of coral reefs worldwide calls for a development of strategies to restore reefs that have undergone severe coral loss, as well as limit the impact of future bleaching events (Baums, 2008; Bellwood, Hughes, Folke, & Nyström, 2004; C. N. Young, Schopmeyer, & Lirman, 2012). Coral reef conservation efforts primarily rely on one tool: the establishment of marine protected areas (MPAs; Bellwood et al., 2004; Young et al., 2012). MPAs are designated areas where human access and activities are restricted, with beneficial effects already observed on reefs worldwide (Bellwood et al., 2004; Cinner et al., 2016; Lester et al., 2009; Palumbi, 2003).

Recent studies on hydrodynamics predictions and coral population genetics suggests that the establishment of MPA networks should take into account the interconnection between reefs based on sea currents, particularly at how coral larvae disperse across ocean waters (Krueck et al., 2017; Lukoschek, Riginos, & van Oppen, 2016). This would optimize management efforts and create synergistic effects between distant protected areas. Furthermore, connectivity can contribute to coral cover recovery after a bleaching event, even though this depends on the frequency of heat waves, as well as on the dispersal characteristics and growth rates of different coral species (Robinson, Wilson, & Graham, 2019).

In addition, MPA design should account for variability of climatic conditions, as these data provide insights on the degree of exposure to environmental stress (Magris, Pressey, Weeks, & Ban, 2014; Maina et al., 2011; McLeod, Salm, Green, & Almany, 2009). Up to now, however, information on environmental stress is rarely used in spatial planning of marine conservation, which might explain why MPAs have not conferred resistance against the extreme heat waves associated with previous mass bleaching events (Hughes et al., 2017; Magris et al., 2014; OECD, 2017).

The same issues are encountered by coral nursery establishment plans (Baums, 2008). Coral nurseries are underwater gardens where colonies grow under protected conditions, specifically for restoration transplantation of damaged reefs. The efficiency of this method relies on the capability of the imported coral to survive in the new environment, an aspect that is often underestimated (Baums, 2008; Baums et al., 2019; van Oppen et al., 2017).

The difficulties in applying recommendations from innovative fields of coral reef research to conservation management have several causes. The main one is communication, as research findings discussed in scientific publications are often too technical and hard to interpret from a conservation standpoint (Bainbridge, 2014; Beger et al., 2014). Another cause is the lack of representativity, as research work often covers specific temporal and spatial scales that are not necessarily relevant in other conservation contexts (Fisher et al., 2011; Rose et al., 2018).

1.1.6. Evolutionary adaptation to the rescue?

Some coral reefs that have been recurrently exposed to thermal stress appear to have developed resistance against heat (Dance, 2019; Fine, Gildor, & Genin, 2013; Hughes et al., 2019; Krueger et al., 2017; Penin, Vidal-Dupiol, & Adjeroud, 2013). It has been largely debated whether such observations could be due to acclimatization or adaptation (Logan, Dunne, Eakin, & Donner, 2014; Palumbi, Barshis, Traylor-Knowles, & Bay, 2014).

In acclimatization, an individual adjusts its phenotype (morphological or physiological) in response to an environmental change. Traits modified under acclimatization are generally reversible and non-heritable (except for epigenetic traits; Lind & Spagopoulou, 2018). Several studies have shown that acclimatization, by itself, would hardly allow corals to cope with the increase in water temperature expected in the next years (Howells et al., 2013; Logan et al., 2014).

In contrast, adaptation is an evolutionary phenomenon where DNA mutations provoke phenotypic changes that confer a selective advantage in a given environment (Kawecki & Ebert, 2004). Adaptation is heritable and its potential resides within the population (and not within each individual, as acclimatization), which might explain why the limits of adaptation are harder to define (Logan et al., 2014). Characterizing the adaptive potential of corals could have crucial implications for conservation (Baums et al., 2019; McLeod et al., 2009; van Oppen et al., 2017, 2015; K. L. Wilson, Tittensor, Worm, & Lotze, 2020). For example, reefs hosting heat stress adapted corals could be protected from local stressors (*e.g.* pollution) by establishing MPAs. Information on adaptive potential could also be applied to reef restoration plans, for example to support the choice of coral breeds for nurseries.

Adaptation studies usually rely on testing the physiological performances of corals after experimental exposure to heat stress (Howells et al., 2013; Krueger et al., 2017; Palumbi et al., 2014; Sampayo et al., 2016; Ziegler, Seneca, Yum, Palumbi, & Voolstra, 2017). Such exposure can be natural, through the reciprocal transplantation between reefs, or artificial, via aquarium conditioning. These methods, which are often used in combination, offer a direct measure of the selective advantages provided by adaptation (Pardo-Diaz, Salazar, & Jiggins, 2015). However, these assays are laborious and usually focus on colonies from a couple of reefs exposed to contrasting thermal patterns (Howells et al., 2013; Krueger et al., 2017; Palumbi et al., 2014). Complementary approaches are therefore necessary to explore the adaptive potential across the staggering diversity of environmental conditions, and genetic variants, appreciable across an extensive reef system (Riginos, Crandall, Liggins, Bongaerts, & Treml, 2016).

1.2. Seascape genomics

1.2.1. Local adaptation: from land to sea

Landscape genomics is a discipline of population genomics dedicated to the study of local adaptation through field-based investigations (Balkenhol et al., 2017; Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015). The basic concept of local adaptation is that organisms persisting in a stressful environment carry the genetic traits that allow them to cope with the selective pressure (Kawecki & Ebert, 2004; Rellstab et al., 2015; Riginos et al., 2016). The landscape genomics method consists in describing the environmental variability of a landscape, while simultaneously characterizing the genomic diversity of the resident population (Rellstab et al., 2015). Statistical methods are then used to investigate the presence of genetic variants whose frequencies correlate with specific environmental gradients (Joost et al., 2007). These genetic variants are considered as potentially adaptive, as they might correspond to a mutation conferring a selective advantage (Balkenhol et al., 2017). Landscape genomic experiments have already been performed on a range of taxa, from livestock to wild mammals (Colli et al., 2014; Lv et al., 2014; Stucki et al., 2017; Vajana et al., 2018), as well as insects (Crossley, Chen, Groves, & Schoville, 2017; Dudaniec, Yong, Lancaster, Svensson, & Hansson, 2018) and plants (Abebe, Naz, & Léon, 2015; Pluess et al., 2016; Yoder et al., 2014).

Seascape genomics is a subset of landscape genomics that is dedicated to marine populations (Riginos et al., 2016). The method has already been applied to some species (*e.g.* on American lobster, Benestran et al., 2016; Atlantic salmon, Vincent, Dionne, Kent, Lien, & Bernatchez, 2013 and European eel, Laporte et al., 2016), but the uptake in the marine environment has been slower than on land (Riginos et al., 2016). Among the reasons for this delay stand the conceptual and practical difficulties (detailed in the following sections) encountered when trying to quantify the environment and the genetic structure in an aquatic ecosystem (Riginos et al., 2016; Selkoe, Scribner, & Galindo, 2015). As a consequence, seascape genomics studies on corals are rare and often performed on suboptimal datasets (*e.g.* low number of genotypes, samples and/or sampling locations; Bay & Palumbi, 2014; Lundgren, Vera, Peplow, Manel, & van Oppen, 2013; Thomas, Kennington, Evans, Kendrick, & Stat, 2017).

1.2.2. Remote sensing of the seascape

The most important difference between a seascape and a landscape is the strength of physical flow (Riginos et al., 2016; Selkoe et al., 2015). Water flow dominates the fluid environment and therefore causes patterns of climatic variability that are more asymmetrical than those observed on land (Riginos et al., 2016; Selkoe et al., 2016). Moreover, currents show a strong non-stationary behaviour that must be considered, as hydrographic regimes vary seasonally, annually or on a multi-year basis (Selkoe et al., 2015). These characteristics can lead to major

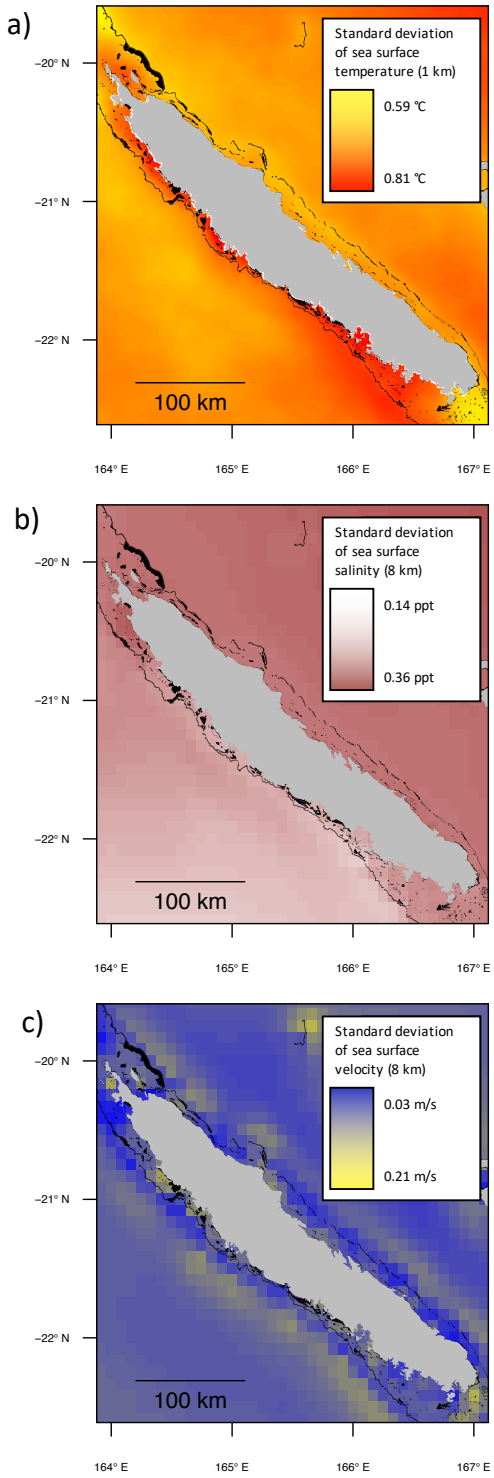
environmental differences between sites located a few kilometres apart, or even that a same site could drastically change its conditions at different time (Riginos et al., 2016).

These considerations have important implications in local adaptation studies (Leempoel et al., 2017; Riginos et al., 2016). The spatial resolution of the study area should be adapted to the activity range of the studied species. In sessile species such as corals, for instance, divergent selective forces can appear at reefs located a few kilometres apart (*e.g.* Bay & Palumbi, 2014), while the same climatic contrast would not drive selection in mobile species

as sharks. The same applies to temporal resolution, which should also be tailored to the characteristics of the studied species. For example, coral bleaching is not driven by anomalous thermal variation in general, but rather by anomalous thermal variation during the hot season (Liu et al., 2003). Last, the temporal extent of the environmental characterization needs to match the timing of the evolutionary processes in the studied species. Clonal organisms as corals, for instance, reach sexual maturity after several years (Harvell & Grosberg, 1988) and adaptation is therefore likely to occur over decades (Logan et al., 2014).

Under a practical point of view, an environmental characterization fulfilling all the aforementioned criteria can only be obtained via satellite imagery (Fig. 1-4; Rellstab et al., 2015; Riginos et al., 2016). Datasets derived from remote sensing technology provide daily records covering several years to decades at a resolution up to one kilometre (EU Copernicus Marine Service,

Figure 1-4. The *seascape* behind seascape genomics. Each map shows an example of an environmental gradient across the seascape of New Caledonia derived from remote sensing products: standard deviation of sea surface temperature (a), sea surface salinity (b) and sea surface velocity (c). Land is represented in grey, coral reef in black.

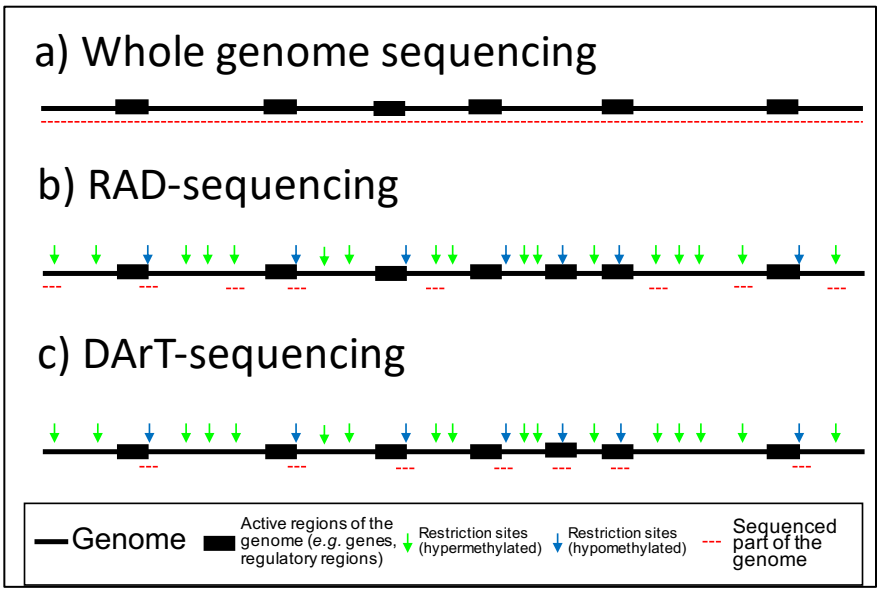


2017). Moreover, these data are available at a global scale and are generally free. Among the variables available under this format are sea surface temperature, sea surface salinity, chlorophyll concentration, sea water velocity, suspended particulate matter and photosynthetic available radiations (EU Copernicus Marine Service, 2017; NASA, 2016).
Remote sensing derived datasets are usually available with different levels of processing (EU Copernicus Marine Service, 2017). Sea surface temperature, for instance, can be obtained at the rawest levels (known as levels 2-3) where pixel values are produced by instantaneous satellite captures; the obtained resolution is high (1 km) but the image contains gaps due to satellite position and cloud cover (Dash et al., 2012). Conversely, more elaborated products (known as *level 4*) use algorithms to impute missing pixels by aggregating captures and refining the measures with *in situ* data; the resulting dataset has a lower spatial resolution (5 km), a lower temporal resolution (daily, weekly), but is gap free (Dash et al., 2012; EU Copernicus Marine Service, 2017). *The European Space Agency Sea Surface Temperature Climate Change Initiative* is among the most complete level 4 datasets available, and has delivered daily records starting from 1981 at a global extent with a resolution of 5 km (Merchant et al., 2019). A valuable alternative is the *Global high-resolution sea surface temperature* dataset, which combines different satellite inputs to obtain a level 4 product that covers a shorter temporal window (since 2002) but at higher spatial resolution (1 km; Chin, Vazquez-Cuervo, & Armstrong, 2017).

1.2.3. Genomics of marine organisms

Most marine species are considered non-model organisms in bioinformatics (Riginos et al., 2016). Consequently, genomic analyses dispose of a limited range of methodologies and resources, compared with well-studied land species (*e.g.* livestock species; The IMAGE Consortium, 2019). In seascape genomics (as in many fields of bioinformatics), the genotyping strategy is a compromise between two contrasting needs: 1) high number of genomic markers and 2) low cost per sample (Rellstab et al., 2015).
On the one hand, high numbers of genomic markers are necessary as genetic variants implicated in local adaptation represent only a small fraction of the genome (Luikart, England, Tallmon, Jordan, & Taberlet, 2003). On the other hand, low costs for genotyping are necessary because of the large sample sizes (often several hundreds of individuals) that are usually required in seascape genomics (Rellstab et al., 2015; Riginos et al., 2016). Single nucleotide polymorphisms (SNPs) are the most commonly used type of genetic markers in landscape and seascape genomics (Rellstab et al., 2015). This is because SNPs are relatively easy to obtain in large quantities, even in non-model species (Luikart et al., 2003). The cost associated with SNPs depends largely on the sequencing strategy (Fig. 1-5).
The ideal sequencing strategy for obtaining a large number of SNPs is unquestionably the whole genome sequencing (WGS) approach (Fig. 1-5a; Ekblom & Galindo, 2011). This method provides the most comprehensive catalogue of SNPs possible (theoretically all of the SNPs in

Figure 1-5. The *genomics* behind seascape genomics. The three panels display different sequencing methods for genotyping single nucleotide polymorphisms (SNPs). In whole genome sequencing (a), the entire genome is genotyped. In restriction site associated DNA (RAD) sequencing (b), genotyping only involves the genomic fragments in proximity of restriction sites randomly occurring across the genome (blue and green arrows). In diversity array technology (DArT) sequencing (c), genotyping concerns genomic fragments that are located next to restriction sites and hypomethylated (blue arrows). Hypomethylation is a genomic signature of actives regions, such as genes (black boxes) or regulatory sequences.



a genome), but is extremely demanding in terms of costs and computational effort, especially when dealing with a species never sequenced before (Davey et al., 2011; Ekblom & Galindo, 2011).

A valuable alternative is provided by sequencing strategies that reduce genome complexity (Ekblom & Galindo, 2011). These methods orientate the sequencing effort towards specific genomic regions, therefore lowering costs and computational requirements. Restriction sites associated DNA markers sequencing (RAD-seq; Fig. 1-5b) is a reduced genome complexity technique that is frequently applied to non-model species (Davey et al., 2011). The reasons behind its popularity are the limited costs and the little *a priori* knowledge about the studied species required (Ekblom & Galindo, 2011). The method consists of sequencing genomic stretches (~100 nucleotides each) adjacent to restriction sites. Restriction sites are short genomic sequences that occur at random positions throughout the genome, (*i.e.* next to genes, in intergenic regions, in genomic repetitions, etc.), but these positions are generally conserved between individuals. SNPs discovered using RAD-seq are therefore spread across the entire genome. Unfortunately, most of the sequences in an eukaryotic genome do not have a function (known or unknown; Palazzo & Gregory, 2014), thus most of the SNPs discovered via RAD-seq are not informative for adaptive processes (Lowry et al., 2017).

There are, however, variants of the RAD-seq approach that mitigate this issue. For example, in the diversity array technology sequencing (DArT-seq: Fig. 1-5c) restriction sites that have highly methylated nucleotides are excluded from sequencing (Kilian et al., 2012). Hypermethylation is frequent in genomic repetitions (*i.e.* regions with no or unknown function; Rabinowicz et al., 1999). Consequently, the set of SNPs discovered via DArT-seq is enriched with genes and active regions of the genome (*e.g.* regulatory sequences), and

therefore constitutes a more appropriate starting point for studying local adaptation (Gawroński et al., 2016).

Regardless of the sequencing strategy employed, the availability of a reference genome can be decisive for the interpretation of seascape genomics studies (Rellstab et al., 2015). DNA sequences carrying SNPs can be mapped onto a reference genome to determine their position (Ekblom & Galindo, 2011). Knowing the distance (linkage) between two SNPs allows for the evaluation of their independence from each other, while identifying the genes surrounding an adaptive SNP might indicate the potential molecular target of selection (Brodie, Azaria, & Ofran, 2016; Rellstab et al., 2015). Furthermore, the presence of a reference genome allows for the exclusion of all the SNP sequences that do not belong to the studied species. This is particularly important in corals, as a DNA sample of the coral holobiont can also contain sequences from the symbionts' genomes (Meyer & Weis, 2012). Over the last decade, an increasing number of coral genomes have been sequenced (*Acropora digitifera, Acropora millepora, Pocillopora damicornis, Stylophora pistillata, Galaxea fascicularis, Fungi sp., Goniastrea aspera* and *Porites lutea*; Cunning, Bay, Gillette, Baker, & Traylor-Knowles, 2018; Fuller et al., 2019; Robbins et al., 2019; Shinzato et al., 2011; Voolstra et al., 2017; Ying et al., 2018), as well as Symbiodiniacea genomes (Aranda et al., 2016; Lin et al., 2015; Shoguchi et al., 2018, 2013).

1.2.4. Genotype-environment association analysis

The genotype-environment association analysis is core to landscape genomic studies (Rellstab et al., 2015). There is a wide range of software that have been developed to perform this task, differing in the statistical approaches used and in how genomic data (*e.g.* single-locus genotypes vs. multiple-loci genotypes) and environmental information (*e.g.* unique gradients vs. composite gradients) are handled (Balkenhol et al., 2017; Rellstab et al., 2015). In addition, each software uses specific methods to account for the problem related to the confounding role of neutral genetic structure (further discussed in section 1.2.6. ; Rellstab et al., 2015). Hereunder are highlighted the main features of two software intensively used in landscape and seascape genomics: SamBada and LFMM.

SamBada is a software that calculates genotype-environment associations using logistic regression models (Joost et al., 2007; Stucki et al., 2017). Each genotype-environment combination is handled independently, even though associations with multiple environmental variables are possible (Stucki et al., 2017). To cope with the issue of population structure, variables describing neutral genetic variation (*e.g.* principal components of the genotypes matrix; Novembre et al., 2008) can be included as co-variables in the statistical model. The strengths of SamBada are its computational speed and ease of use, particularly as it is implemented in an R package, thus facilitating the processing of genomic and environmental input data (Duruz et al., 2019).

LFMM stands for a latent factor mixed models, which is the statistical method used for describing the genotype-environment associations (Frichot, Schoville, Bouchard, & François, 2013). Similar to SamBada, genotype-environment combinations are handled independently, but with substantial differences in the way models are estimated (Frichot et al., 2013; Rellstab et al., 2015). In LFMM, neutral genetic variation is introduced as random factor (latent factor) and the model parameters are estimated using a stochastic (Markov Chain Monte Carlo) algorithm (Frichot et al., 2013). The main consequence is that multiple runs are required and therefore computational speed is slower than in SamBada (Stucki et al., 2017). On the other hand, the software has shown to be robust under a variety of demographic models (Lotterhos & Whitlock, 2015) and is also implemented in a R package (Frichot & François, 2015).

1.2.5. Strengths of seascape genomics

Seascape genomics constitutes and exploratory method that studies the local adaptation in such a way that is substantially different from traditional experimental approaches (Rellstab et al., 2015; Riginos et al., 2016). The main particularities are that (1) the seascape is used as a natural experimental set-up and that (2) genomic analyses are employed to investigate the presence of adaptation (Riginos et al., 2016). In contrast, in traditional approaches such as aquarium experiments, the selective constraint (*e.g.* temperature variation) is set by the researcher and presence of adaptation is usually estimated out of physiological responses (*e.g.* bleached vs. non-bleached; Krueger et al., 2017; Pardo-Diaz et al., 2015).

In seascape genomics, sampling occurs at multiple sites, which portrays a variety of adaptive responses existing in a population (Rellstab et al., 2015; Riginos et al., 2016). Is it therefore possible (and recommended) to study local adaptation to multiple environmental variables, or even against the same variables at different spatial and temporal resolutions (Leempoel et al., 2017; Rellstab et al., 2015). In traditional experiments, usually only one kind of environmental stress can be tested. In addition, seascape genomics allows for the creation of surrogate variables representing selective constraints that are hard to reproduce under an experimental set-up, such as the disturbances due to the proximity to densely populated areas, boat traffic or touristic regions (Leempoel et al., 2017; Riginos et al., 2016).

The *genomic* component of the seascape genomics approach confers two advantages in comparison with traditional experimental approaches. The first is that it provides the bases for formulating hypotheses on the molecular side of adaptation, and therefore disclose a secondary axis of investigation (Pardo-Diaz et al., 2015; Rellstab et al., 2015). In traditional experiments, this axis can be investigated only through additional analyses (*e.g.* transcriptome profiling; Pardo-Diaz et al., 2015). The second is that sampling tissues for DNA analyses does not require expensive equipment nor complicated protocols, and is usually non-invasive (Carroll et al., 2018). This is particularly important for sampling permits, especially when working with endangered animals as corals. Traditional experiments, in

contrast, usually require the sacrifice of the sampled individuals, and are therefore subjected to more rigid regulations.

Compared with traditional experimental approaches, seascape genomic studies use sampling at multiple sites covering ecologically meaningful scales (Leempoel et al., 2017). This spatially broad view of local adaptation is advantageous when transposing results in a conservation perspective (Forester, Landguth, Hand, & Balkenhol, 2018). For instance, it is possible to make predictions on the presence of potentially adaptive genotypes across the whole study area, even in locations never visited during sampling (Rochat & Joost, 2019). Furthermore, seascape genomics studies are inextricably connected with population genomics analyses (Balkenhol et al., 2017), since adaptive genomic variation cannot be investigated unless the neutral genetic variation is identified. The advantage here is that seascape genomics produces a collateral information on population structure, which represents another valuable insight for conservation (Palumbi, 2003).

1.2.6. Weaknesses of seascape genomics

The main weakness of seascape genomics, and more generally of the landscape genomics approach, are the false discoveries (Rellstab et al., 2015). These can occur at high rates when the characterization of the seascape and its population lacks of crucial information (Leempoel et al., 2017). For instance, an environmental variable relevant for adaptation might be missing, and genotypes might be mistakenly associated with another variable collinear to the missing selective constraint. The main hurdle, however, concerns neutral population structure. In fact, most natural populations display heterogenous spatial patterns of genetic diversity, mainly due to demographic processes (Lotterhos & Whitlock, 2015; Novembre et al., 2008). Such patterns can be superposed with environmental gradients, and therefore cause the misleading association of neutral genotypes with a climatic constraint (Rellstab et al., 2015). This problem is particularly relevant to seascape genomics as sea currents simultaneously shape environmental gradients and drive migration and dispersal (Riginos et al., 2016).

In addition, because of the large sample sizes required and the general lack of genomic resources for marine species, seascape genomic experiments usually employ cost-effective genotyping strategies such as those that reduce genome complexity (Rellstab et al., 2015). These genotyping strategies have limited genomic resolution, which complicates the interpretation of the molecular side of an adaptive signal. For instance, a SNP found to be associated with an environmental gradient is not likely to be the mutation providing the selective advantage itself (the causative mutation), but a SNP that is physically close (Lowry et al., 2017; Pardo-Diaz et al., 2015). The search for the causative mutation will therefore cover a relatively large genomic window (several kilobases), encompassing different genes with distinct functions (Rellstab et al., 2015). In this kaleidoscope of genomic annotations, it is often difficult for seascape genomic results to be conclusive, and further validation via

complementary experimental frameworks (*e.g.* aquaria experiments) is usually recommended (Pardo-Diaz et al., 2015).
Sampling strategies could mitigate some of the weaknesses mentioned above (Leempoel et al., 2017; Manel, Albert, & Yoccoz, 2012). For instance, the location of the sampling sites could be optimized to maximize environmental contrasts, to avoid measured gradients to be collinear or to anticipate the confounding effect of population structure (Riginos et al., 2016). A rational sampling strategy might also identify uninformative sampling locations (for instance, those with redundant environmental conditions), reduce the costs of sampling and thus free up budgets to increase the genomic resolution of the genotyping strategy. While theoretical insights on optimizing sampling strategies exist, practical guidelines that demonstrate improved statistical power are rare (De Mita et al., 2013; Lotterhos & Whitlock, 2015; Manel et al., 2012). How many samples to collect, how many sampling sites to establish and how to distribute sites across the study area are open questions that landscape and seascape genomics researchers are recurrently confronted with (Rellstab et al., 2015).

1.3. Aim of the thesis

As coral bleaching is resulting in the death of reefs worldwide, characterizing corals' adaptive potential is now of paramount importance. This characterization should be considered under a conservation perspective, so that effective conservation strategies can be organized accordingly. Seascape genomics can contribute to fulfilling these needs. Thus, the main research goal of this thesis is to **apply the seascape genomics approach to corals to empower coral reef conservation strategies.**
Among the strengths of the seascape genomics approach, we find the interdisciplinarity between population-level and molecular-level analyses. This thesis takes advantage of this aspect by undertaking **two parallel axes of research for coral adaptation against heat stress**. The first axis focuses on the **molecular side of adaptation** and investigates which of the genes participating in coral responses against heat stress are involved in evolutionary processes. The second axis concerns **adaptive potential at the population scale and its implications for conservation**. The goal here is the creation of objective indices that identify, within a given region, reefs that are more or less susceptible to host corals adapted to heat stress.
Landscape and seascape genomics are evolving disciplines with considerable margins for improvement. One critical point is the sampling strategy. The present research tackled this issue by **defining practical guidelines to optimize the sampling strategy in landscape genomics**.

Globally, the goals can be summarized by the following four research questions:

1. How does the sampling strategy drive statistical power of landscape genomics analyses?
2. Can seascape genomics characterize the adaptive potential of corals against heat stress?
3. What are the molecular mechanisms implicated in heat stress adaptation in coral?
4. How can information on coral adaptive potential support reef conservation?

1.4. Layout of the thesis

This thesis is composed of four scientific articles, either already published or submitted for publication in peer-reviewed journals. The first two articles are *in silico* investigations that are preparatory to the research performed in articles three and four, in the frame of the SABLE project (see section 1.4.2 and Figs. 1-6).

1.4.1. Preparatory studies

For the first two articles (A and B), input data used for the analyses came from computer simulations (article A) or from re-analysis of previously published datasets (article B). **Article A is entitled "Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach".** It focuses on sampling strategy, and on how statistical power of landscape genomics can be driven by sample size and by the number and spatial position of sampling sites. To achieve this, we elaborated a simulative framework to create artificial landscape genomic datasets, each differing by sampling strategy decisions. Each simulated dataset underwent a landscape genomics analysis followed by an evaluation of statistical power and false discoveries. This work defined practical guidelines to rationalize the decisions on how many samples to collect and where from. Although this article is not specific to corals, it provides indicators that are applicable to landscape or seascape genomics on any species, and this information was necessary for the work conducted in the context of the SABLE project.
Article B is entitled "Seascape genomics as a new tool to empower coral reef conservation strategies: An example on north-western Pacific *Acropora digitifera*". The aim of this study was to test the application of seascape genomics to corals. The dataset used was taken from an existing publication, characterizing the genetic structure of a coral population in the Ryukyu Archipelago (Japan; Shinzato, Mungpakdee, Arakaki, & Satoh, 2015). We ran a seascape genomics study combining the genomic data from this publication with the environmental descriptors derived from satellite imagery. The adaptive signals detected were used to predict the probability of heat stress adaptation for corals of the region. In addition, we employed remote sensing data of sea surface currents to evaluate connectivity between reefs of the study area. Predictions on adaptation and connectivity were synthesized in

objective indices for conservation. This work was the core basis of the thesis as it provided the *proof-of-concept* for the application of seascape genomics to corals.

1.4.2. The SABLE project

SABLE is a project funded by the United Nations Environment Programme (UNEP) and the International Coral Reef Initiative (ICRI; Fig. 1-6). SABLE results from a scientific collaboration between EPFL and the joint research unit ENTROPIE[1] at the Institute of research for development (IRD) of Nouméa (New Caledonia). The goal of SABLE was to apply a seascape genomics approach to characterize the adaptive potential of three coral species of New Caledonia (Southwestern Pacific), and to use these findings to support reef conservation strategies.

In 2018, a three-month long field campaign was organized to collect coral samples for seascape genomics (Box. 1-1). The sampling plan was designed following the guidelines defined in the preparatory article A. Seascape genomics analyses were then performed, following the methods defined in article B. The results of the SABLE analyses were presented in two additional articles that developed the seascape genomics approach by focusing on two related axes of research: the molecular basis of adaptation (article C) and the implications for conservation (article D).

Article C is entitled "Seascape genomics reveals candidate molecular targets of heat stress adaptation in three coral species". This research applied the seascape genomics approach to identify SNPs potentially driving adaptation against heat stress in three coral species. We characterized the genomic neighbourhood of heat-stress associated SNPs to investigate whether genes or the corresponding molecular functions recurred within and between species. This research highlighted certain molecules that are candidate molecular targets for heat-stress adaptation, some of which participate with well-established cellular pathways of heat stress resistance in corals.

Article D is entitled "*Coral cover surveys corroborate predictions on reef adaptive potential to thermal stress*". This work applied the indices of connectivity and probability of adaptation to heat stress developed in article B to the case-study of New Caledonia. We then compared current conservation strategies (in particular, MPAs) with the conservation priorities highlighted by seascape genomics. In addition, we retrieved field surveys describing over 15 years of changes in living coral cover across the Archipelago. The combined analysis of field surveys and environmental conditions showed an association between heat stress and decrease of coral cover. However, the analysis indicated that the strength of this association was mitigated at reefs predicted to show a high probability of adaptation to heat stress, and

[1] The ENTROPIE laboratory (Ecologie Marine Tropicale des Océans Pacifique et Indien) is a French joint research unit backed by French Institute of Research for Development – IRD, Centre National de la Recherche Scientifique – CNRS, and University of Reunion Island – UR.

at those expected to receive larger amounts of incoming favourable propagules via sea currents.

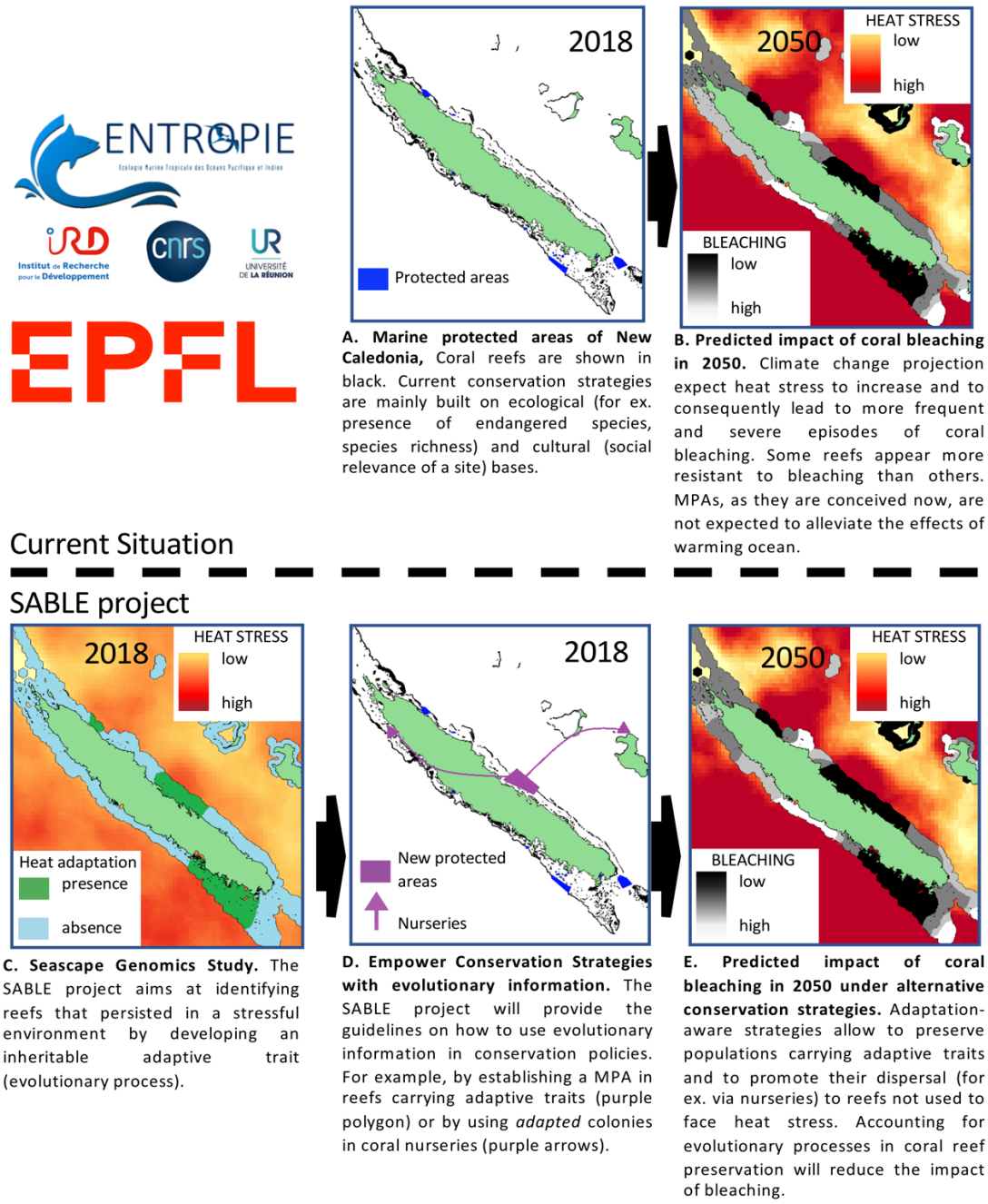
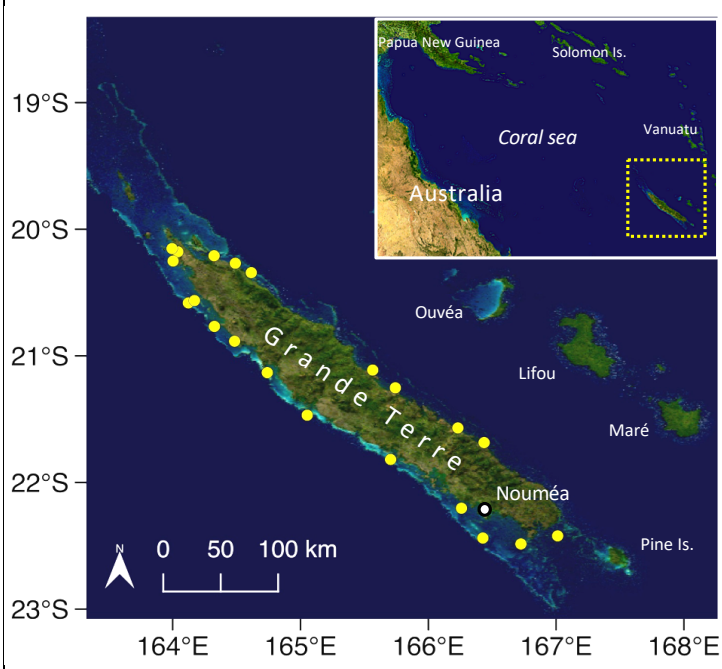


Figure 1-6. A Seascape genomics Approach to improve coral reefs conservation strategies against BLEaching (SABLE). In 2017, the International Coral Reef Initiative (ICRI) and the United Nations Environment Program (UNEP) launched a call for proposals for projects dedicated to the innovation in coral reef management strategies. The SABLE project was one of the five selected for funding out of 233 proposals. SABLE was an initiative of the joint research unit ENTROPIE and the EPFL. SABLE focused on the coral reefs of New Caledonia, listed as UNESCO world heritage sites and featuring the world's second largest barrier reef, after the Australian Great Barrier Reef. The goal of SABLE was to use the seascape genomics approach to characterize the adaptive potential of flagship corals species from New Caledonia, and to use this information to support conservation strategies. The diagram here below displays the theory of change underlying the SABLE project. Note that this figure is for illustrative purposes and does not contain real data.

Box 1-1. The field campaign in New Caledonia.



Map of New Caledonia, with sampling sites of the SABLE project in yellow.

New Caledonia is a special collectivity of France in the Southern Pacific. The capital Nouméa hosts roughly two thirds of the ~270,000 inhabitants. The rest of the territory is considerably less densely occupied and is prevalently inhabited by the indigenous Kanak population.

Grande Terre is the main island of the Archipelago and is enclosed by more than 1,000 km of barrier reefs, the second world largest after the Australian Great Barrier Reef. Because of its pristine state, the coral reef of New Caledonia is listed as UNESCO World Heritage.

Whitetip reef shark in a coral reef of the Southern Lagoon of Grande Terre.

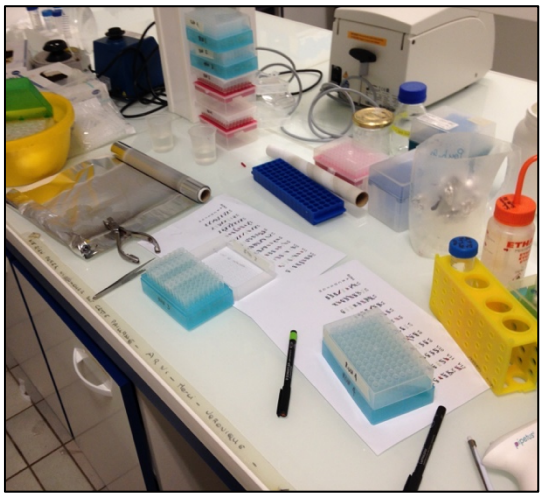
Coral reef in the Southern Lagoon of Grande Terre.

Field crew of the SABLE project. From left to right: V. Berteaux-Lecellier, G. Mou-Tham, M. Clarque, C. Peignon, myself, J. Baly and G. Lecellier.

In February 2018 I travelled to New Caledonia to participate to the field campaign of the SABLE project. My stay lasted three months, during which we organized the sampling expeditions from the headquarters of the French Institute of Research for Development (IRD) in Nouméa. In total, we sampled corals from twenty distinct reefs surrounding Grande Terre. At every sampling site, two members of the team dived in search of corals belonging to the species of interest: *Acropora millepora* and *Pocillopora damicornis* and *Pocillopora acuta.* Once found a colony of interest, the divers took a photograph and then snapped a branch of ~1 cm length. The sample was then immediately brought to the boat, where two members of the crew transferred it into a labelled tube filled with 80% ethanol and stored it on ice. In total, we collected 800 coral samples that underwent DNA purification at the IRD laboratories in Nouméa at the end of the field campaign.

Diver collecting an Acropora millepora sample.

Field crew at work to launch the boat in Thio.



Laboratory benchtop during DNA purification

Dinner at Kanak homestay accommodation in Tiabet.

Some of the sampling sites were reached in single-day boat expeditions from Nouméa, but most often the sampling expeditions lasted several days and required the overnight stay in hotels or homestay accommodations along the way. These expeditions went beyond the scientific purposes of the campaign, as we could witness the hospitality of the inhabitants of New Caledonia. At the end of the sampling campaign we had travelled for roughly 3,000 km across Grande Terre and navigated for 750 km inside the lagoon.

Lunch break on the east coast of Grande Terre.

Sunset in Nouméa.

[photo credits: Oliver Selmoni]

2. Article A

Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach

Oliver Selmoni[1], Elia Vajana[1], Annie Guillaume[1], Estelle Rochat[1] and Stéphane Joost[1].

[1]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

Postprint version of the article published in *Molecular Ecology Resources:*
Selmoni, O., Vajana, E., Guillaume, A., Rochat, E., and Joost, S. (2020). Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach. Mol. Ecol. Resour. 20. doi:10.1111/1755-0998.13095.

Contribution of the candidate:
As first author of this article, I initiated the research, developed the codes for running the simulations and wrote the first version of the manuscript. The co-authors provided advice concerning the methods used and critically revised the manuscript before submission.

During the preparation of the field campaign for the SABLE project, some questions recurrently returned to the table: "how many colonies should be sampled per species?", "how many colonies should be sampled per site?", "how many sampling sites should be visited?" and "how do we choose the location of sampling sites?". The answers to these questions referred to the general rules of thumb based on previous work, and to theoretical guidelines defining the good practices in sampling strategies (e.g. "stratify sampling across the climatic and/or biological spaces"; Manel et al., 2012). These recommendations, however, were not proof-based.
Two works (De Mita et al., 2013; Lotterhos & Whitlock, 2015) tackled the issue of testing the efficiency of different sampling strategies in genotype-environment association studies using a simulative approach. The simulative approach consists in computing artificial populations, controlled by precise demographic (e.g. migration rate) and evolutionary parameters (e.g. selection strength). Next, different sampling strategies are re-iterated for these populations and the sampled genotypes undergo the landscape genomics analysis. The goal of this workflow is to evaluate the effect of changes in sampling strategy on statistical power and false discoveries. The main drawback of these studies was that they accounted for a unique environmental variable. This is problematic as collinearity between environmental gradients is one of the main sources of false discoveries in landscape genomics analyses (Leempoel et al., 2017). *Furthermore, methods used to optimize the sampling design along a single gradient*

are hard to transpose to a multivariate environment. The aim of the article presented hereunder was to cope with these issues.
The main challenge laid in computing artificial populations counting thousands of individuals, characterised by thousands of genetic variants, controlled by specific demographic dynamics and exposed to multiple environmental constraints. Even when using state-of-the-art software for demographic simulations (e.g. CDPOP, Landguth & Cushman, 2010), the computational requirements would have been prohibitive. This is why we opted for a heuristic solution: instead of computing genotypes by simulating the evolution of a population generation after generation, we developed a probabilistic framework to infer genotypes based on pre-established demographic and environmental gradients. This approach required some approximations (as every heuristic approach) but allowed for testing sampling strategies on thousands of distinct landscape genomic datasets.
This article describes practical guidelines for choosing the appropriate sample size and number of sampling sites depending on the demographic scenario expected in the population of interest. Furthermore, this work proposes two methods to optimize sampling across a multivariate landscape, one of which is also conceived to anticipate the confounding effects of population structure. The methods presented here were transposed to a seascape genomics perspective and used to plan the sampling strategy for the SABLE project (article C).

2.1. Abstract

An increasing number of studies are using landscape genomics to investigate local adaptation in wild and domestic populations. The implementation of this approach requires the sampling phase to consider the complexity of environmental settings and the burden of logistic constraints. These important aspects are often underestimated in the literature dedicated to sampling strategies.
In this study, we computed simulated genomic datasets to run against actual environmental data in order to trial landscape genomics experiments under distinct sampling strategies. These strategies differed by design approach (to enhance environmental and/or geographic representativeness at study sites), number of sampling locations and sample sizes. We then evaluated how these elements affected statistical performances (power and false discoveries) under two antithetical demographic scenarios.
Our results highlight the importance of selecting an appropriate sample size, which should be modified based on the demographic characteristics of the studied population. For species with limited dispersal, sample sizes above 200 units are generally sufficient to detect most adaptive signals, while in random mating populations this threshold should be increased to 400 units. Furthermore, we describe a design approach that maximizes both environmental and geographical representativeness of sampling sites and show how it systematically outperforms random or regular sampling schemes. Finally, we show that although having

more sampling locations (between 40 and 50 sites) increase statistical power and reduce false discovery rate, similar results can be achieved with a moderate number of sites (20 sites). Overall, this study provides valuable guidelines for optimizing sampling strategies for landscape genomics experiments.

2.2. Introduction

Landscape genomics is a subfield of population genomics, with the aim of identifying genetic variation underlying local adaptation in natural and managed populations (Balkenhol et al., 2017; Joost et al., 2007; Rellstab et al., 2015). The approach consists of analyzing genomic diversity and environmental variability simultaneously in order to detect genetic variants associated with a specific landscape composition. Studies of this kind usually incorporate an analysis of population structure, such that neutral genetic variation can be distinguished from adaptive variation (Rellstab et al., 2015). Over the last few years, the landscape genomic approach is becoming more widely used (see Tab. 2-1; Balkenhol et al., 2017; Rellstab et al., 2015). It is being applied to a range of species, including livestock (Colli et al., 2014; Lv et al., 2014; Pariset, Joost, Marsan, & Valentini, 2009; Stucki et al., 2017; Vajana et al., 2018), wild animals (Harris & Munshi-South, 2017; Manthey & Moyle, 2015; Stronen et al., 2015; Wenzel, Douglas, James, Redpath, & Piertney, 2016), insects (Crossley et al., 2017; Dudaniec et al., 2018; Theodorou et al., 2018), plants (Abebe et al., 2015; De Kort et al., 2014; Pluess et al., 2016; Yoder et al., 2014) and aquatic organisms (DiBattista et al., 2017; Hecht, Matala, Hess, & Narum, 2015; Laporte et al., 2016; Riginos, Crandall, Liggins, Bongaerts, & Treml, 2016a; Vincent, Dionne, Kent, Lien, & Bernatchez, 2013).

Sampling strategy plays a pivotal role in experimental research, and must be theoretically tailored to the aim(s) of a study (Rellstab et al., 2015; Riginos et al., 2016). In the context of landscape genomics, the sampling design should cover a spatial scale representative of both the demographic processes and the environmental variability experienced by the study population (Balkenhol et al., 2017; Leempoel et al., 2017; Manel et al., 2010; Rellstab et al., 2015). This is imperative to be able to properly account for the confounding effect of population structure, to provide a biologically meaningful contrast between the environmental variables of interest and to definitely allow the search for actual adaptive variants (Balkenhol et al., 2017; Manel et al., 2010; Rellstab et al., 2015). Consequently, extensive field sampling is generally required and needs to be coupled with high-throughput genome sequencing to characterize samples at a high number of loci (Balkenhol et al., 2017; Rellstab et al., 2015). Beyond these theoretical aspects, pragmatic choices need to be made with regards to financial and logistic constraints that are often imposed (Manel et al., 2010; Rellstab et al., 2015). A sampling strategy is constituted of: i) sampling design (the spatial arrangement of the sampling locations, D); ii) the number of sampling locations (L); and iii)

sample size (the number of individuals sampled, N; Tab. 2-1). The care with which these parameters are defined affects the scientific output of an experiment as well as its costs (Manel et al., 2010; Rellstab et al., 2015).

The landscape genomics community has traditionally focused on formulating theoretical guidelines for collecting individuals throughout the study area. In this literature, particular emphasis has been placed on how spatial scales and environmental variation should be accounted for when selecting sampling sites (Leempoel et al., 2017; Manel et al., 2012, 2010;

Table 2-1. Sampling design in landscape genomics studies. A non-exhaustive list of landscape genomics studies, highlighting different species and their related sampling strategies.

* Numbers from the Vincent et al. report (2013) concerning the *non-pooled* samples.

Study	Species	Sampling Design (D)	Sampling Locations (L)	Sample Size (S)
Colli et al. 2014	Goat	Spatial and breed representativeness	10 sites	43
Pariset et al. 2009	Goat	Spatial and breed representativeness	16 regions	497
Stucki et al. 2017, Vajana et al. 2018	Cattle	Spatial representativeness	51 regions	813
Harris and Munshi-South, 2017	White-footed Mouse	Habitat representativeness	6 sites	48
Stronen et al., 2015	Wolf	Opportunistic, population representativeness	59 sites	59
Wenzel et al., 2016	Red Grouse	Spatial representativeness	21 sites	231
Crossley et al., 2017	Potato Beetle	Habitat representativeness	16 sites	192
Dudaniec et al., 2018	Damselfly	Environmental and spatial representativeness	25 sites	426
Theodorou et al., 2018	Red-tailed bumblebee	Habitat representativeness	18 sites	198
Abebe et al., 2015	Barley	Spatial representativeness	10 regions	260
De Kort et al., 2014	Black alder	Spatial and habitat representativeness	24 populations	356
Pluess et al., 2016	Eurpean beech	Spatial and environmental representativeness	79 populations	234
Yoder et al., 2014	Barrelclover	Spatial representativeness	202 sites	202
DiBattista et al., 2017	Stripey Snapper	Spatial representativeness	51 sites	1,016
Hecht et al., 2015	Chinook salmon	Spatial representativeness	53 sites	1,956
Laporte et al., 2016	Eurpeal Eel	Spatial and environmental representativeness	8 sites	179
Vincent et al., 2013	Atlantic Salmon	Spatial representativeness	26* rivers	641*

Rellstab et al., 2015; Riginos et al., 2016). Theoretical simulations have shown that performing transects along environmental gradients or sampling pairs from contrasting sites which are spatially close reduced false discovery rates caused by demographic processes confounding effects (De Mita et al., 2013; Lotterhos & Whitlock, 2015). However, in these studies the environment was described using a single variable, which oversimplifies the choice of sampling sites. In fact, in a real landscape genomics application, several variables are usually analyzed in order to explore a variety of possible environmental pressures causing selection (Balkenhol et al., 2017). The concurrent use of several environmental descriptors also allows to control for the bias associated with collinear conditions (Rellstab et al., 2015). Furthermore, these studies focused on the comparison of different statistical methods with the drawback of confronting only a few combinations of the elements determining the sampling strategy (De Mita et al., 2013; Lotterhos & Whitlock, 2015). Last but not least, the number of samples used in the simulations (between 540 and 1800; Lotterhos & Whitlock, 2015) appear to be unrealistic for use in most of real landscape genomic experiments (Tab. 2-1) and thus the guidelines proposed are scarcely applicable in practice.

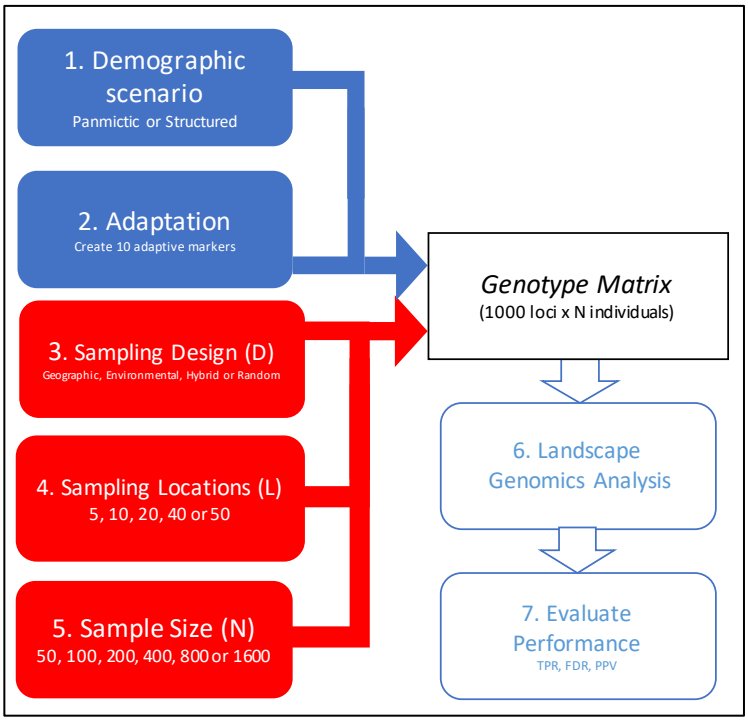
For these reasons, there is a need to identify pragmatic and realistic guidelines such that a sampling strategy is designed to maximize statistical power, minimize false discoveries, and optimize efforts and money expenses (Balkenhol et al., 2017; Rellstab et al., 2015). In particular, the fundamental questions that need to be addressed are: i) how to determine the spatial arrangement of sampling locations; ii) how to organize sampling effort (for instance preferring many samples at few sites, or rather fewer samples at many sites); and iii) how many samples are required to obtain sufficient statistical power (Rellstab et al., 2015; Riginos et al., 2016).

In this paper, we investigate how the outcome of landscape genomic analyses is driven by the sampling strategy. We ran simulations using a fictive genetic dataset encompassing adaptive genotypes shaped by real environmental variables. The simulations accounted for antithetic demographic scenarios encompassing strong or weak population structure. We proposed sampling strategies that differed according to three elements: sampling design approach (D), number of sampling locations (L) and sample size (number of samples, N). For each of these three elements, we measured their relative impacts on the analyses' true positive rates (TPR) and false discovery rates (FDR), as well as their impact on the predictive positive value (PPV; Marshall, 1989) of the strongest adaptive signals.

2.3. Material and Methods

The iterative approach we designed to test the different sampling strategies required that a new genetic dataset encompassing neutral and adaptive variation was created at every run of the simulations. A simulated genomic dataset can be constructed by means of software

Figure 2-1. Workflow for each iteration of the simulative approach. The seven steps taken for every iteration. Starting with the blue boxes, the genetic set-up is established by selecting the demographic scenario (panmictic or structured), which determines the neutral structure, and by picking the environmental variables implied in adaptation. The environmental variable of interest and the strength of selection is randomly sampled for each of the 10 adaptive markers. Following this, the sampling strategy (here shown with red boxes) is set as a combination of design approach (geographic, environmental, hybrid or random), number of sampling locations (5, 10, 20, 40 or 50 locations) and sample size (50, 100, 200, 400, 800 or 1600 samples). This results in the creation of a genotype matrix that undergoes a landscape genomics analysis. At the end of iterations, statistical power (TPR) and false discovery rate (FDR) of the analysis and statistical predictive positive value of the strongest associations (PPV) are calculated to assess the performance of the sampling strategy.



performing coalescent (backward-in-time) or forward-in-time simulations (Carvajal-Rodríguez, 2008). However, methods using coalescent simulations (for ex. SPLATCHE2; Ray, Currat, Foll, & Excoffier, 2010) did not match our needs as they cannot compute complex selective scenarios (for instance those involving multiple environmental variables; Carvajal-Rodríguez, 2008). We could not use forward-in-time methods either, as they are slow and therefore not compatible with the computational requirements of our simulative approach (Carvajal-Rodríguez, 2008). For these reasons, we developed a customized framework in the R environment (version 3.3.1; R Core Team, 2016) to compute both neutral and adaptive genetic variation based on gradients of population membership and environmental variations, respectively (Fig. 2-1).

Prior to running the simulations across the complete dataset (the multivariate environmental landscape of Europe), we tested our approach on a reduced dataset and compared it to a well-established forward-in-time simulation software (CDPOP, version 1.3; Landguth & Cushman, 2010). This step allowed us to define the optimal parameters required to simulate two types of demographic scenarios: panmictic (no dispersal constraints, random mating) and structured (dispersal and mating limited by distance).

We then proceeded with the simulations on the environmental dataset of Europe. At each iteration, a new genetic background encompassing neutral and adaptive variation was computed (Fig. 2-1 steps 1 and 2). Subsequently, a sampling strategy was applied as a combination of sampling design (D), number of sampling locations (L) and sample size (N)

(Fig. 2-1, steps 3, 4 and 5), resulting in the generation of a genetic dataset that, coupled with environmental data, underwent a landscape genomics analysis (Fig. 2-1, step 6). At the end of each iteration, three diagnostic parameters were calculated: true positive rate (TPR, *i.e.* statistical power) and false discovery rate (FDR) for the analysis, as well as the predictive positive value (PPV) of the strongest genotype-environment associations (Fig. 2-1, step 7).
At the end of the simulations, we analyzed how each element of sampling strategy (D, L, N) affected the rates of the three diagnostic parameters (TPR, FDR, PPV) under the two demographic scenarios (with or without dispersal constraints). All scripts and data used to perform this analysis are publicly available on Dryad (doi:10.5061/dryad.m16d23c).

2.3.1. Environmental data

As a base for our simulations, we quantified the environmental settings of Europe (Fig. S2-1). We retrieved eight climatic variables from publicly available sources (annual mean temperature, mean diurnal range, temperature seasonality, mean temperature of wettest quarter , annual precipitation, precipitation seasonality, precipitation of warmest quarter and altitude; Tab. S2-1; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005; Ryan et al., 2009). In order to work on a relevant geographical scale (Leempoel et al., 2017) while maintaining an acceptable computational speed, the landscape was discretized into grid cells of 50x50 km, using QGIS toolbox (version 2.18.13; QGIS development team, 2009). This resulted in 8,155 landscape sites. Average values of environmental variables were computed for each cell of the landscape using the QGIS zonal statistics tool.

2.3.2. Computation of genotypes

For the creation of the genotype matrices, we developed an R-pipeline based on probability functions to compute genotypes from population membership coefficients and environmental values (Box S2-1). The theoretical fundaments of this method are based on the observation that when the population is structured, neutral alleles tend to show similar spatial patterns of distribution (a feature commonly exploited in Fst outlier tests; Luikart et al., 2003; and principal component analyses of genotype matrices; Novembre et al., 2008). Conversely, when a marker is under selection, its genotypic/allelic frequencies correlate with the environmental variable of interest (this is the basic concept of Landscape Genomics; see Balkenhol et al., 2017). For every iteration, 1,000 loci are computed: 10 are set to “adaptive”, while the remaining 990 to “neutral”. They are computed as follows:

- Neutral markers (Box S2-1a): a parameter (*m)* is set to define the number of population membership gradients used in the simulations, where higher values of *m* result in more complex population structures. Every population membership gradient is simulated by randomly picking one to five landscape locations to represent the center of the gradient. For each landscape location, the geographical distance to the gradient centers (calculated

using the R *dist* function) constitutes the membership coefficient. Next, a linear transformation converts this coefficient (Fig. S2-1) for each sampling site into the probability of carrying a private allele for the population described ($pA|PS$). A second parameter (c, Box. S2) define this transformation, with values between 0.5 (random population structure) and 0 (strong population structure). The probability of $pA|PS$ is then used to draw (using the R-stat *sample* function) the bi-allelic genotype for each individual. This procedure is re-iterated for every neutral locus assigned to a specific population membership coefficient. Each of the 990 neutral loci is then assigned to one of the m population membership coefficients (probability of assignment equal to $\frac{(1-c)}{\sum_{i=1}^{m}(1-c_i)}$) using the R *sample* function.

- Adaptive markers (Box S2-1b): the probability of carrying an adaptive allele ($pA|Env$) is calculated through a linear transformation of a specific environmental gradient. This transformation is defined by two parameters. The first parameter (s_1) determines the amplitude of the transformation, and ranges between 0 (strong selective response) and 0.5 (neutral response; Box S2-2). The second parameter (s_2) shifts the baseline for allele frequencies, and ranges between -0.2 and 0.2 (weakening and strengthening the selective response, respectively; Box S2-2). Each of the ten adaptive loci are randomly associated with one environmental variable. This implies that some environmental conditions can be associated with several genetic markers, while others with none. For every adaptive locus, the bi-allelic genotype is drawn (using the R-stat *sample* function) out of $pA|Env$.

2.3.3. Evolutionary scenarios and parametrization

Two distinct demographic scenarios were chosen for this study: one involving a population that is not genetically structured (hereafter referred to as the "panmictic population scenario"), and one involving a structured population (hereafter referred to as the "structured population scenario"; see Box S2-2). In order to define the values of parameters m, c, s_1 and s_2 that allow the production of these two demographic scenarios, we ran a comparison of our customized simulation framework against simulations obtained using a well-established forward-in-time simulation software for landscape genetics called CDPOP (version 1.3; Landguth & Cushman, 2010).

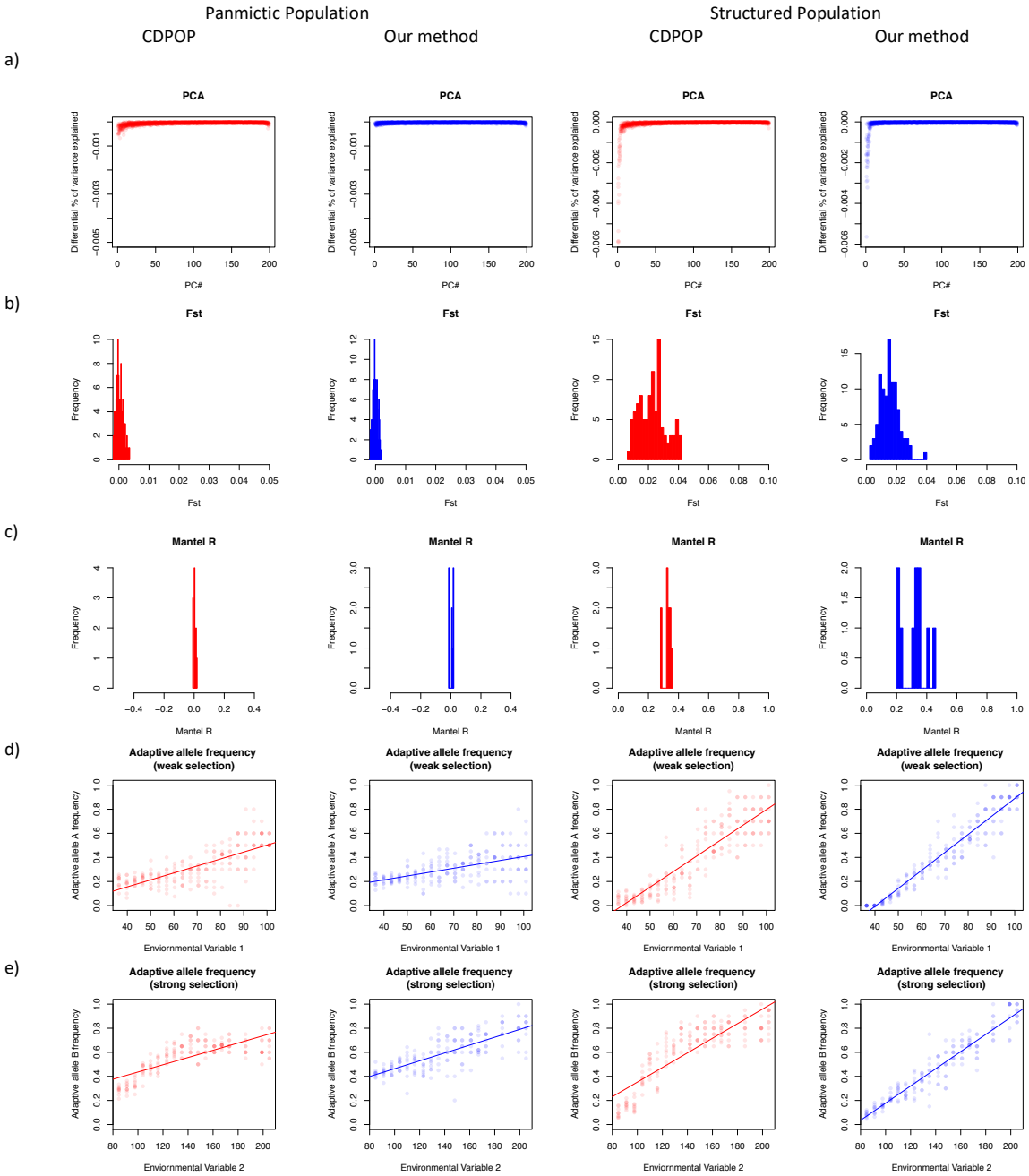
This comparison was performed on a reduced dataset composed of a 10-by-10 cell grid, covered with two dummy environmental variables extracted from the bioclim collection (Hijmans et al., 2005; Fig. S2-1a, b). Each cell could host up to 5 individuals, where each individual was characterized at 200 SNPs. In this set-up, we ran CDPOP using two distinct settings: the first that allowed for completely random dispersal and mating movements of individuals (*i.e.* panmictic population scenario), while the second setting restricted movements to neighboring cells using a dispersal-cost based on distance (*i.e.* structured population scenario). In both scenarios, we applied identical mortality constraints related to

the two environmental variables, and set for each of them a genetic variant modulating fitness (Fig. S2-1c, d). Fitness responses were constructed on an antagonistic pleiotropy model (*i.e.* adaptive tradeoffs, Lowry, 2012), using different intensities to represent moderate (Fig. S2-1c) and strong selective constraints (Fig. S2-1d). The following default CDPOP parameters were employed for the remaining settings: five age classes with no sex-specific mortality, reproduction was sexual and with replacement, no genetic mutations, epistatic effects or infections were allowed. The simulations ran for 100 generations and ten replicates per demographic scenario were computed.

In parallel, we ran our customized algorithm to compute genotypes, using the same simplified dataset as above. We iteratively tested all the possible combinations (hereafter referred to as "simulative variants") of the parameters *m* (values tested: 1, 5, 10, 15, 20, 25), *c* (all possible ranges tested between: 0.1, 0.2, 0.3, 0.4, 0.5), s_1 (values tested: 0, 0.1, 0.2, 0.3, 0.4, 0.5) and s_2 (values tested: -0.2, -0.1, 0, 0.1, 0.2), and replicated each combination ten times. Following this, we investigated which of the simulative variants provided the closest match with the allele frequencies observed in the CDPOP runs. The comparisons were based on three indicators of neutral structure:

1) Principal component analysis (PCA) of the genotype matrix (Fig. 2-2a): a PCA of the genotype matrix was performed using the *prcomp* R function for each simulation (of both the CDPOP and the present customized method), where the differential of the variation explained by each principal component was then calculated. When the population is structured, the first principal component usually shows strong differences in the percentage of explained variation compared with the other components (Novembre et al., 2008). In contrast, when the population structure is absent, minor changes in this differential value emerge. The curve describing this differential value was then used for a pairwise comparison between the ten replicates of each CDPOP scenario and the ten replicates of each simulative variant (from the customized method). The curves were compared by calculating the root mean square error (RMSE), then the average RMSE was used to rank simulative variants.
2) F statistic (Fst; Fig. 2-2b): five areas, which spanned four cells each, were selected to represent subpopulations of the study area. Four areas located at the four corners of the 10-by-10 cell grid and the fifth located at the center. For each simulation, we computed the pairwise Fst (Weir & Cockerham, 1984) between these sub-populations using the *hierfstat* R package (version 0.04; Goudet, 2005). An Fst close to 0 indicates the absence of a genetic structure between sub-populations, while under a structured scenario this value tends to raise (Luikart et al., 2003). The distribution of all the Fst values for the ten CDPOP replicates were compared to the distribution of the Fst of ten replicates of each simulative variant using the Kullback-Leibler Divergence (KLD; Kullback & Leibler, 1951) analysis implemented in the *LaplacesDemon* R package (version 16.1.1; Statisticat & LCC, 2018). KLD was then used to rank simulative variants.

Figure 2-2. Comparison of genotypes simulated with CDPOP and our method. Two distinct demographic scenarios were conceived, one with random mating (panmictic population) and one with dispersal costs related to distance (structured population). For each of them, CDPOP simulated the evolution of the population over 100 generations (red graphs) and replicated the same scenario 10 times. Simultaneously, we replicated the same scenarios using our simulative approach and show here the closest match (also replicated 10 times) to CDPOP simulations (blue graphs). Five methods for evaluating the genetic makeup are presented. In a), a principal component analysis is applied to the genotype matrix and the differential of the percentage of explained variation by each component is plotted for every replicate. In b), a pairwise Fst analysis between five subpopulations is performed for every replicate and the resulting distribution of Fst is shown. In c), Mantel correlation is calculated between a matrix of genetic and of geographic distances. The resulting Mantel R for every replicate is shown. In d) and e), the allelic frequency of adaptive genotypes is shown as a function of the environmental variables causing selection (representing a case of moderate and strong selection, respectively).



3) Mantel test (Fig. 2-2c): for each simulation, we computed the genetic and geographic distance between all individuals of the population applying the R *dist* function to the genotype matrix and the coordinates, respectively. Next, we calculated the Mantel correlation (mR; Mantel, 1967) between these two distance matrices using the *mantel.rtest* function implemented in the *ade4* R package (version 1.7, Dray & Dufour, 2007). When mR is close to 0, it indicates the absence of correlation between the genetic and geographical distances, suggesting the absence of genetic structure (*i.e.* panmictic population scenario). In contrast, an mR closer to -1 or +1 indicates that genetic distances match geographic distances, as we would expect in a structured population scenario (Mantel, 1967). The average mR was calculated for each simulative variant and compared to the average mR measured in the two CDPOP scenarios. The resulting difference in mR (ΔmR) was used to rank simulative variants.

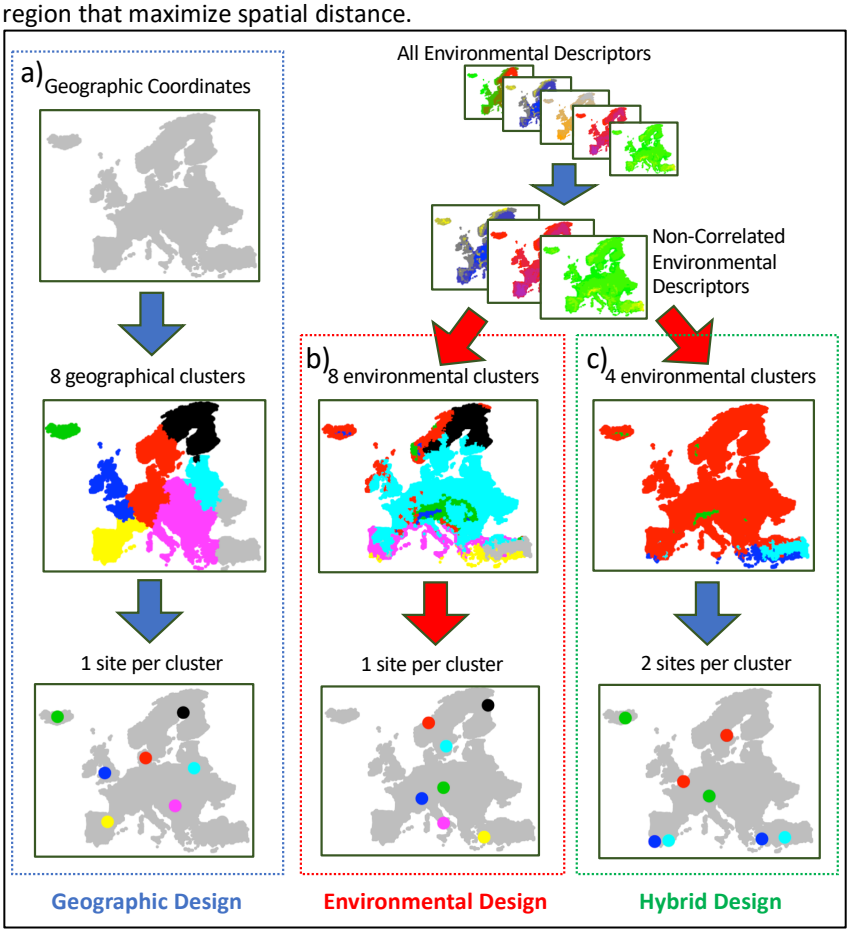
The three ranking coefficients (RMSE, KLD and ΔmR) were scaled using the *scale* R function and averaged, and the resulting value was used to rank simulative variants. In this way, it was possible to find one simulative variant with the best ranking when compared to the CDPOP panmictic population scenario, and another with the best ranking when compared to the CDPOP structured population scenario. These two simulative variants provided the values of *m* and *c* for the simulations on the complete dataset.

Subsequently, we focused on the comparison of the values for the parameters defining the adaptive processes: s_1 and s_2. For each CDPOP demographic scenario, we searched for the s_1 and s_2 combination that resulted in a simulative variant that best matched the allelic frequencies of each of the two genotypes implied in selection (moderate and strong). The environmental variable of interest was distributed in 20 equal intervals and within each interval the allelic frequencies of the adaptive genotype were computed. This resulted in the computation of a regression line for each simulation that described the allelic frequency of the adaptive genotype as a function the environmental variable causing the selective constraint (Fig. 2-2d-e). Next, we calculated the RMSE to compare this regression line between the CDPOP scenarios and the respective simulative variant (*i.e.* those with the optimal *m* and *c* according to the previous analyses) under different s_1 and s_2 combinations. For the two demographic scenarios, the ranges of s_1 and s_2 were ranked according to RMSE to represent a moderate to strong selection in the simulations for the complete dataset.

2.3.4. Sampling design

Four types of sampling design are proposed: three of them differently account for the characteristics of the landscape while one randomly selects the sampling locations. The first is “geographic” (Fig. 2-3a) and is defined through a hierarchical classification of the sites based on their geographic coordinates. The desired number of sampling locations (L) determines the number of clusters and the geographical center of each cluster is set as a sampling

Figure 2-3. The three sampling design approaches accounting for landscape characteristics. The three maps illustrate how the eight sampling sites are chosen under three different sampling designs. Under a geographic strategy (A), sample location is selected using only geographic coordinates in order to maximize distance between sites. The environmental design (B) is computed using environmental variables (after filtering out highly correlated variables), in order to maximize the climatic distance between the chosen sites. The hybrid strategy (C) is a combination of the first two designs: first the landscape is divided into distinct environmental regions before choosing sites within each region that maximize spatial distance.



location. The goal of this strategy is to sample sites located as far apart as possible from each other in the geographical space to guarantee spatial representativeness.

The second design type is "environmental" (Fig. 2-3b). It is based on the computation of distances depending on the values of environmental variables. The latter are first processed by a correlation filter: when two variables are found correlated to each other (R>±0.5), one of them (randomly chosen) is excluded from the dataset. The remaining un-correlated descriptors are scaled (sd=1) and centered (mean=0) using R *scale* function. The scaled values are used to perform a hierarchical clustering between the landscape sites. Like the previous design, the desired number of sampling locations (L) defines the number of clusters. For each cluster, the environmental center is defined by an array containing the mean of the scaled environmental values. Then, the Euclidean distances between this array and the scaled values of each site of the cluster are computed. On this basis, the most similar sites to each center are selected as sampling locations. This strategy aims to maximize environmental contrast between sampling locations and thus favors the detection of adaptive signals (Manel et al., 2012; Riginos et al., 2016).

The third design is "hybrid" (Fig. 2-3c) and is a combination of the first two. It consists of dividing the landscape into *k* environmental regions and selecting within each of these regions two or more sampling locations based on geographic position. Initially, the environmental variables are processed as for the environmental design (correlation-filter and scaling) and

used for the hierarchical classification of the landscape sites. The next step is separating the landscape sites in *k* environmental regions based on this classification. The allowed value of *k* ranges between 2 and half of the desired number of sampling locations (L). We use the R package NbClust (version 3.0, Charrad, Ghazzali, Boiteau, & Niknafs, 2015) to find the optimal value of *k* within this range. The optimal *k* is then used to determine the *k* environmental regions. Next, the number of sampling locations (L) is equally divided across the *k* environmental regions. If *k* is not an exact divisor of L, the remainder of L/*k* is randomly assigned to environmental regions. The number of sampling locations per environment region (L_{ki}) can therefore be equal among environmental regions or, at worst, differ by one (for ex. If L=8 and *k*=4: L_{k1}=2, L_{k2}=2, L_{k3}=2, L_{k4}=2; if L=10 and *k*=4: L_{k1}=3, L_{k2}=3, L_{k3}=2, L_{k4}=2). Sampling locations within environmental regions are chosen based on geographical position. Geographical clusters within each environmental region are formed as in the geographic design, setting L_{ki} as the number of clusters. The landscape site spatially closer to the center of each geographical cluster is selected as sampling location. In such a way, the procedure allows the replication of similar environmental conditions at distant sites, being therefore expected to disentangle neutral and adaptive genetic variation and to promote the detection of variants under selection (Manel et al., 2012; Rellstab et al., 2015; Riginos et al., 2016).
The fourth type of design is "random": the sampling locations (L) are randomly selected across all the available landscape sites. In our simulations, we tested each type of sampling design with numbers comparable to the ones used in real experiments (see Tab. 2-1). We used 5 levels of sampling locations L (5, 10, 20, 40 and 50 locations) and 6 of sample sizes N (50, 100, 200, 400, 800 and 1600 individuals). In iterations for which the sample size is not an exact multiple of the number of sites (for ex., 20 sites and 50 individuals), the total number of individuals was changed to the closest multiple (here 40 individuals). The scripts including these procedures were written in R using the functions embedded within the *stats* package (R Core Team, 2016).

2.3.5. Landscape genomics analysis

We computed association models for each iteration with the SamBada software (version 0.6.0; Stucki et al., 2017). First, the simulated matrix of genotypes is filtered through a customized R function with minor allele frequency <0.05 and major genotype frequency >0.95 to avoid including rare or monomorphic alleles and genotypes, respectively. Secondly, a principal component analysis (PCA) is run on the filtered genotype matrix to obtain synthetic variables accounting for population structure (hereafter referred to as population structure variables; Patterson, Price, & Reich, 2006). The analysis of the eigenvalues of the PCA is carried out in order to assess whether the population structure is negligible for downstream analysis or not (Patterson et al., 2006). At each iteration, the algorithm runs a Tracy-Widom significance test of the eigenvalues, as implemented in the AssocTests R package (version 0.4, Wang, Zhang, Li, & Zhu, 2017). Significant eigenvalues indicate the presence of non-negligible

population structure: in these situations, the corresponding principal components will be used as co-variables in the genotype-environment association study.

After filtering, SamBada is used to detect candidate loci for local adaptation. The software is able to run multivariate logistic regression models (Joost et al., 2007) that include population structure as a co-variable, while guaranteeing fast computations (Duruz et al., 2019; Rellstab et al., 2015; Stucki et al., 2017). To ensure compatibility with our pipeline and increase computational speed, we integrated the SamBada method into a customized python script (version 3.5; van Rossum, 1995) based on the Pandas (McKinney, 2010), Statsmodels (Seabold & Perktold, 2010) and Multiprocessing (Mckerns, Strand, Sullivan, Fang, & Aivazis, 2011) packages. *P*-values related to the two statistics (G-score and Wald-score) associated with each association model are computed and subsequently corrected for multiple testing using the R *q-value* package (version 2.6; Storey, 2003). Models are deemed significant when showing a q<0.05 for both tests. When multiple models are found to be significant for the same marker, only the best one is kept (according to the G-score). The pipeline was developed in the R-environment using the *stats* library.

2.3.6. Simulations and evaluation of the performance

Each combination of demographic scenarios, sampling designs, number of sampling locations and sample sizes was replicated 20 times for a total of 4,800 iteration (Tab. 2-2). A new genetic matrix was randomly redrawn for each iteration to change the selective forces implying local adaptation and the demographic set-up determining the neutral loci. At the end of each iteration, three diagnostic parameters were computed:

- True Positive Rate of the analysis (TPR or statistical power): percentage of true associations detected to be significant;
- False Discovery Rate of the analysis (FDR): percentage of false association among the significant ones;
- Positive Predictive Value (PPV; Marshall, 1989) of the ten strongest associations: significant associations were sorted according to the association strength (β, the value of the parameter associated to environmental variable in the logistic model calculated by SamBada). PPV represents the percentage of true associations among the best ten associations according to β.

After the simulations, we calculated median (Mdn) and inter-quartile range (IQR) of TPR, FDR and PPV under the different levels of the three elements underlying the sampling strategy (*i.e.* sampling design, number of sampling locations and sample size; Tab. 2-2). Furthermore, we estimated how changes in these three elements drove alterations in TPR, FDR and PPV (*i.e.* effect size). We focused only on main effects (*i.e.* effects of single elements of sampling strategy) since interactions effects (*i.e.* effects obtained combining two elements of the sampling strategy) appeared as minor after a preliminary visual inspection (Fig. S2-2). Since TPR, FDR and PPV did not follow a normal distribution, we applied a bootstrap resampling

Table 2-2. Table of factors varying in the simulative approach. Two different demographic scenarios are possible, one in which there is no neutral genetic structure (panmictic population) and one in which there is a structured variation (structured population). We then used sampling strategies emulating those observed in real experiments. Three different sampling design approaches accounting for landscape characteristics are proposed: one maximizing the spatial representativeness of samples (geographic), one maximizing the environmental representativeness (environmental) and one that is a combination of both (hybrid). A fourth sampling design picks sampling locations randomly. The numerical ranges we employed were comparable to those from real experiment: 5 levels for number of sampling locations spanning from 5 to 50 sites, and 6 levels of sample sizes (*i.e.* total number of samples) from 50 to 1600 samples. For each combination of the aforementioned factors, 20 replicates were computed differing in the number and types of selective forces driving adaptation. In total, 4,800 simulation were computed.

Factor	**# levels**	**Levels**
Demographic Scenarios	2	Panmictic Population, Structured Population
Sampling Design (*D*)	4	Geographic, Environmental, Hybrid, Random
Sampling Locations (*L*)	5	5, 10, 20, 40, 50
Sampling Size (*N*)	6	50, 100, 200, 400, 800, 1600
Replicates	20	
Total	4800	

technique (r=5000) to estimate their means and the related uncertainties under the different levels of each element of the sampling strategy (Dixon, 2001; Nakagawa & Cuthill, 2007). This step was performed in R, using the *boot* library (version 1.3; Canty & Ripley, 2017; Davison & Hinkley, 1997). The effect size was then calculated as the difference in the mean values of TPR, FDR or PPV (and the related 95% confidence interval; Nakagawa & Cuthill, 2007) under different levels of the elements defining the sampling strategy. In the case of numerical elements (*i.e.* number of sampling locations and sample size), effect sizes were calculated as the changes in TPR, FDR and PPV along with the increments of the ordinal factor levels (for ex*.:* change in TPR between sample sizes of 100 to 200, 200 to 400, 400 to 800, etc.). In the case of sample design, where the factor levels are not ordinal, we compared each design approach against a random sampling scheme.

2.4. Results

2.4.1. Parameters of simulations

For the panmictic population scenario, the simulative variant best matching the CDPOP results was obtained with the coefficients $m = 1$ and $c = 0.5$, whereas for the structured population scenario, the simulative variant was best at $m = 10$ and $c = Unif(0.2, 0.4)$ (Fig. 2-2a-c, Box S2-2, Tab. S2-2a-b). In the panmictic population scenario, we found that the moderate selection case was best emulated by $s_1 = 0.4$ and $s_2 = -0.2$ and the strong selection by $s_1 = 0.3$ and $s_2 = +0.1$. In the structured population scenario, the moderate

selection found its best match in the simulative variant with $s_1 = 0$ and $s_2 = -0.1$ while the strong selection in the one set with $s_1 = 0$ and $s_2 = +0.2$ (Fig. 2-2d-e, Box S2-2, Tab. S2-2c-d).

2.4.2. True Positive Rate

In general, the panmictic population scenario simulations showed higher TPR (Mdn_{PAN}=40% [IQR=0-90%]) than simulations performed under the structured population scenario (Mdn_{STR}=0% [IQR=0-40%]; Fig. 2-4a-c). For both scenarios, the largest effect sizes on TPR were generally related to changes in sample size (Tab. 2-3a). Smaller sample sizes (N= 50, 100) resulted in TPR close or equal to zero for both population scenarios (Fig. 2-4c). Under the structured population scenario, an increase of TPR started from N=200 (Tab. 2-3a), leading to an initial increase of ~4% TPR for every 10 additional samples. At N=400, this increment progressively became less abrupt until reaching a maximal value at N=800 (Mdn=100% [IQR=60-100%]; Fig. 2-4c; Tab. 2-3a). By comparison, the panmictic population scenario showed an increase in TPR starting at N=400, with a more constant and less abrupt rate of

Table 2-3. Effect sizes of element defining the sampling strategy on TPR, FDR and PPV. The table shows the changes in the averages of the three diagnostic parameters of the analysis (TPR: true positive rate, a; FDR: false discovery rate, b; PPV: positive predictive value of among the ten strongest significant association models, c) for every element determining the sampling strategy (sampling design, number of locations and sample size) under the two demographic scenarios, panmictic and structured population. In the case of sampling design, the changes refer to a comparison against the random sampling design (and positive values represent increase in the diagnostic parameter in comparison to the random case, vice versa for negative values). In the situations concerning number of sampling sites and sample size, two levels are compared and positive values indicate the increase of the diagnostic parameter under the second term of comparison (vice versa for negative values). In parentheses, the 95% confidence intervals are shown.

		a) TPR		b) FDR		c) PPV	
		Panmictic	Structured	Panmictic	Structured	Panmictic	Structured
Design	**Geo**	0.0052 (0.0049-0.0056)	0.0101 (0.0094-0.0107)	-0.0167 (-0.0174- -0.016)	0.0085 (0.0079-0.0091)	0.0158 (0.0151-0.0164)	0.0018 (0.0011-0.0025)
	Env	0.1037 (0.1032-0.1042)	0.1173 (0.1167-0.118)	-0.045 (-0.0457- -0.0444)	-0.0453 (-0.0459- -0.0447)	0.0569 (0.0562-0.0576)	0.1139 (0.1132-0.1146)
	Hyb	0.1145 (0.114-0.115)	0.1351 (0.1344-0.1357)	-0.0327 (-0.0333- -0.032)	-0.0555 (-0.0561- -0.0549)	0.048 (0.0473-0.0487)	0.1271 (0.1264-0.1278)
Nb. of Sites	**5-10**	0.0746 (0.0742-0.0751)	0.321 (0.3205-0.3215)	-0.0751 (-0.0757- -0.0744)	-0.2288 (-0.2293- -0.2283)	0.0849 (0.0842-0.0856)	0.3526 (0.3521-0.3532)
	10-20	0.0281 (0.0275-0.0286)	0.1241 (0.1234-0.1248)	-0.0459 (-0.0466- -0.0451)	-0.1008 (-0.1014- -0.1001)	0.0453 (0.0445-0.046)	0.1326 (0.1319-0.1333)
	20-40	-0.0043 (-0.0049- -0.0038)	0.0286 (0.0278-0.0293)	-0.0021 (-0.0029- -0.0013)	0.0063 (0.0057-0.007)	~0	-0.0132 (-0.0139- -0.0124)
	40-50	-0.0044 (-0.005- -0.0038)	0.0353 (0.0345-0.0361)	0.0061 (0.0053-0.0068)	-0.0636 (-0.0642- -0.0629)	-0.0097 (-0.0105- -0.0089)	0.0759 (0.0751-0.0767)
Sample Size	**50-100**	~0	0.08 (0.0799-0.0802)	-0.0025 (-0.0026- -0.0024)	-0.35 (-0.3506- -0.3493)	0.0025 (0.0024-0.0026)	0.3502 (0.3495-0.3508)
	100-200	0.0165 (0.0164-0.0165)	0.3911 (0.3906-0.3916)	-0.134 (-0.1345- -0.1336)	-0.3488 (-0.3496- -0.348)	0.1337 (0.1332-0.1342)	0.3582 (0.3574-0.359)
	200-400	0.1089 (0.1087-0.109)	0.2126 (0.212-0.2132)	-0.3896 (-0.3904- -0.3889)	0.0933 (0.0927-0.094)	0.39 (0.3893-0.3908)	0.0138 (0.0131-0.0144)
	400-800	0.2801 (0.2798-0.2805)	0.0937 (0.0931-0.0944)	-0.2217 (-0.2224- -0.2211)	0.1813 (0.1808-0.1818)	0.2264 (0.2258-0.2271)	0.025 (0.0244-0.0256)
	800-1600	0.3031 (0.3026-0.3035)	0.0137 (0.013-0.0143)	-0.0411 (-0.0415- -0.0406)	0.1107 (0.1103-0.1111)	0.0836 (0.0832-0.0841)	0.018 (0.0174-0.0186)

increase (Fig. 2-4c, Tab. 2-3a). Under this scenario, a N=1600 was not sufficient to yield maximal TPR (Mdn=80% [IQR=60-90%]; Fig. 2-4c).

The effect sizes on TPR related to the number of sampling locations were less pronounced, compared to those of sample size (Tab. 2-3a; Fig. 2-4b). Under both population scenarios, the largest increases in TPR were observed when passing from L=5 to L=10 (+7% and + 32% TPR under panmictic and structured scenarios, respectively; Fig. 2-4b, Tab. 2-3a). At higher numbers of sampling sites (L=20, 40 and 50) the incremental rate of TPR was less evident but still positive under the structured scenario and close to zero under the panmictic one (Tab. 2-3a, Fig. 2-4b).

Similar to the influence of sampling locations, the type of sampling design had a minor effect on TPR when compared to the effect that sample size had (Tab. 2-3a; Fig. 2-4a). When compared to the random approach, a hybrid design approach was seen to increase the TPR by +11% and +14% under panmictic and structured population scenarios, respectively (Fig. 2-4a, Tab. 2-3a). The environmental design displayed slightly lower effect sizes on TPR (+10% and +12% under panmictic and structured population scenarios, respectively; Fig. 2-4a, Tab. 2-3a), while those of the geographic design were close to zero (Fig. 2-4a, Tab. 2-3a).

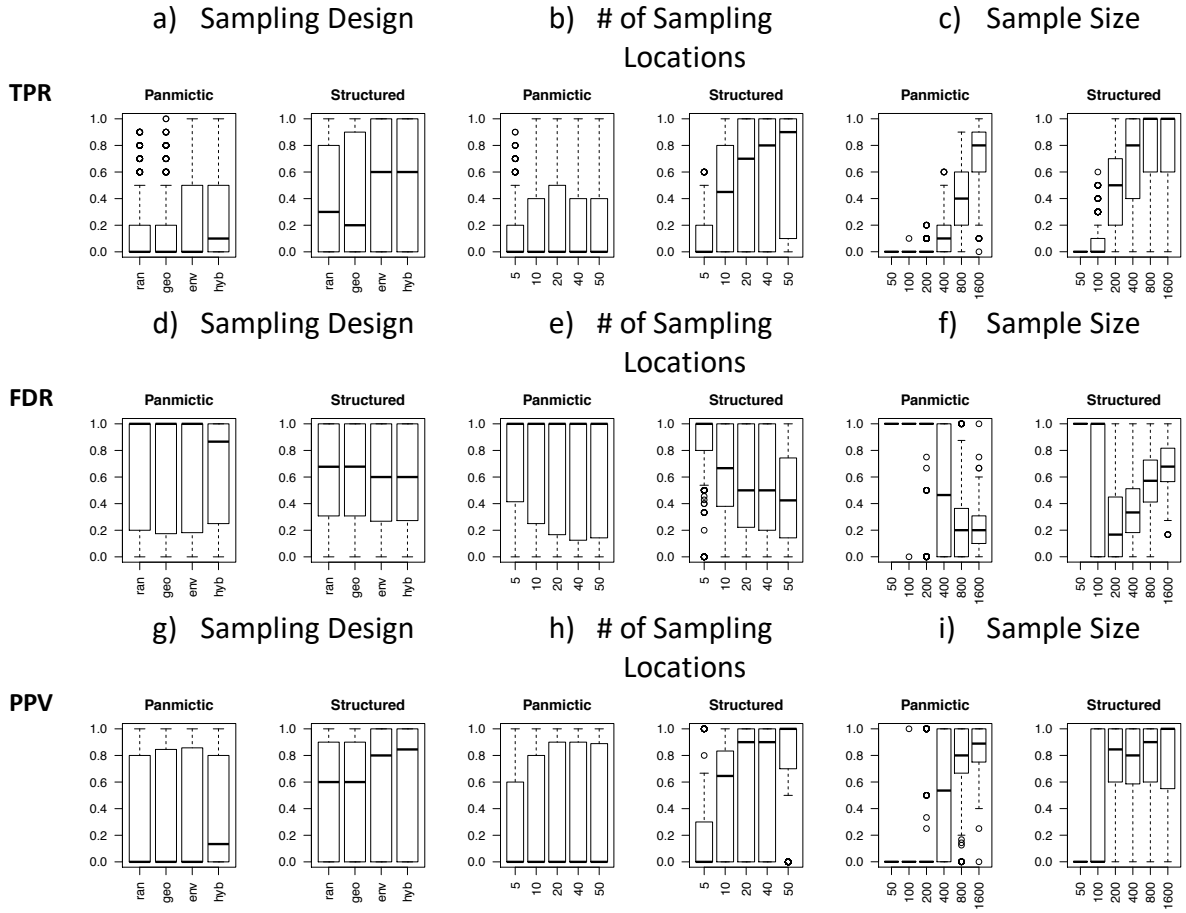
2.4.3. False Discovery Rate

False discoveries generally appeared at a higher rate under a panmictic population scenario (Mdn_{PAN}=100% [IQR=20-100%]) than under a structured population scenario (Mdn_{STR}=63% [IQR=20-100%]; Fig. 2-4d-f). Sample size had the largest effects on FDR for both population scenarios (Tab. 2-3b; Fig. 2-4f). For the panmictic population scenario, median FDR was 100% at smaller sample sizes (N=50, 100 and 200; Fig. 2-4f), but between N=200 and N=400, the FDR began to decrease by ~2% for every ten additional samples taken (Tab. 2-3b). The reduction in FDR was less abrupt after N=400, and quasi-null after N=800 (Tab. 2-3b). At N=1600, median FDR was 20% [IQR=10-30%] (Fig. 2-4f). The structured population scenario produced a different pattern: the largest median FDR was found at smaller sample sizes (N=50 and 100), before a steep decrease was observed closer to N=200 (Fig. 2-4f, Tab. 2-3b). At larger sample sizes (N=400, 800, 1600), FDR showed a logarithmic increase in growth rate where, at its most abrupt (between N=100 and 400), there was an increase of +0.9% FDR for every ten additional samples (Fig. 2-4f, Tab. 2-3b). For the structured population scenario, N=1600 resulted in a median FDR of 68% [IQR=57-82%].

Under both population scenarios, the effect of sampling location number on FDR was weaker than the effect of sample size (Tab. 2-3b, Fig. 2-4e). Similar to the pattern for TPR, the effects were stronger when passing from L=5 to L=10 (leading to a decrease of FDR of 7% and 23% under panmictic and structured population scenarios, respectively; Fig. 2-4e, Tab. 2-3b), than between higher numbers of sampling locations (L=20, 40, 50; Tab. 2-3b).

Sampling design showed effects on FDR, but it was not as strong as the influences of sample size and sampling locations (Tab. 2-3b, Fig. 2-4d). When compared to a random sampling

Figure 2-4. Effects of sampling strategy on the landscape genomics simulations. The plots display how the performance of landscape genomics experiments is driven by changes in the elements defining the sampling strategy. Three diagnostic parameters are used to measure the performance of each strategy: true positive rate (TPR; a-c) and false discovery rate (FDR; d-f) for the analysis and the positive predictive value of the ten strongest significant association models (PPV; g-i). For each diagnostic parameter, we show the effect of sampling design (a, d, g; ran=random, geo=geographic, env=environmental, hyb= hybrid), number of sampling locations (b, e, h; 5, 10, 20, 40 or 50 sites) and sample size (c, f, I; 50, 100, 200, 400, 800, 1600 individuals) under two demographic scenarios: panmictic and structured population.



scheme, both the environmental and hybrid sampling designs showed comparable decreases in FDR (hybrid design: -3% and -6%; environmental design: -4% and -5% under panmictic and structured population scenarios, respectively), while the geographic one showed negligible changes (Fig. 2-4d, Tab. 2-3b).

2.4.4. Positive Predictive Value

The PPVs of the ten strongest significant associations (hereinafter simply referred to as PPV) was generally higher under the structured population scenario (Mdn_{PAN}=70% [IQR=0-100%]) than under the panmictic population scenario (Mdn_{PAN}=0% [IQR=0-80%]; Fig. 2-4g-i). As with TPR and FDR, changes in sample size had the strongest influence on PPV under both

population scenarios (Tab. 2-3c, Fig. 2-4i). Under the panmictic population scenario, median PPV was ~0% for the smaller sample sizes (N=50, 100 and 200; Fig. 2-4i), after which patterns of increase were observed: from N=200 to 400 there was an increase of PPV of ~+2% for every 10 additional samples, and from N=800 to 1600 PPV continued to increase though it was less abrupt, resulting in a median PPV of 88% [IQR=75-100%] at N=1600 (Fig. 2-4i, Tab. 2-3c). Under the structured population scenario, fewer samples were required to observe a similar increment: while PPV was close to 0 for N=from N=50 to N=100 the PPV increased by +7% for every ten additional samples (Fig. 2-4i, Tab. 2-3c). The increment of PPV became gradually weaker when transitioning between higher levels (N=400, 800 and 1600) and led to a median PPV of 100% [IQR=57.5-100%] at N=1600.

Similar to TPR and FDR, the effect of sampling location number on PPV was weaker than the one of sample size (Tab. 2-3c, Fig. 2-4h). This effect was particularly evident under the structured population scenario, where an increase of the number of sampling locations strongly raised PPV (Fig. 2-4h). The strongest PPV increment was observed between L=5 and 10, where each additional sampling location raised the PPV by +7% (Fig. 2-4h, Tab. 2-3c). With more sampling locations (L=20, 40 and 50) the rate of increase of PPV remained but was weaker (Fig. 2-4h, Tab. 2-3c). In the panmictic population scenario, an increase in the number of sampling locations produced weaker changes in PPV (Fig. 2-4h, Tab. 2-3c).

The sampling design used resulted in rate changes for PPV, despite being less strong than when compared to the other elements of the sampling strategy (Tab. 2-3c, Fig. 2-4g). When compared to a random sampling scheme, the hybrid design and the environmental design increased PPV of +6% and +5% under the panmictic population scenario and of +13% and +12% under the structured one, respectively (Fig. 2-4g, Tab. 2-3c). In contrast, the geographic design did not result in pronounced changes of PPV (Fig. 2-4g, Tab. 2-3c).

2.5. Discussion

The simulations presented in this study highlight that sampling strategy clearly drives the outcome of a landscape genomics experiment, and that the demographic characteristics of the studied species can significantly affect the analysis. Despite some limitations that will be discussed below, the results obtained make it possible to answer three questions that researchers are confronted with when planning this type of research investigation.

2.5.1. How many samples are required to detect any adaptive signal?

In line with the findings of previous studies (*e.g.* Lotterhos & Whitlock, 2015), our results suggest that sample size is the key factor in securing the best possible outcome for a landscape genomics analysis. Where statistical power is concerned, there is an

unquestionable advantage in increasing the number of samples under the scenarios tested. When focusing on the panmictic population scenario, we found a lack of statistical power in simulations for N≤200, while detection of true positives increased significantly for N≥400 (Fig. 2-4c). As we progressively doubled sample size (N=800, 1600), TPR linearly doubled as well (Fig. 2-4c). Under the structured population scenario, this increase in statistical power started at N≥100 and followed a logarithmic trend that achieved the maximum power at N≥800 (Fig. 2-4c).

These results show that it is crucial to consider the population's demographic background to ensure sufficient statistical power in the analyses, as advised by several reviews in the field (Balkenhol et al., 2017; Manel et al., 2012; Rellstab et al., 2015). In fact, the allelic frequencies of adaptive genotypes respond differently to a same environmental constraint under distinct dispersal modes (Fig. 2-2d-e). When individual dispersal is limited by distance (structured population scenario), the allelic frequencies of adaptive genotypes are the result of several generations of selection, resulting in a progressive disappearance of non-adaptive alleles from areas where selection acts. When the dispersal of individuals is completely random (panmictic population scenario), the same selective force only operates within the last generation, such that even non-adaptive alleles can be found where the environmental constraint acts. Under these premises, a correlative approach for studying adaptation (such as SamBada) is more likely to find true positives under a structured population scenario rather than under a panmictic one.

The dichotomy between structured and panmictic populations also emerges when analyzing false discovery rates. Under the panmictic population scenario, increasing the number of individuals sampled reduced FDR, while the inverse pattern was seen under a structured population scenario (Fig. 2-4f). The issue of high false positives rates under structured demographic scenarios is well acknowledged in landscape genomics (De Mita et al., 2013; Rellstab et al., 2015). Population structure results in gradients of allele frequencies that can mimic and be confounded with patterns resulting from selection (Rellstab et al., 2015). As sample size increases, the augmented detection of true positives is accompanied by the (mis-)detection of false positives. Under the panmictic population scenarios, these confounding gradients of population structure are absent (Fig. 2-2a-c) and high sample sizes accentuate the detection of true positives only (Fig. 2-4f).

Working with FDR up to 70% (Fig. 2-4f) might appear excessive, but this should be contextualized in the case of landscape genomics experiments. The latter constitute the first step toward the identification of adaptive loci, which is generally followed by further experimental validations (Pardo-Diaz et al., 2015). Most landscape genomics methods test single-locus effects (Rellstab et al., 2015). This framework is efficient for detecting the few individual loci that provide a strong selective advantage, rather than the many loci with a weak individual-effect (for instance those composing a polygenic adaptive trait; Pardo-Diaz et al., 2015). For this reason, when researchers are faced with a high number of significant associations, they tend to focus on the strongest ones (Rellstab et al., 2015), as we did here

by measuring the PPV of the ten strongest associations. By relying on this diagnostic parameter, we could show that increasing sample size ensures that the genotypes more strongly associated with environmental gradients are truly due to adaptive associations (Fig. 2-4i). Under these considerations, acceptable results are obtainable with moderate sample sizes: a median PPV of at least 50% was found with simulations with N=400 and N=200 under panmictic and structured population scenario, respectively.

Each landscape genomic experiment is unique in terms of environmental and demographic scenarios, which is why it is not possible to propose a comprehensive mathematical formula to predict the expected TPR, FDR and PPV based solely on sample size. When working with a species with a presumed structured population (for instance, wild land animals), we advise against conducting experiments with fewer than 200 sampled individuals, as the statistical requirements to detect true signals are unlikely to be met. Panmixia is extremely rare in nature (Beveridge & Simmons, 2006), but long-range dispersal can be observed in many species such as plants (Nathan, 2006) and marine organisms (Riginos et al., 2016). When studying species of this kind, it is recommendable to increase sample size to at least 400 units.

2.5.2. How many sampling sites?

Increasing the number of samples inevitably raises the cost of an experiment, largely resulting from sequencing and genotyping costs (Manel et al., 2010; Rellstab et al., 2015). Additionally, field work rapidly increases the cost of a study in cases where sampling has to be carried out across landscapes with logistic difficulties and physical obstacles. Therefore, it is both convenient and economical to optimize the number of sampling locations to control for ancillary costs.

De Mita et al. (2013) suggested that increasing the number of sampling locations would raise power and reduce false discoveries. The present study partially supports this view. A small number of sampling locations (L=5) was found to reduce TPR and PPV while increasing FDR, compared to using more sampling locations (L=10, 20, 40 and 50; Fig. 2-4b, e, h). This is not surprising, because when sampling at a small number of locations the environmental characterization is likely to neglect some contrasts and ignore confounding effects between collinear variables (Leempoel et al., 2017; Manel et al., 2010). This was particularly evident under the structured population scenario (Fig. 2-4b, e, h). In contrast, we found that higher numbers of sampling locations (L=40 and 50) provided little benefits in terms of TPR, FDR and PPV, compared to a moderate number of locations (L= 20; Fig. 2-4b, e, h). These discrepancies with previous studies are probably due to differences in the respective simulative approaches applied (we use several environmental descriptors instead of one) and the characteristics of the statistical method we employed to detect signatures of selection. In fact, as a number of sampling locations is sufficient to portray the environmental contrasts of the study area, adding more locations does not bring additional information and therefore does not increase statistical power. The implications of the information described above are considerable since

the cost of field work can be drastically reduced with marginal countereffects on statistical power and false discoveries.

2.5.3. Where to sample?

Compared with random or opportunistic approaches, sampling designs based on the characteristics of the study area are expected to improve the power of landscape genomics analysis (Lotterhos & Whitlock, 2015). We developed three distinct methods to choose sampling locations accounting for geographical and/or environmental information (geographic, environmental and hybrid designs). We compared these design approaches between themselves and with random sampling schemes. The approach based on geographic position (geographic design) resulted in statistical power similar to the random designs (Fig. 2-4a, d, f), while those based on climatic data (environmental and hybrid design) displayed remarkably higher TPRs and PPV and slightly lower FDR (Fig. 2-4a, d, f). These beneficial effects on the analysis were accentuated under the structured demographic scenario.

These results match previous observations: methods conceived to take advantage of environmental contrasts facilitate the detection of adaptive signals (Manel et al., 2012; Riginos et al., 2016). Furthermore, the hybrid design prevents the sampling of neighboring sites with similar conditions, therefore avoiding the superposition between adaptive and neutral genetic variation (Manel et al., 2012). This is likely to explain why the hybrid design slightly outperformed the environmental approach (Fig. 2-4a, d, f).

For these reasons, we strongly advise in using a sampling scheme accounting for both environmental and geographical representativeness. Without bringing any additional cost to the analysis, this approach can boost statistical power of up to 14% under a complex demographic scenario (Tab. 2-3a), in comparison to a regular (geographic) or random sampling scheme.

2.5.4. Limitation

The preliminary run of comparison with a well-established forward-in-time simulation software (CDPOP) displayed the pertinence of our customized simulative approach (Fig. 2-2). The neutral genetic variation appeared as random under the panmictic population scenario (no skew on the PC graph, Fst close to 0, mR close to 0) and structured under the structured population scenario (skew in the PCA graph, Fst higher than 0, mR different from 0; Fig. 2-2a-c). Adaptive allele frequencies also matched theoretical expectations: the responses along the environmental gradients were more stressed under the structured population scenario than under the panmictic one (Fig. 2-2d-e).

Nonetheless, the use of *forward-in-time* simulations on the complete dataset (used by De Mita et al., 2013; Lotterhos & Whitlock, 2015) would probably have resulted in more realistic

scenarios. In order to be used in a framework as the one proposed here, the *forward-in-time* methods should be compatible with a large number of spatial locations (*i.e.* potential sampling sites), hundreds of individuals per location and a genetic dataset counting at least one thousand loci, of which 10 set as adaptive against distinct environmental variables. Importantly, all these requirements should be fulfilled at a reasonable computational speed (with our method, for instance, genotypes are computed in a few seconds). As far as we know, there are no existing software meeting these criteria.

The framework we presented here is based on an artificial genomic architecture encompassing 10 adaptive loci and 990 neutral ones. Given the generally high rates of false positives in landscape genomics (Balkenhol et al., 2017; Rellstab et al., 2015), it is hard to estimate a realistic percentage of SNPs implied in local adaptation from the literature. Besides, this percentage is driven by various factors specific to the biology of the studied species/population (for ex. life cycle duration, genome size, the mutation rate, population size, extent of selective pressures; Dittmar, Oakley, Conner, Gould, & Schemske, 2016) and to the methods applied (for ex. genotyping strategy; Rellstab et al., 2015). Furthermore, not all adaptive genotypes are the same (Dittmar et al., 2016) and, as a consequence, diversified landscape genomics methods exist. Our framework relied on SamBada, a well-established method that assumes that 1) genotype-environment association follows a logistic response and 2) a few genotypes have large effects (Stucki et al., 2017). Not all the landscape genomics methods are based on these assumptions though, and the guidelines described in this work might not be relevant for all these methods.

2.5.5. Conclusions

The present work provides guidelines for optimizing the sampling strategy in the context of landscape genomic experiments. Our simulations highlight the importance of considering the demographic characteristic of the studied species when deciding the sampling strategy to be used. For species with limited dispersal, we suggest working with a minimum sample size of 200 individuals to achieve sufficient power for landscape genomic analyses. When species display long-range dispersal, this number should be raised to at least 400 individuals. The costs induced by a large number of samples can be balanced by reducing those related to field work. In cases where a moderate number of sampling locations (20 sites) is sufficient to portray the environmental contrasts of the study area, there is only minimal statistical benefit for sampling a larger number of sites (40 or 50). Furthermore, we describe an approach for selecting sampling locations while accounting for environmental characteristics and spatial representativeness, and show its benefic effects on the detection of true positives.

2.6. Acknowledgements

We thank the anonymous reviewers for the useful comments and suggestions provided during the redaction of this paper. We acknowledge funding from the IMAGE (Innovative Management of Animal Genetic Resources) project funded under the European Union's Horizon 2020 research and innovation program under grant agreement No. 677353.

3. Article B

Seascape genomics as a new tool to empower coral reef conservation strategies: An example on north-western Pacific Acropora digitifera

Oliver Selmoni[1], Estelle Rochat[1], Gaël Lecellier[2,3], Véronique Berteaux-Lecellier[2] and Stéphane Joost[1].

[1]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

[2]UMR250/9220 ENTROPIE IRD-CNRS-UR, Labex CORAIL, Noumea, New Caledonia.

[3]UVSQ, Université de Paris-Saclay, Versailles, France.

Postprint version of the article published in *Evolutionary Applications:*
Selmoni, O., Rochat, E., Lecellier, G., Berteaux-Lecellier, V., and Joost, S. (2020). Seascape genomics as a new tool to empower coral reef conservation strategies: an example on north-western Pacific Acropora digitifera. Evol. Appl., 588228. doi:10.1101/588228.

Contribution of the candidate:
As first author of this article, I initiated the research, retrieved and processed genetic data from the original publication, ran the seascape genomics analyses, developed the conservation indices and wrote the first version of the manuscript. The co-authors provided advice concerning the methods used, interpretation of the results and critically revised the manuscript before submission.

Environmental genomic studies characterizing the adaptive potential of corals are rare. Moreover, these studies hardly fit the category of seascape genomics, as some used low number of genotyping markers (e.g. Lundgren et al., 2013), had restricted numbers of sampling sites (e.g. Bay & Palumbi, 2014), or small sample sizes (e.g. Thomas et al., 2017). This lack of precedent studies raised some doubts on the suitability of seascape genomics to characterize local adaptation in corals. For example, would the spatial resolution of remote sensing data be sufficient to capture selective forces in the fine-scale seascape of a reef? (Riginos et al., 2016). And would the coarse genomic resolution of cost-effective sequencing strategies (e.g. RAD-seq) be sufficient to genotype rare adaptive variants? (Lowry et al., 2017). And even assuming that seascape genomics would work with corals, how could the results be useful and comprehensible in a conservation perspective? (Beger et al., 2014).

In this work, we used the genomic dataset from a previous publication studying the population structure of coral from the Ryukyu Archipelago (Japan; Shinzato et al., 2015). The dataset was suited for seascape genomics as sample size (155 colonies) and number of sampling sites (11) were sufficiently large, based on the guidelines developed in article A. Genotyping was performed using a whole-genome-sequencing technique. However, the limited number of SNPs with sufficient confidence-call resulted in a dataset with genomic resolution comparable to a RAD-seq genotyping. We ran a seascape genomics analysis combining this genomic dataset with environmental gradients derived from remote sensing data.

The main challenge in this work was to transpose the findings of seascape genomics analysis into objective indices for the reef conservation of the Ryukyu Archipelago. The framework we designed was inspired by two tools. The first is called "spatial area of genotype probability" (SPAG; Rochat & Joost, 2019), and uses models of genotype-environment association to predict the probability of presence of putative adaptive genotypes (in our case, against heat stress). The second tool was the prediction of reef connectivity, as dispersal is crucial for the spread of heat-stress adapted genotypes (Matz, Treml, Aglyamova, & Bay, 2018; Matz, Treml, & Haller, 2019). The combination of these two tools defined an adaptive potential index against heat stress that is objective and quantifiable. An index with these characteristics can easily fit in well-established prioritization criteria for reef conservation (Ball, Possingham, & Watts, 2009; Beger et al., 2014).

This article constitutes a proof-of-concept showing that seascape genomics can be applied to corals and can provide useful insights for reef conservation. The methods described in this article were the foundations for the work carried out in the SABLE project.

3.1. Abstract

Coral reefs are suffering a major decline due to the environmental constraints imposed by climate change. Over the last 20 years, three major coral bleaching events occurred in concomitance with anomalous heat waves, provoking a severe loss of coral cover worldwide. The conservation strategies for preserving reefs, as they are implemented now, cannot cope with global climatic shifts. Consequently, researchers are advocating for preservation networks to be set-up to reinforce coral adaptive potential. However, the main obstacle to this implementation is that studies on coral adaption are usually hard to generalize at the scale of a reef system.

Here, we study the relationships between genotype frequencies and environmental characteristics of the sea (seascape genomics), in combination with connectivity analysis, to investigate the adaptive potential of a flagship coral species of the Ryukyu Archipelago (Japan). By associating genotype frequencies with descriptors of historical environmental conditions, we discovered six genomic regions hosting polymorphisms that might promote resistance against heat stress. Remarkably, annotations of genes in these regions were

consistent with molecular roles associated with heat responses. Furthermore, we combined information on genetic and spatial distances between reefs to predict connectivity at a regional scale.

The combination of these results portrayed the adaptive potential of this population: we were able to identify reefs carrying potential heat stress adapted genotypes and to understand how they disperse to neighbouring reefs. This information was summarized by objective, quantifiable, and mappable indices covering the whole region, which can be extremely useful for future prioritization of reefs in conservation planning. This framework is transferable to any coral species on any reef system, and therefore represents a valuable tool for empowering preservation efforts dedicated to the protection of coral reefs in warming oceans.

3.2. Introduction

Coral reefs are suffering a severe decline due to the effects of climate change (Hughes et al., 2017). Loss of reef is already showing catastrophic consequences for marine wildlife that depend on these structures (Pratchett et al., 2018), with disastrous aftermaths expected for human economies as well (Moberg & Folke, 1999). One of the major threats to the persistence of these ecosystems is coral bleaching (Bellwood et al., 2004): a physiological response induced by environmental stress that provokes hard skeleton corals, the cornerstone of reefs, to separate from the symbiotic microbial algae essential for their survival (Mydlarz et al., 2010).

Over the last 20 years, episodes of coral bleaching struck world-wide and resulted in a local coral cover loss of up to 50% (Hughes, Anderson, et al., 2018; Hughes et al., 2017). Heat stress is considered the main driver of coral bleaching (Hughes et al., 2017), but additional stressors causing coral decline were also identified (*e.g.* ocean acidification, water eutrophication, sedimentation and overfishing; Anthony et al., 2008; Ateweberhan et al., 2013; Maina et al., 2008).

Conservation efforts to mitigate the threat of coral bleaching tend to focus on restoring reefs that have undergone severe losses, as well as limit the impact of future bleaching events (Baums, 2008; Bellwood et al., 2004; C. N. Young et al., 2012). To achieve these aims, two main strategies are currently used: establish marine protected areas (MPAs) at reefs, and develop coral nurseries (Baums, 2008; Bellwood et al., 2004; C. N. Young et al., 2012). MPAs are designated zones in which human access and activities are restricted in order to alleviate the effects of local anthropogenic stressors (Lester et al., 2009). Coral nurseries are underwater gardens of colonies that can then be transplanted to restore damaged reefs (Baums, 2008; C. N. Young et al., 2012). For both conservation strategies, researchers advocate the use of methods that account for demographic connectivity such that the location of a conservation measure can also promote resistance and resilience for

neighbouring sites (Baums, 2008; Krueck et al., 2017; Lukoschek et al., 2016; Palumbi, 2003; Shanks, Grantham, & Carr, 2003). Despite the observed beneficial effects of these conservation policies worldwide (Cinner et al., 2016; Rodgers et al., 2017; Selig & Bruno, 2010), these solutions do not confer resistance against the heat stress associated with the last mass bleaching events (Baums, 2008; Hughes et al., 2017). Coral reefs that had experienced previous heat stress were found to be more resistant to subsequent heat waves (Hughes et al., 2019; Krueger et al., 2017; Penin et al., 2013; Thompson & van Woesik, 2009), but to date this information is neglected in conservation actions (Baums, 2008; Maina et al., 2011; OECD, 2017). There is an urgent need to understand whether these observations are due to evolutionary processes and, if so, to determine how the underlying adaptive potential could be included in predictions of climate change responses and in conservation programs (Baums, 2008; Logan et al., 2014; Maina et al., 2011; van Oppen et al., 2015).

To this end, seascape genomics tools are likely to play an important role. Seascape genomics is the marine counterpart of landscape genomics, a branch of population genomics that investigates adaptive potential through field-based experiments (Balkenhol et al., 2017). Samples that are collected across a landscape are genotyped using next-generation-sequencing techniques, describing thousands of genetic variants, while simultaneously the environmental variables of the study area are characterized, usually using remote-sensing data to describe specific local climatic conditions (Leempoel et al., 2017). Genomics and environmental information are then combined to detect genetic polymorphisms associated with particular conditions (*i.e.,* potentially adaptive genotypes against a specific condition; Rellstab et al., 2015). This approach has been applied to many terrestrial species, and is increasingly being used to analyse marine systems in what is referred to as *seascape genomics* (exhaustively reviewed in Riginos, Crandall, Liggins, Bongaerts, & Treml, 2016). To our knowledge no seascape genomics experiment has yet been applied to reef corals. In fact, adaptation of these species has been mostly studied via transplantation assays coupled with aquarium conditioning, which is a time- and resource-demanding approach that is often restricted to a couple of reefs experiencing contrasting conditions (Howells et al., 2013; Krueger et al., 2017; Palumbi et al., 2014; Sampayo et al., 2016; Ziegler et al., 2017). Genotype-environment associations studies have also been conducted on corals, but have used either a limited number of markers (<10 Single Nucleotide Polymorphism, SNPs) in Lundgren, Vera, Peplow, Manel, & van Oppen, 2013), a restricted number of locations (two in Bay & Palumbi, 2014), or focused on populations with restricted gene flow (Thomas et al., 2017). Contrary to these previous studies, a seascape genomics approach should cover ecologically meaningful spatial scales and be able to distinguish the pressures caused from different climatic conditions, as well as account for confounding effects of demographic processes (Balkenhol et al., 2017). Of note, recent studies showed that combining population genomics analyses with demographic simulations allows to estimate adaptive potential in corals and provide valuable information for reef preservation (Matz et al., 2018, 2019). A

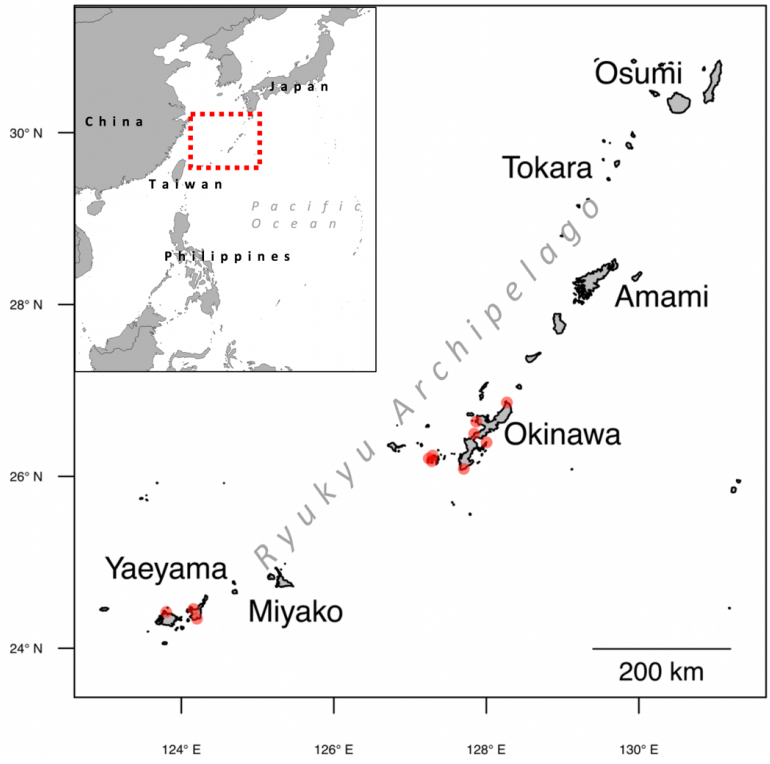


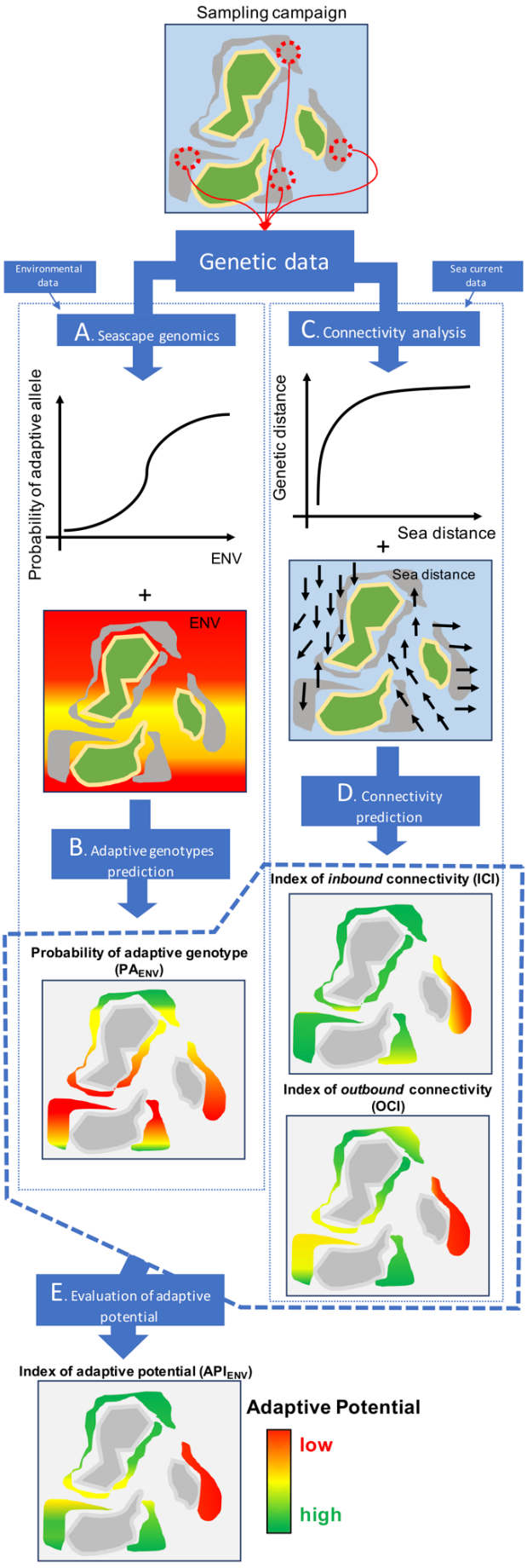
Figure 3-1. Study area. The Ryukyu Archipelago extends for more than 1,000 km in the north-western Pacific Ocean. The red circles display the 11 sites where samples were collected for the seascape genomics analysis (adapted from Shinzato, Mungpakdee, Arakaki, & Satoh, 2015).

similar approach can be used to transpose findings of seascape genomics studies to inform conservation strategies.

In the present study, we applied a seascape genomics framework to detect coral reefs that are carrying potentially heat stress adapted genotypes, and in turn, to show how conservation policies could implement the results. Our study focuses on *Acropora digitifera* of the Ryukyu Archipelago in Japan (Fig. 3-1), an emblematic species of the Indo-Pacific and flagship organism for studies on corals genomics (Shinzato et al., 2011). We first analysed the convergence between genomic and environmental information to i) detect SNPs potentially conferring a selective advantage, and ii) develop a model describing connectivity patterns. Next, we took advantage of these findings to infer which reefs were more likely to be carrying heat stress adapted genotypes and to evaluate their interconnectedness with the rest of the reef system. Finally, we propose an approach to implement the results obtained into conservation planning. Overall, our work provides tools for the interface between conservation genomics and marine environmental sciences, which are likely to empower preservation strategies for coral reefs into the future.

3.3. Material and Methods

Our framework is structured on two axes of analysis and prediction: one focusing on the presence of putative heat stress adapted genotypes (seascape genomics), and the other on population connectivity (Fig. 3-2). The seascape genomics analysis (Fig. 3-2A) combines



genomic data with environmental information to uncover potentially adaptive genotypes at sampling sites. The models describing these relationships are then used to predict, at the scale of the whole study area, the probability of the presence of heat stress adapted genotypes (Fig. 3-2B). In the connectivity study (Fig. 3-2C), we designed a model describing how

Figure 3-2. Workflow between the steps of the approach. The starting point for the analysis is the generation of genetic data describing the genotypes observed at different sampling locations (in this example, 4 sampling sites). In the seascape genomics analysis (A), these data are combined with environmental information to uncover genotypes whose frequencies are associated with specific climatic conditions (ENV). Such genotypes are defined as potentially adaptive against the environmental condition of interest. The model describing this link is then applied to environmental data at the scale of the whole reef system (B), to predict the probability of presence of the adaptive genotypes (green: high probability; red: low probability). The genetic data are also combined with sea current information to build a connectivity model (C) describing how sea distances correspond to genetic separation between sampling sites. This model is fitted with sea distance between all the reefs of the study area to predict (D) patterns of connectivity from (outbound) and to (inbound) each reef (green: high connectivity; red: low connectivity). Finally, predictions of the presence of adaptive genotypes and connectivity patterns are combined to assess the adaptive potential across the study area (E): reefs that are connected with sites that are likely to carry the adaptive genotype will have a higher adaptive potential (green), while those that are isolated will have lower adaptive potential (red).

distances based on sea currents (calculated on the basis of remote sensing data) correspond to the genetic separation between corals at these sites. This model is then used to predict connectivity of sites at the study area scale (Fig. 3-2D). Finally, the predictions of where the heat stress adapted genotypes are more likely to exist, and of how the reef system is interconnected, allow the assessment of adaptive potential across the whole study area (Fig. 3-2E).

3.3.1. Genomic dataset

The genomic data used come from a publicly available dataset consisting of 155 geo-referenced colonies of *A. digitifera* from 12 sampling locations (13±5 colonies per site) of the Ryukyu Archipelago in Japan (Fig. 3-1; Bioproject Accession PRJDB4188). These samples were sequenced using a Whole-Genome Sequencing approach in the scope of a population genomics study. Details on how samples were collected and processed for genomic analysis can be found in Shinzato et al. (2015).

Raw genomic data were processed using the Genome Analysis Toolkit framework (GATK; McKenna et al., 2010) following the recommended pipeline (the "GATK Best Practices"; Van der Auwera et al., 2013) with the necessary modifications for coping with the absence of reliable databases of known variants for this species. In short, the *A. digitifera* reference genome (v. 1.1, GenBank accession: GCA_000222465.2; Shinzato et al., 2011) was indexed using bwa (v. 0.7.5a, Li & Durbin, 2009), samtools (v. 1.9, Heng Li et al., 2009) and picard-tools (v. 1.95, http://broadinstitute.github.io/picard) and raw sequencing reads were aligned using the bwa *mem* algorithm. The resulting alignments were sorted, marked for duplicate reads, modified for read-group headers and indexed using picard-tools. Next, each alignment underwent an independent variant discovery using the GATK HaplotypeCaller tool (using the ERC mode and setting the --minPruning flag to 10) and genotypes were then jointly called by the GATK GenotypeGVCFs tool in random batches of 18 samples to match our computational power (18 CPUs). The variant-calling matrices of the different batches were then joined and filtered in order to keep only bi-allelic Single Nucleotide Polymorphisms (SNPs) using the GATK CombineVariants and SelectVariants tools, respectively. This resulted in a raw genotype matrix counting ~1.2 M of SNPs. Subsequently, we used the GATK VariantAnnotator tool to annotate variants for Quality-by-depth and filtered for this value (<2), read coverage (>5 and <100 within a sample), minor allele frequency (>0.05), major genotype frequency (<0.95) and missing rate of both individuals and SNPs (<0.1) using the GATK VariantFiltrationTool and custom scripts in the R environment (v. 3.5.1, R Core Team, 2016). Finally, we filtered for linkage disequilibrium using the *snpgdsLDpruning* function of the SNPrelate R package (v. 1.16, LD threshold=0.3; Zheng et al., 2012). This pipeline produced the filtered genotype matrix consisting of 136 individuals and 7,607 SNPs.

Natural hybridization and transient species boundaries have been observed in *Acropora* species (van Oppen, Willis, Van Rheede, & Miller, 2002), and might cause bias in the analysis

of adaptation and connectivity. For this reason, we investigated the presence of these phenomena by running a preliminary analysis of fixation index (F_{ST}) variation by genomic position using the R KRIS package (v. 1.1; Chaichoompu et al., 2018). Since we found no genomic islands of low-recombination (*i.e.* high F_{ST}; Nosil et al., 2009) when comparing the populations of Kerama, Yaeayama and Okinawa (Fig. S3-1) we excluded the possibility of presence of genetically isolated groups in the dataset. Importantly, previous studies on this coral population did not report hybridization with other species, neither the presence of cryptic species nor isolated sub-populations (Nakajima, Nishikawa, Iguchi, & Sakai, 2010; Nishikawa, 2008; Shinzato et al., 2015).

3.3.2. Environmental data

Seascape genomics analyses require an exhaustive characterization of the environment in order to distinguish the effect of collinear gradients (Leempoel et al., 2017; Rellstab et al., 2015; Riginos et al., 2016). Six georeferenced datasets describing atmospheric and seawater conditions were retrieved from publicly available resources (EU Copernicus Marine Service, 2017; National Oceanic and Atmospheric Administration, 2017; Tab. S3-1). All these datasets provided records over several years (on average 15) before the genetic data were sampled (2010; Shinzato et al., 2015), covering the entire study area (Fig. 3-1) with a spatial resolution ranging from ~9 km to 4 km (Tab. S3-1). Four of these datasets (sea surface temperature, salinity, chlorophyll concentration and current velocity) were captured at a daily temporal resolution, while the other two (suspended particulate matter and photosynthetically available radiations) provided monthly averages. We processed these variables in the R environment using the *raster* package (v. 2.8, Hijmans, 2016) to compute for each: i) the overall average; ii) the highest monthly average, iii) the lowest monthly average. For the four variables captured at a daily temporal resolution, we also computed the standard deviations associated with the three averages.

Furthermore, sea surface temperature measurements were used to compute the bleaching alert frequency (BAF), representing the percentage of days (over the 23 years of remote sensing) during which the heat stress (Liu et al., 2003) accumulated over 2 weeks exceeded 4 °C. Sea surface temperature and salinity records were combined in a polynomial equation to produce estimates of seawater alkalinity (Lee et al., 2006). Bathymetry data (Ryan et al., 2009) were used to retrieve the depth at sampling locations. Finally, population density data (CIESIN Columbia University, 2010) were averaged in a 50 km buffer area to produce a surrogate-variable for anthropogenic pressure (Welle, Small, Doney, & Azevedo, 2017). In total, 39 environmental variables were computed.

We used the geographic coordinates associated with each sample to characterize the environmental conditions using the QGIS point sampling tool (v. 2.18.25, QGIS development team, 2009). For the predictive step of our study (Fig. 3-2C) at the scale of the whole reef system we retrieved the shapes of the reefs of the region (UNEP-WCMC, WorldFish-Center,

WRI, & TNC, 2010) and reported them into a regular grid (cell size of 5x5 km) using QGIS. For the reef cells smaller than 5 km^2, we calculated the actual area (in km^2), as it will be required for the computation of connectivity and adaptive potential indices. Reefs from the neighbouring regions (Taiwan and Philippines, Fig. 3-1) were also included to avoid border-effects in computations. Environmental conditions were assigned to each reef cell using the average function of the QGIS zonal statistics tool.

3.3.3. Seascape genomics

The seascape genomics analysis was carried out to investigate the possible correlation between environmental variables and the frequency of particular genotypes. Associations of this kind might reveal an environmental constraint requiring adaptation in *A. digitifera*, as well as the genetic features conferring the selective advantage.

We performed the genotype-environment association analysis using the logistic regression method implemented within the SamBada software (v. 0.7; Duruz et al., 2019; Stucki et al., 2017). The SamBada approach allows proxy variables of genetic structure to be included in the analysis in order to avoid possible confounding effects (patterns of neutral genetic variation mimicking signals of adaptation to the local environment; Holderegger et al., 2008). Here we performed a discriminant analysis of principal components (DAPC) on the SNPs genotype matrix using the R package *adegenet* (v. 2.1.1; Jombart, 2008). This procedure highlighted a main separation between two groups of samples along the first discriminant function (Fig. S3-2). The latter was therefore used as co-variable in association models.

The genotype-environment association analysis with SamBada evaluated 890,019 association models (39 environmental variables matched against the 3 genotypes of the 7,607 bi-allelic SNPs). For each association-model related to the same environmental variable, p-values of G-scores (G) and Wald-scores (W) were corrected for multiple testing using the R *q-value* package (v. 2.14, Storey, 2003). Association models scoring $q<0.01$ for both statistics were deemed significant. If a SNP was found in more than one significant association (*e.g.* with collinear environmental variables), only the best model (according to the value of G) was kept. This best association model is hereafter referred to as the significant genotype-environment association (SGEA).

3.3.4. Annotation of seascape genomics results

Since landscape/seascape genomics analysis can suffer the issue of false positives, it is important to use a complementary method to strengthen SGEAs (Rellstab et al., 2015). In this work, we annotated the genomic neighbourhood of each SGEAs and verified whether the molecular functions of the genes surrounding a SNP were coherent with a presumptive adaptive role.

We set the maximum size of the search window to ±250 kbs around the concerned SNP of each SGEA. This maximal window size was selected because genes(s) possibly linked to a mutation may lay up to hundreds of kbs away (Brodie et al., 2016; Visel, Rubin, & Pennacchio, 2009), and this window size corresponds approximately to the scaffold N50 statistics of the reference genome (*i.e.* half of the genome is contained within scaffolds of this size or longer). For every SGEA, the annotation procedure was performed as follows. Based on the NCBI annotation of the reference genome (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Acropora_digitifera/100/), we retrieved all the predicted genes falling within the ±250 kbs window. Next, we retrieved the predicted protein sequences related to these genes and ran a similarity search (blastp, (Madden & Coulouris, 2008) against metazoan protein sequences in the swissprot database (release 2019_07; Boeckmann et al., 2003). For every predicted gene, only the best significant match (E-score threshold < 10^{-7}) was kept. Finally, predicted genes were annotated with the eukaryotic cluster of orthologous genes (KOG; Jensen et al., 2008) annotation from the matching swissprot entry. For every KOG we calculated the relative frequency across the *A. digitifera* genome. This was obtained by dividing the genome into 500 kbs windows and by calculating the percentage of windows in which the KOG was observed.

3.3.5. Probability of presence of heat stress adapted genotypes

The seascape genomics analysis pointed out genotypes expected to confer a selective advantage under a determined environmental condition. Furthermore, the SamBada approach provided, for every SGEAs, the parameters of a logistic regression that links the probability of occurrence of the adaptive genotype with the value of the environmental variable (Fig. S3-3; Stucki et al., 2017). These logistic models can therefore be used to estimate the probability of presence of the genetic variant for any value of the environmental variable at any reef of the Ryukyu Archipelago (Joost, 2006; Rochat & Joost, 2019).
For SGEAs related to a same environmental pressure, these single genotype probabilities can be combined into an average probability (*i.e.* the arithmetical mean) of carrying genotypes adapted to a specific condition (PA_{env}). In this study, we applied this calculation to a group of SGEAs related to heat stress (high bleaching alert frequency) that showed functional annotations coherent with a role in heat response (SGEA3, 5-8 and 13; Tab. 3-1). The resulting value was the probability of carrying heat stress adapted genotypes (PA_{heat}).

3.3.6. Sea current data

The starting point for the connectivity analysis and prediction was the evaluation of how pairs of reefs are expected to be connected by water flow. This step was carried out by processing remote sensing data on water current to construct a matrix that defines the costs of transitions from one reef to another.

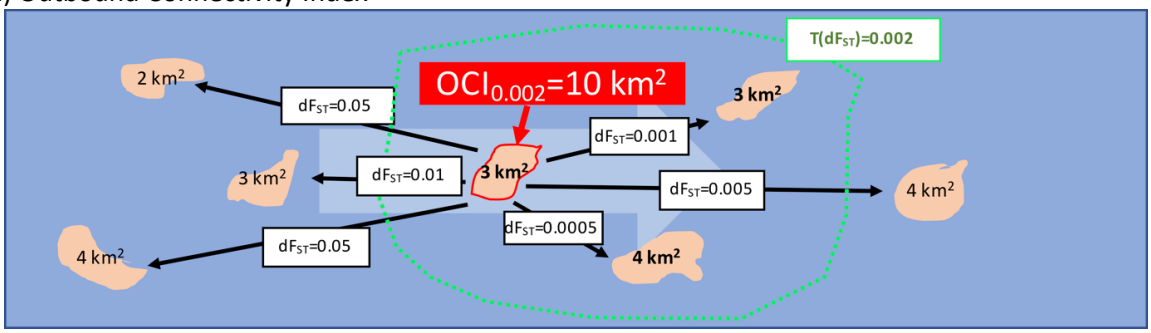
Daily records of sea surface current were retrieved from publicly available databases (zonal and meridional surface velocities from the *global-reanalysis-phy-001-030* product; EU Copernicus Marine Service, 2017) and used to compute the direction and speed of currents in the R environment using the *raster* library. By using the resample function of the R *raster* library, we downscaled these data from original 0.083° (~9.2 km) to 0.015° (~1.6 km) and corrected land pixels (*i.e.* removing sea current values) using a high-resolution bathymetry map (Ryan et al., 2009). These day-by-day records of sea currents (from 1993 to 2010) were then stacked to retrieve, for each pixel of the study area, the cumulative speed toward each of the eight neighbouring pixels. For every pixel, the cumulative speed in each of the eight directions was divided by the total speed (the sum of the eight directions) to obtain the probability of transition in each direction (the conductance). This information was used to calculate dispersal costs (the inverse of the square conductance) and was summarized in a transition matrix in the format of the R *gdistance* package (v. 1.2, van Etten, 2018).
For the connectivity analysis (Fig. 3-2C), the transition matrix was used to calculate sea distances (*i.e.* the least-cost path) between sampling sites of the genotyped colonies. For the connectivity predictions (Fig. 3-2D), we calculated the sea distances between all the reefs of the study area (the 5x5 km cells described in the environmental variables section). Importantly, for each pair of reefs (for instance $reef_1$ and $reef_2$) two sea distances were computed, one for each direction (*i.e.* from $reef_1$ to $reef_2$ and from $reef_2$ to $reef_1$). The result of this calculation was an asymmetrical square matrix describing sea distance between any reef cell.

3.3.7. Connectivity analysis

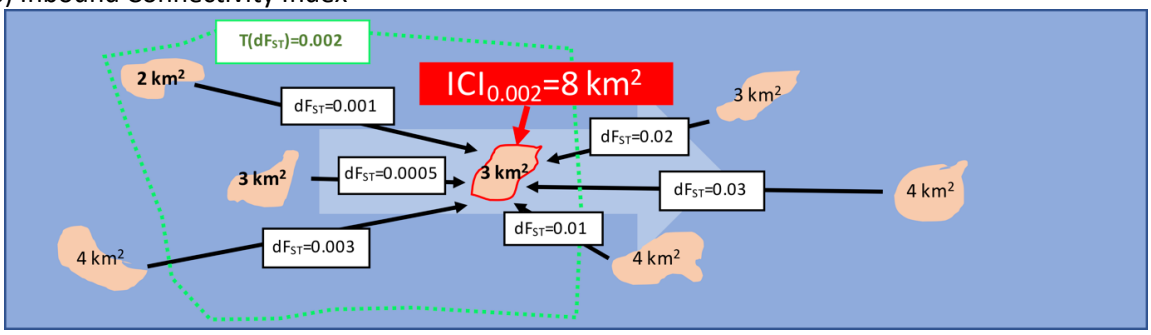
The connectivity analysis was performed to estimate how a unit of sea distance between two reefs is translated in terms of genetic separation between *A. digitifera* colonies. This step is necessary because sea distance does not account for the biological differences (for instance differential larval survival period) between different species.
Genetic distances between sampling sites were calculated using the pairwise F-statistics (F_{ST}; Weir & Cockerham, 1984) as implemented in the R *hierfstat* library (v. 0.04; Goudet, 2005). When there is no dispersal constraint between two sub-populations, the related F_{ST} is equal to zero. Conversely, when dispersal is constrained, F_{ST} increases up to a maximum value of one (isolated sub-populations). To avoid bias due to low sample sizes, we only considered sample sites with more than 10 samples each (7 out of 12).
Next, we built a linear model (hereafter referred to as the connectivity model) to estimate F_{ST} from the shortest sea distance (least cost path) between each pair of sample sites. As a comparison, we built a connectivity model using Euclidean distances of coordinates (aerial distances) as independent variable while maintaining F_{ST} as response variable. The quality of models was estimated by calculating the coefficients of determination (R^2) and the Akaike information criterion (AIC; Bozdogan, 1987).

Figure 3-3. Calculation of connectivity and adaptive potential indices. The three maps display a hypothetical seascape with seven reefs (in rose) of different extent and connected by sea current flowing from left to right (large light blue arrow). On each map, a different index is calculated for the same focal reef (highlighted in red): a) outbound connectivity index (OCI), b) inbound connectivity index (ICI), c) adaptive potential index (API). The black arrows display the estimated directional genetic separation (dF_{ST}) for corals traveling from (a) and toward (b, c) the focal reef. The calculation of the indices requires that a threshold value for dFst is set (in this example, T(dFst)=0.002, the green border) in order to define the reefs neighbouring the focal one. OCI (a) represents the total area (in km^2) of neighbouring reefs (destinations) that can be reached from the focal reef (departure). ICI (b) represents the total area of neighbouring reefs (departures) that can reach the focal reef (destination). API (c) is a special case of ICI, where the area of the neighbouring reefs is weighted by their probability of presence of adapted genotypes (PA).

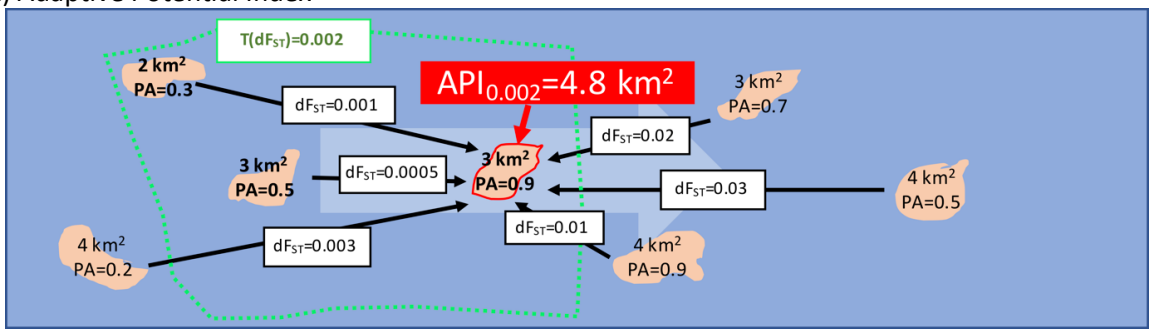
a) Outbound Connectivity Index



b) Inbound Connectivity Index



c) Adaptive Potential Index



3.3.8. Connectivity predictions

The model that was developed during the connectivity analysis describes how a unit of sea distance is translated into a unit of genetic separation (F_{ST}) in *A. digitifera* (Fig. S3-4). Since we previously characterized the sea distances between any reef of the Ryukyu Archipelago, here we translated such physical distances into predicted degrees of genetic separation. This

transformation was applied to the asymmetrical square matrix describing sea distances between any reef cell of the study area. The resulting matrix contains the corresponding directional estimates of genetic separation (dFst; Fig. S3-5) and is employed to calculate two indices that summarize connectivity for every reef cell:

- *outbound connectivity index* (OCI; Fig. 3-3a): OCI describes how a specific reef (departure reef) is expected to disperse toward neighbouring reefs (destination reefs). More specifically, OCI represents the total area (in km^2) of neighbouring destination reefs that can be reached from the departure reef within a determined dF_{ST} distance.
- *inbound connectivity index* (ICI; Fig. 3-3b): ICI describes how a specific reef (destination reef) is expected to receive recruits from neighbouring reefs (departure reefs). More specifically, ICI represents the total area (in km^2) of neighbouring departure reefs that can reach the destination reef within a determined dF_{ST} distance.

These connectivity indices and their interpretation are subordinate to the dF_{ST} threshold applied in the calculation. For this reason, it is crucial to set this threshold by considering the size of the study area and the distribution of the dF_{ST} values observed (Fig. S3-5). In this work, we set the dF_{ST} threshold to 0.02. In fact, a smaller dF_{ST} (for instance 0.01; Fig. S3-5) would have informed on local connectivity only (within neighbouring islands) and neglect connectivity at the scale of the Ryukyu Archipelago. In contrast, a higher dF_{ST} (for instance 0.05, Fig. S3-5) would have exceeded the study area boundaries, causing bias (border effects) in the calculation of the indices for reefs of the southern Islands (Yaeyama and Miyako) of the Archipelago.

3.3.9. Evaluation of the adaptive potential against heat stress

The adaptive potential against heat stress was evaluated by combining the predictions of the presence of heat stress adapted genotypes (PA_{heat}) and connectivity patterns (ICI) in an *index of adaptive potential* against heat stress (API_{heat}, Fig. 3-2E). Indeed, API_{heat} is a special case of ICI calculated as the sum of the weighted area (in km^2) of all the reefs connected under a specific dF_{ST} threshold to the focal reef (Fig. 3-3c). The weight applied to each reef corresponded to the probability of carrying heat stress adapted genotypes (PA_{heat}). For the dF_{ST} threshold, we used the same value (0.02) as employed in the ICI and OCI calculations.

3.4. Results

3.4.1. Seascape genomics

We detected 18 significant genotype-environment associations (SGEA, q_G and q_W<0.01, Tab. 3-1) spanning across 17 distinct scaffolds of the *A. digitifera* reference genome. Among them, 14 were related to bleaching alert frequency (BAF), two to lowest average monthly salinity (SSS) and two to lowest monthly average alkalinity (AT).

Table 3-1. Significant genotype-environment associations (SGEA). The seascape genomics analysis using the SamBada method detected 18 significant (qG and qW<0.01) genotype-environment associations (SGEA). This table shows, for each SGEA, the genomic position of the concerned SNP (in the format scaffoldID:position; Position), the q-values related to the G-score (G) and the Wald-score (W) of the association model, the concerned environmental variables (BAF: bleaching alert frequency, SSS.LM: lowest average monthly salinity, AT.LM: lowest average monthly alkalinity; Env. Var.), the eukaryote cluster of orthologous genes (KOGs) annotated within ± 50 kb (light grey), ± 100 kb (grey) and ±250 kb (dark grey) around the concerned SNP. For every KOG annotation, the frequency of the term across the reference genome is given in brackets.

ID	Position	q-values	Env.Var	Annotations KOGs (±50 kb, ±100 kb, ±250 kb)
SGEA1	NW_015441080.1: 208400	G: 1.13E-09 W: 5.53E-05	BAF	KOG4193: G- protein-coupled receptor (0.0593), KOG0777: geranylgeranyl diphosphate synthase (0.0014) | KOG0120: splicing factor U2AF (0.0014), KOG0157: Cytochrome p450 (0.0108), KOG3656:receptor (0.3245), KOG2358:NFU1 (0.0018)
SGEA2	NW_015441080.1: 963851	G: 1.49E-07 W: 1.84E-03	BAF	KOG4291: sushi domain containing 2 (0.0022) | KOG4475:PTK7 protein tyrosine kinase 7 (0.0718), KOG3880: Involved in vacuolar transport and vacuole pH homeostasis (0.012) | KOG3588: chondroitin sulfate (0.0242), KOG1836: Laminin, alpha (0.0094), KOG4523: mef2b neighbor (6e-04), KOG3848: PleXin Domain containing (0.003)
SGEA3	NW_015441121.1: 665651	G: 1.72E-07 W: 2.03E-04	BAF	KOG0192: protein kinase (0.0357), KOG0619: leucine rich repeat (0.0615), KOG3744: jnk1 mapk8-associated membrane protein (9e-04)
SGEA4	NW_015441261.1: 566971	G: 3.36E-06 W: 1.26E-03	SSS.LM	---
SGEA5	NW_015442197.1: 32233	G: 4.06E-06 W: 1.04E-03	BAF	KOG0351: DNA helicase (0.0409), KOG4373: Exonuclease 3'-5' domain containing 2 (0.0134)
SGEA6	NW_015441195.1: 470076	G: 5.82E-06 W: 1.84E-03	BAF	KOG2989: Coiled-coil domain-containing protein (0.0019), KOG0278: serine threonine kinase receptor associated protein (0.0019), KOG0583: serine threonine-protein kinase (0.0221), KOG0351: DNA helicase (0.0409), KOG4373: Exonuclease 3'-5' domain containing 2 (0.0134) | KOG2745: mitochondrial carrier (9e-04), KOG1497: cop9 signalosome complex subunit (9e-04) | KOG4441 :kelch-like (0.0512), KOG2111: WD repeat domain phosphoinositide-interacting protein (0.0043), KOG1028: synaptotagmin (0.0141), KOG3656: receptor (0.3245), KOG0452: iron-responsive element binding protein 2 (9e-04), KOG2106: Echinoderm microtubule associated protein like (0.0064)
SGEA7	NW_015441282.1: 27616	G: 5.82E-06 W: 8.09E-04	BAF	KOG0351: DNA helicase (0.0409), KOG4373: Exonuclease 3'-5' domain containing 2 (0.0134)
SGEA8	NW_015441785.1: 16151	G: 6.59E-06 W: 1.04E-03	BAF	KOG0351: DNA helicase (0.0409), KOG4373: Exonuclease 3'-5' domain containing 2 (0.0134)
SGEA9	NW_015441192.1: 602343	G: 1.06E-05 W: 8.11E-03	AT.LM	KOG4341: F-box and leucine-rich repeat protein (0.0184), KOG4581: UbiA prenyltransferase domain containing 1 (0.001) | KOG3627: protease (0.0241), KOG0759: Mitochondrial (0.0027), KOG3953: splA ryanodine receptor domain and SOCS box containing (8e-04)
SGEA10	NW_015441113.1: 326020	G: 1.40E-05 W: 7.54E-03	BAF	KOG4585: transposon protein (0.0527) | KOG3102: Phosphodiesterase (0.001), KOG3261: Mediates the side-chain deamidation (0.0018), KOG0910:Thioredoxin (0.0018), KOG4834:Chromosome 17 open reading frame 49 (0.0018), KOG1046:aminopeptidase (0.0057)
SGEA11	NW_015441391.1: 251497	G: 2.44E-05 W: 1.84E-03	BAF	KOG0811: SYNtaxin (0.0039) | KOG1840: Kinesin light chain (0.0176)
SGEA12	NW_015441600.1: 9407	G: 9.89E-05 W: 6.22E-03	SSS.LM	KOG4585: transposon protein (0.0527), KOG0196: Eph receptor (0.0114)
SGEA13	NW_015441190.1: 582812	G: 1.49E-04 W: 6.35E-03	BAF	KOG0351: DNA helicase (0.0409), KOG4585: transposon protein (0.0527) | KOG4088: Signal sequence receptor delta (9e-04)
SGEA14	NW_015441072.1: 291659	G: 1.67E-04 W: 6.14E-03	BAF	KOG3656: receptor (0.3245) | KOG0603: ribosomal protein s6 (0.0013), KOG2101: sorting nexin (0.004), KOG3738: UDP-N-acetyl-alpha-D-galactosamine polypeptide (0.0035), KOG2145: tryptophanyltRNA synthetase (0.0013), KOG4193: G- protein-coupled receptor (0.0593), KOG4729: LATrophilin (0.0205), KOG0583: serine threonine-protein kinase (0.0221)
SGEA15	NW_015441328.1: 255377	G: 1.73E-04 W: 6.22E-03	BAF	KOG4585: transposon protein (0.0527) | KOG1075: Retrotransposon protein (0.0045), KOG4776: Craniofacial development protein (0.0054), KOG0017: Retrotransposon protein (0.0022), KOG1721: Zinc finger protein (0.11)
SGEA16	NW_015442007.1: 107968	G: 2.26E-04 W: 6.22E-03	AT.LM	KOG0638: 4-hydroxyphenylpyruvate dioxygenase (0.0039)
SGEA17	NW_015441133.1: 148591	G: 6.04E-04 W: 6.91E-03	BAF	KOG4331: promini n (0.0036) | KOG3656: receptor (0.3245) | KOG1966: Sodium hydrogen exchanger (0.0015), KOG2618: Chromosome 3 open reading frame 37 (0.0015), KOG0694: protein kinase c (0.0136), KOG2363: ribonuclease p (0.0015), KOG1366: Complement component (0.0031)
SGEA18	NW_015442144.1: 516	G: 7.83E-04 W:7.90E-03	BAF	---

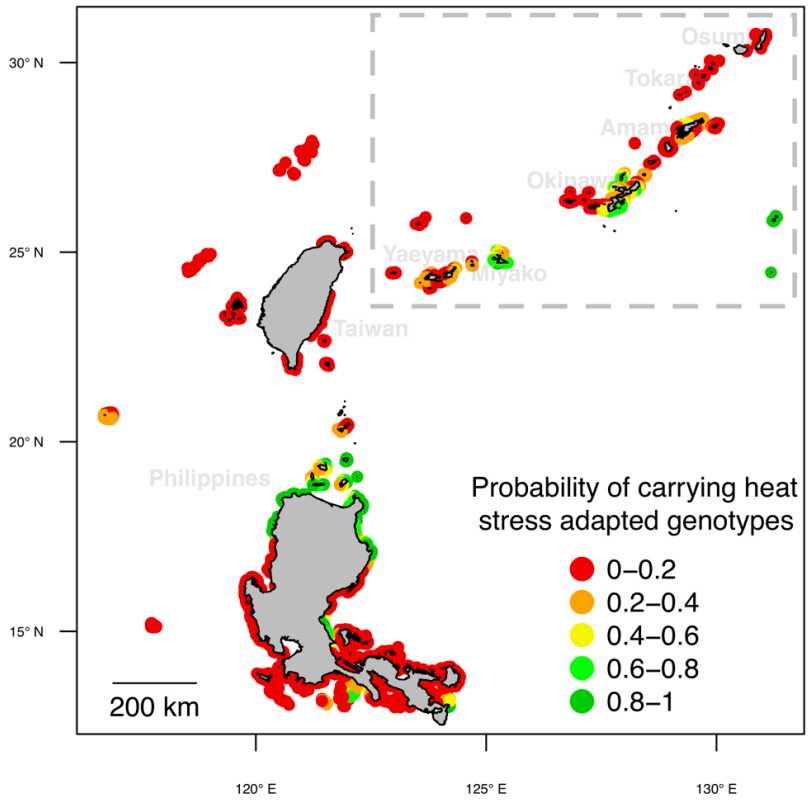


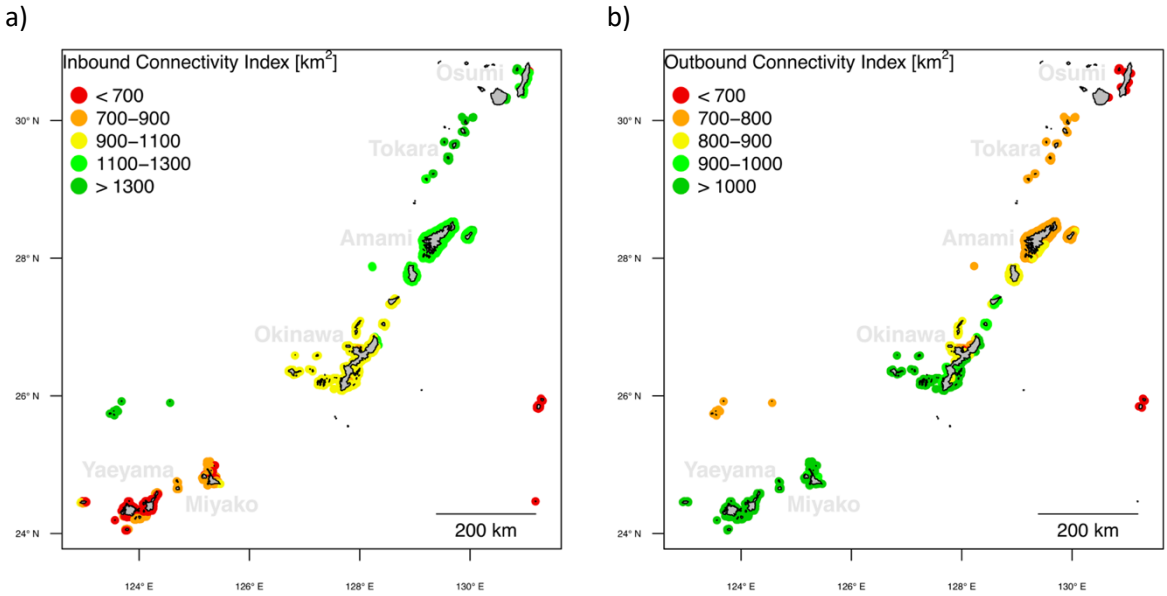
Figure 3-4. Probability of carrying heat stress adapted genotypes (PA_{heat}). The map shows the probability of presence of the genotypes expected to be linked to adaptation against heat stress across the study area and the neighbouring regions. Seven significant gene-environment associations (SGEA1, 3, 5-8 and 13, Tab. 3-1) describing the association between distinct genotypes and bleaching alert frequency were used to predict expected genotypes frequencies. These expected frequencies were then averaged to compute the cumulated probability of adaptive genotypes. The dashed box highlights the position of the Ryukyu Archipelago.

The functional annotations surrounding SNPs involved in SGEAs showed that in nine cases the closest genes belonging to eukaryotic clusters of orthologs (KOGs) fell within a ± 50 kb window, in two within ± 100 kb, in five within ± 250 kb and in two over ± 250 kb (Tab. 3-1). In total, 64 KOGs were annotated and some recurred in SNPs from different SGEAs, such as *DNA helicases* (in SGEA5-8 and 13, all related to BAF), *transposon protein* (SGEA10, 12, 13 and 15), *Exonuclease 3'-5' domain containing* (SGEA5-8, all related to BAF), *serine threonine-protein kinase* (SGEA6 and 14, both related to BAF) and *G- protein-coupled receptor* (SGEA1 and 14, both related to BAF). The remaining KOGs were observed only once, and among those expected at lowest frequency (<0.001 per ± 250 kb window) across the *A. digitifera* genome we found *jnk1 mapk8-associated membrane protein* (SGEA3, associated with BAF), *mitochondrial carrier and iron-responsive element binding protein 2* (SGEA5, associated with BAF), *splA ryanodine receptor domain* (SGEA9, associated with AT), *Signal sequence receptor delta* (SGEA13, associated with BAF).

3.4.2. Probability of presence of heat stress adapted genotypes

The SGEAs of the seascape genomics analysis were then used as the starting point for predicting the probability of presence of heat stress adapted genotypes (PA_{heat}) across the reefs of the region. For the calculation of this probability, we employed six SGEAs (SGEA3, 5-8

Figure 3-5. Connectivity indices. The maps show the potential connectivity to (a) and from (b) every reef of the Ryukyu Archipelago. In (a), the inbound connectivity index (ICI) represents the total area (in km^2) of the reefs that are connected to the focal reef with a dF_{ST}<0.02 (dF_{ST} toward the focal reef). Reefs with a high ICI are expected to receive recruits from a larger neighbourhood. In (b), the outbound connectivity index (OCI) displays the total area of the reefs that are connected from the focal reef with a dF_{ST}<0.02 (dF_{ST} from the focal reef). Reefs with a high OCI are expected to disperse toward a larger neighbourhood.



and 13) related to bleaching alert frequency that displayed functional annotations coherent with a role in heat stress resistance (Tab. 3-1).

The average of PA_{heat} ranged from 0 to 1 (Fig. 3-4). In Ryukyu Archipelago, PA_{heat} was higher in Miyako ($\overline{PA}_{heat_{Miyako}} = 0.47 \pm 0.21$) and Okinawa ($\overline{PA}_{heat_{Okinawa}} = 0.33 \pm 0.21$), lower in Amami ($\overline{PA}_{heat_{Amami}} = 0.18 \pm 0.12$) and Yaeyama ($\overline{PA}_{heat_{Yaeyama}} = 0.18 \pm 0.09$), and close to zero in the north of the region (Tokara and Osumi; $\overline{PA}_{heat_{Tokara}} = 0.02 \pm 0.03$, $\overline{PA}_{heat_{Osumi}} = \sim 0$; Fig. 3-4). Outside the Ryukyu Archipelago, a high PA_{heat} (>0.8) was predicted in northern Philippines while reefs around Taiwan displayed in general low PA_{heat} (<0.2; Fig. 3-4).

3.4.3. Connectivity modelling

The connectivity model used for the calculation of the connectivity indices accounted for 72% of the F_{ST} variation (R^2=0.72, AIC=-234; Fig. S3-4a) and resulted as a more accurate model when compared to the one based on aerial distance (R^2=0.66, AIC=-230, Fig. S3-4b).

The ICI variation followed a north to south decrease (Fig. 3-5a). The reefs around the islands in the north of the archipelago (Osumi, Tokara and Amami) were generally those with the highest ICI ($\overline{ICI}_{Tokara} = 1615 \pm 229\ km^2$; $\overline{ICI}_{Amami} = 1209 \pm 28\ km^2$; $\overline{ICI}_{Osumi} = 1164 \pm 336\ km^2$; Fig. 3-5a). In the central area (Okinawa), ICI was lower ($\overline{ICI}_{Okinawa} =$

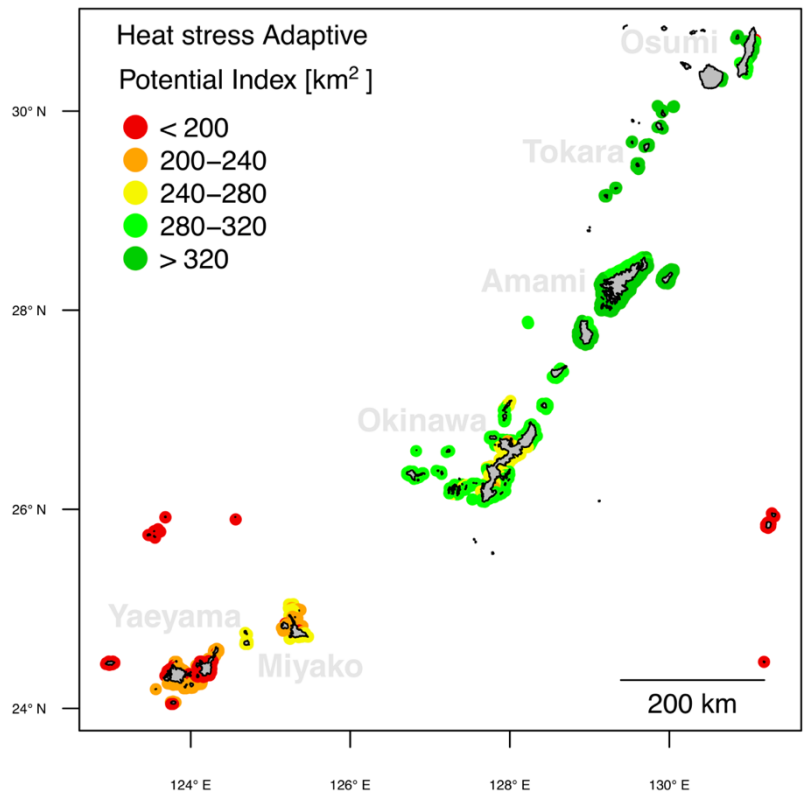


Figure 3-6. Index of adaptive potential against heat stress (API_{heat}). The map displays the index of adaptive potential against heat stress (high bleaching alert frequency, BAF) for every reef of the study area. This index represents the sum of weighted areas of reefs connected to the focal reef with a pFst<0.02 (pFst toward the focal reef). The weight applied corresponds to the probability of carrying heat stress adapted genotypes (PA_{heat}). Reefs with a large API are expected to receive more heat stress adapted recruits.

$999 \pm 42\ km^2$), while the lowest ICI values were observed in the southern area (Yaeyama and Miyako; $\overline{ICI}_{Miyako} = 777 \pm 71\ km^2$; $\overline{ICI}_{Yaeyama} = 674 \pm 76\ km^2$; Fig. 3-5a).
With regards to OCI, we observed a decrease in index with increasing latitude (Fig. 3-5b). OCI was highest in the southern half of the archipelago (Yaeyama, Miyako and Okinawa; $\overline{OCI}_{Yaeyama} = 1014 \pm 2\ km^2$; $\overline{OCI}_{Miyako} = 1008 \pm 14\ km^2$; $\overline{OCI}_{Okinawa} = 936 \pm 91 km^2$; Fig. 3-5b). A lower OCI was observed in the northern part (Amami and Tokara; $\overline{OCI}_{Amami} = 766 \pm 51\ km^2$; $\overline{OCI}_{Tokara} = 706 \pm 2\ km^2$; Fig. 3-5b), while the extreme north of the Archipelago (Osumi) had a very low OCI ($\overline{OCI}_{Osumi} = 6 \pm 4\ km^2$; Fig. 3-5b).

3.4.4. Evaluation of the adaptive potential

The variations of API_{heat} were generally structured along the latitudinal axis (Fig. 3-6). Reefs in the northern part of the Archipelago (Tokara, Amami and Osumi) generally showed the highest API_{heat} values ($\overline{API}_{heat_{Tokara}} = 335 \pm 6\ km^2$; ($\overline{API}_{heat_{Amami}} = 317 \pm 10\ km^2$; $\overline{API}_{heat_{Osumi}} = 296 \pm 86\ km^2$; Fig. 3-6). In the central part of the study area (Okinawa), API_{heat} was lower ($\overline{API}_{heat_{Okinawa}} = 279 \pm 12\ km^2$; Fig. 3-6) and in the southern part the lowest API_{heat} values were observed, ($\overline{API}_{heat_{Yaeyama}} = 200 \pm 17\ km^2$; $\overline{API}_{heat_{Miyako}} = 237 \pm 24\ km^2$; Fig. 3-6).

3.5. Discussion

3.6.2. Adaptation to heat stress

Heat stress is expected to be one of the major threats to coral reef survival, where the research for adaptive traits is becoming of paramount importance (Baums, 2008; Logan et al., 2014; Maina et al., 2011). In the present study, the seascape genomics analysis of *A. digitifera* of the Ryukyu Archipelago revealed the presence of 14 genomic regions hosting genetic variants that might confer a selective advantage against heat stress (Tab. 3-1). None of the SNPs related to the SGEA lay directly within a coding sequence of a putative gene, but this is rarely the case for causative-mutations (Brodie et al., 2016). In fact, genetic variants in intergenic regions that play a modulatory action on the expression of neighbouring genes are more frequent and can influence loci at a distance of 1-2 Mb (Visel et al., 2009). The fragmentation of the reference genome forced us to limit our search window to ±250 Kb around each SNP, yet we still found annotations corroborating a response to heat stress.
The SNP in SGEA3 was found to be related to KOG3744 (*jnk1 mapk8-associated membrane protein;* Tab. 3-1). This KOG is rare across the genome of *A. digitifera* (with an expected frequency of 0.0009 per 500 kbs window) and previous research corroborates the hypothesis that this gene plays a role in thermal adaptation. In fact, mitogen-activated-protein kinases (MAPKs) are proteins known to be involved in cellular responses to stress across a range of taxa (Neupane, Nepal, Benson, MacArthur, & Piya, 2013), and the c-Jun-N-terminal kinase (JNK) has previously been shown to be activated under thermal stress in the coral *Stylopohora pistillata* (Courtial et al., 2017).
In SGEA5-8 and 13 one KOG recurred in the annotations: KOG0351 (*DNA helicase*; Tab. 3-1). The expected frequency of this KOG is 0.04 per 500 kbs window and remarkably we found five of them in five distinct 500 kbs windows around SGEA associated with heat stress. Of note, in these five SGEAs *DNA helicase* were consistently the closest KOGs annotated around the SNPs concerned (Tab. 3-1). KOG0351 annotate a particular type of DNA helicases (swissprot IDs: Q91920, Q14191) known as “helicases Q” or “RecQ”, which are involved in the DNA repairing mechanism caused by UV-light damage in prokaryotes (Courcelle & Hanawalt, 1999), and for which light-stress driven effects were observed in eukaryotic cells as well (Sharma, Doherty, Brosh, & Jr, 2006). The modulation of this mechanism might therefore play a role in increasing *A. digitifera* resistance against light-stress associated with heat waves.

3.6.3. Connectivity patterns

Coral dispersal is driven by water flow (Paris-Limouzy, 2011), which is highly asymmetrical in this region (north-east oriented) due to the Kuroshio current (Nishikawa, 2008). As previously observed, the main patterns of migrations in this population occurs from the south-west to the north-east (Shinzato et al., 2015). Reefs in the southern part of the study area (Yaeyama

and Miyako) showed the lowest ICI values (Fig. 3-5a), suggesting a potential lack of recruits arriving from other reefs of the region. In fact, the genetic diversity of southern reefs of the Ryukyu Archipelago is likely to depend on the recruits arriving from the east-coast of Taiwan and the northern Philippines, which are located upstream of the Kuroshio current (Fig. S3-5a; Chen & Shashank, 2009).

In the previous study on this data (Shinzato et al., 2015), reefs from Yaeyama resulted as those with the lowest heterozygosity rates across the study area. This observation was attributed to a population bottleneck caused by the 1998 bleaching event, but it is worth noting that reefs on the west coast of Okinawa showed higher heterozygosity rates despite having suffered recurrent bleaching events since 1998 (Donner et al., 2017). The lower heterozygosity rates in Yaeyama therefore might reflect not only the effects of past bleaching, but also the relative isolation of these islands from the reefs of the region (Fig. 3-5a).

In line with the same previous observations (Shinzato et al., 2015), the OCI value showed (Fig. 3-5b) that the southern reefs (Yaeyama and Miyako) are those expected to be the most prominent source of recruits for the rest of the Archipelago. Given this crucial aspect, it is even more important to preserve southern reefs of the Ryukyu Archipelago from the risks of isolation (*e.g.* inbreeding depression; Keller & Waller, 2002).

3.6.4. Heat stress adaptive potential in the 2016 bleaching event

Reefs in islands of Miyako, Okinawa were those most likely to carry heat stress adapted genotypes (Fig. 3-4). Previous work reported severe bleaching in Okinawa in 1998 (Yamazato, 1999) and that adapted colonies might have resisted (Van Woesik, Irikawa, & Loya, 2004). In contrast, reefs in the northern part of the Archipelago (Amami, Tokara and Osumi) experienced bleaching with moderate severity during the 1998 event (Donner et al., 2017), which might explain why heat stress adapted genotypes are not expected at the same frequency (Fig. 3-4).

The heat stress adaptive potential index (API_{heat}) defines the convergence between the probability of carrying heat stress adapted genotypes with connectivity predictions (Fig. 3-6). Reefs in the northern part of the Archipelago (Amami, Tokara and Osumi) showed a higher API_{heat} compared to those in the southern half of the region (Okinawa, Yaeyama and Miyako). Two reasons may explain this result: 1) these northern reefs are located downstream (on the Kuroshio Current) of two areas where putative adapted reefs are frequent (Okinawa and Miyako; Fig. 3-4); 2) the region of Northern Philippines, hosting high density of putative adapted reefs (Fig. 3-4), is more connected to the northern part of the Ryukyu Archipelago than with the southern part (Fig. S3-6). This may also explain why, despite hosting putative heat stress adapted reefs (Fig. 3-3), the Miyako area showed among the lowest API_{heat} values of the Archipelago (Fig. 3-6).

Table 3-2. Field report of the 2016 mass bleaching event. The table shows the severity and mortality associated with the 2016 bleaching event as reported by Global Coral Reef Monitoring Network (Kimura, Tun, & Chou, 2018). For every region surveyed in this report (identified by an ID and a region name), we show the corresponding region in our study and the associated average API against heat stress (API_{heat}), the probability of presence of heat-stress adapted genotypes (PA_{heat}) and degree of heat stress in 2016 (estimated as the number of days under bleaching alert). Colour scales highlight the variation of the value of each variable.

ID	Region Name	Region Area (this study)	Bleaching [%]	Morality [%]	API_{heat} [km2]	PA_{heat}	Bleaching alert [# of days]
3	Amami Islands	Amami	8.5	2.1	318	0.21	66
4	Okinawa Island, East coast	Okinawa	21	0.7	286	0.52	74
5	Okinawa Island, West coast	Okinawa	13.1	4.3	276	0.30	78
6	Okinawa Outer Islands	Okinawa	48.4	13.5	283	0.60	78
7	Kerama Islands	Okinawa	7.3	5.4	282	0.07	80
9	Miyako Island	Miyako	68.8	31	239	0.52	87
10	Miyako Outer Reefs	Miyako	70.1	67.5	248	0.60	87
11	Ishigaki Island, East coast	Yaeyama	47.9	8.8	198	0.30	84
12	Ishigaki Island, West coast	Yaeyama	63.2	14.8	193	0.09	83
13	Sekisei Lagoon, North	Yaeyama	91.5	46.9	192	0.13	84
14	Sekisei Lagoon, East	Yaeyama	99.5	67.9	204	0.23	84
15	Sekisei Lagoon, Center	Yaeyama	94.9	49.7	206	0.19	84
16	Sekisei Lagoon, South	Yaeyama	98.2	50	218	0.16	84
17	Iriomote Islands	Yaeyama	94.3	34.8	202	0.23	84

In 2016, the first mass bleaching event occurred in Japan since Shinzato and colleagues published the genetic data re-analysed in this work (Kimura et al., 2018). Field surveys related to this bleaching event reported severe bleaching in Yaeyama (intensity up to 99%, mortality up to 68%) and in Miyako (intensity up to 70%, mortality up to 67%; Tab. 3-2). In Okinawa and Amami, the impact of this same bleaching event was moderate to mild (Okinawa: intensity up to 48%, mortality up to 13%; Amami: intensity 8% and mortality 2%; Tab. 3-2). Reefs predicted with low API_{heat} (the southern reefs) appeared to suffer more severe bleaching than those in the northern region (which showed higher API_{heat}; Fig. 3-6), but care must be taken in the interpretation due to the confounding role of sea temperature during 2016 (Tab. 3-2). Indeed, satellite records of sea temperature (EU Copernicus Marine Service, 2017) show that in 2016 the number of days under bleaching alert was higher in the southern part of the Archipelago (Yaeyama: ~84 days; Miyako: ~87 days) than in the northern region (Okinawa: ~76 days; Amami: ~66; Tab. 3-2). Nevertheless, when two sites had a comparable degree of heat stress, higher API_{heat} was generally associated with a reduced severity in bleaching. For instance, reefs in Kerama (Okinawa) and Ishigaki Island West (Yaeyama) suffered 80 and 83 days under bleaching alert in 2016, respectively, but the bleaching intensity in the Ishigaki Island was more than nine times higher than observed for Kerama (63% vs 7%), with a lower API_{heat} (193 km^2 vs 282 km^2; Tab. 3-2). Similarly, despite spending 87 days under bleaching alert, reefs in Miyako (API_{heat}~240 km^2) showed lower bleaching

intensity (70%) compared to those in the Sekisei Lagoon (Yaeyama, >95% bleaching) that were predicted with lower API_{heat} (~200 km^2).
While these field observations seem to corroborate our predictions on adaptive potential, it is important to consider that they do not refer specifically to *A. digitifera,* but to the coral community as a whole (Kimura et al., 2018). Additionally, other local stressors (for instance anthropogenic pollution) might have modulated the bleaching response (Ateweberhan et al., 2013). Future bleaching surveys, with larger sample sizes and bleaching data referring to the specific coral genus, might provide a more reliable ground for validating our predictions.

3.6.5. Limitations and future directions

Seascape/Landscape genomics studies are susceptible to high false discovery rates, especially when the cofounding role of neutral genetic variation is not accounted for (Rellstab et al., 2015). We coped with this issue by running seascape genomics models explicitly integrating demographic processes (Stucki et al., 2017). However, a sampling scheme adapted to seascape genomics (unlike the one used by Shinzato et al. 2015 who did not consider environmental variability) would have further increased sensitivity and lowered false discoveries (Riginos et al., 2016; Selmoni, Vajana, Guillaume, Rochat, & Joost, 2020). In an ideal situation, significant genotype-environment associations should be validated by running experimental assays such common garden or aquaria experiments (Krueger et al., 2017), reciprocal transplantation (Palumbi et al., 2014) and molecular analysis (Courtial et al., 2017) to ascertain the adaptive role.
As regards environmental information, the data we employed had a maximal spatial resolution of ~4 km. It is important to acknowledge that crucial drivers of coral survival (heat stress in particular) can vary considerably under the fine-scale structure (< 1 km) of a seascape (*e.g.* Bay & Palumbi, 2014). Future development of coral seascape genomics should therefore focus on implementing new approaches to describe environmental variation at finer scales (Riginos et al., 2016). For instance, the Landsat-8 satellite (U.S. Geological Survey, 2016) allows to evaluate thermal patterns at less than 100 m of resolution since 2013 (Vanhellemont, 2020), and could therefore represent a valuable input for future studies.
Another element to mention is that we employed a straightforward method to describe coral connectivity in order to facilitate the reproducibility of the analysis. However, there are more sophisticated approaches to describe both genetic and physical distances between reefs that might produce more accurate models of connectivity. For instance, recent works (Matz et al., 2018, 2019) showed that the use of the F_{ST} metric could be replaced with directional estimates of gene flow (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009) and the sea distances could be calculated out of forward-in-time dispersal simulations (Lett et al., 2008).
Finally, when calculating connectivity and adaptation indices, we assumed that the demographic and environmental patterns observed at the twelve sampling sites were representative for those of the whole archipelago and that *A. digitifera* were a ubiquitous

species. These generalizations might be source of bias in the calculation of the indices. For instance, the twelve sampling sites used in this study cover the higher half of BAF range observed across the Ryukyu Archipelago. Because of this, we might be missing adaptive processes necessary to cope with small to moderate heat stress (*i.e.* lower half of the BAF range). To avoid this situation in future studies, we suggest to verify these assumptions before starting the seascape genomics study and to define a sampling strategy that minimizes the risks of collecting an unrepresentative dataset (Selmoni, Vajana, Guillaume, Rochat, et al., 2020).

3.6.6. Application in conservation

Conservation policies require objective and quantifiable information to prioritize areas for intervention efforts (OECD, 2017). In this study, we presented an original framework to calculate indices matching these requirements to describe the connectivity and adaptive potential against heat stress of a flagship coral species of the north-western Pacific. Insights of this kind are essential for effective planning of coral conservation strategies (Baums, 2008; Logan et al., 2014; Palumbi, 2003; van Oppen et al., 2015).

As they are derived from a universal metric of population connectivity (F_{ST}; Weir & Cockerham, 1984), the indices we propose here are computable for any coral species. Thus, connectivity indices for different species can be compared or aggregated for conservation management planning within a region. Furthermore, each of the indices we propose is expressed in a tangible spatial unit (km^2) that allows for comparison between different datasets and areas.

As an example, the predictions from the connectivity indices can be used to support the planning of marine protected areas (MPAs). An ideal placement of an MPA should ensure that the connectivity to the rest of the reef system is optimal (Krueck et al., 2017; Thomas et al., 2014), and the OCI provides this information (Fig. 3-5b). Furthermore, the computation of the ICI (Fig. 3-5a) from a protected area to the rest of the reef system could be used to compare how different locations of MPAs may modify the connectivity to other specific regions.

Similarly, information on adaptive potential could be used to inform conservation strategies. For instance, an MPA could be established to protect reefs with a high PA_{heat} (*i.e.* those likely to carry the traits necessary to persist against heat waves) from local stressors. Alternatively, this information could support the planning and location of coral nurseries to reinforce the adaptive potential of a population (Baums, 2008; van Oppen et al., 2015). For instance, this could be done by transplanting corals from reefs with high PA_{heat} to reef with low API_{heat} (*i.e.* reefs that had not experienced heat stress and are less likely to receive heat-adapted corals via natural migration).

To date, the calculation of these indices can be performed using R scripts and codes (R Core Team, 2016) made publicly available in this research. In the future, however, this framework

should be transposed to a more user-friendly interface to facilitate its use by conservation managers.

3.6.7. Conclusions

This study highlights the value of a seascape genomics approach for supporting the conservation of corals. We applied it to a flagship coral species of the Ryukyu Archipelago and identified genetic variants that may underpin adaptation to heat stress. Coupling this information with a genetic analysis of connectivity made it possible to evaluate the adaptive potential at the scale of the entire study area. The outputs of this analysis are quantitative indices that could be used to support objective prioritization of reefs in conservation plans. This framework is transferable to any coral species on any seascape and therefore constitutes a useful conservation tool to evaluate the genomic adaptive potential of coral reefs worldwide.

3.6. Acknowledgements

We thank Annie Guillaume, François Bonhomme and the anonymous reviewers for the useful comments and suggestions provided during the redaction of this paper.

3.7. Data Archiving Statement

All the data and codes used in this article are publicly available on Dryad (doi:10.5061/dryad.qz612jm90).

4. Article C

Seascape genomics reveals candidate molecular targets of heat stress adaptation in three coral species

Oliver Selmoni[1,2], Gaël Lecellier[2,3], Hélène Magalon[4], Laurent Vigliola[2], Nicolas Oury[4], Francesca Benzoni[5], Christophe Peignon[2], Stéphane Joost[1] and Véronique Berteaux-Lecellier[2]

[1]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

[2]UMR250/9220 ENTROPIE IRD-CNRS-UR, Labex CORAIL, Noumea, New Caledonia.

[3]UVSQ, Université de Paris-Saclay, Versailles, France.

[4]UMR250/9220 ENTROPIE IRD-CNRS-Ifremer-UNC-UR, Labex CORAIL, Université de La Réunion, St Denis, France.

[5]Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

Preprint version of the article submitted to *Molecular Ecology:*
Selmoni, O., Lecellier, G., Magalon, H., Vigliola, L., Oury, N., Benzoni, F., Peignon, C., Joost, S. and Berteaux-Lecellier, V. (2020). Seascape genomics reveals candidate molecular targets of heat stress adaptation in three coral species. *Available on bioRxiv: https://doi.org/10.1101/2020.05.12.090050*

Contribution of the candidate:
As first author of this article, I contributed, in collaboration with the co-authors, to initiate the research, organize the sampling strategy, participate to the field campaign, perform the DNA purification in laboratory, process genomic data, and run the seascape genomics and annotation analyses. I wrote the first version of the manuscript, which was then critically revised by the co-authors before submission.

The work described here is the first part of the research performed in the frame of the SABLE project. In this work, we ran a seascape genomics study with methods similar to those applied to the Japanese case study previously presented (article B). There were, however, three important differences. First, the genomic dataset used in the analysis described hereunder is the outcome of a sampling campaign followed by wet-laboratory work. We followed the guidelines defined in article A to maximize environmental contrasts between sampling sites and thus facilitate statistical inference of genotype-environment association analysis (Leempoel et al., 2017; Manel et al., 2012)*. Second, samples were genotyped using a DArT-seq*

approach, which is a RAD-seq variant particularly suited for detecting SNPs (as the adaptive ones) located in active regions of the genome (Kilian et al., 2012). Third, in SABLE we studied local adaptation of three different species simultaneously, and this provided ground to consolidate the results via inter-specific comparisons. These three features brought the focus into an aspect overlooked in article B: the characterization of the molecular targets of heat stress adaptation.

*The main challenge in this work was the definition of a unique framework to run seascape genomics analyses on three different species. The reason for this is that one of the three species (*Acropora millepora*) has a different reproductive strategy than the other two (*Pocillopora damicornis *and* Pocillopora acuta*; Ayre & Hughes, 2000). Different reproductive strategies imply different degrees of population structure, which in turn imply different statistical power and false discoveries in seascape genomics analyses (Rellstab et al., 2015; Riginos et al., 2016). In addition, sample sizes differed between the three species, bringing another element unbalancing the statistical power of the analyses (Lotterhos & Whitlock, 2015). We tackled this issue by running the seascape genomics analysis using a statistical method with important computational requirements (LFMM; Frichot et al., 2013), but expected it to be robust even under complex demographic scenarios (Rellstab et al., 2015). This approach disclosed SNPs associated with heat stress in each of three species.*

We then evaluated whether genes surrounding heat-stress associated SNPs recurred between species. The main obstacle to this investigation was the lack of uniformity in genome annotations between different species. To cope with this problem, we re-annotated genes in the reference genomes by matching their coding sequences against a common database (Uniprot; Bateman et al., 2015). The use of this database also made it possible to retrieve the gene ontology (GO) terms describing the molecular function of genes (Ashburner et al., 2000). Since GO terms are standardized descriptors, it was possible to objectively evaluate the occurrence of specific molecular functions in genes neighbouring adaptive SNPs of every species.

Information on candidate molecular targets for heat stress adaptation is rare in coral studies (van Oppen et al., 2015)*. The work described in the article presented here paves the way to filling this gap. Furthermore, this article is a practical implementation of the guidelines defined in article A, and refines the seascape genomics approach described in article B.*

4.1. Abstract

Anomalous heat waves are causing a major decline of hard corals around the world and threatening the persistence of coral reefs. There are, however, reefs that had been exposed to recurrent thermal stress over the years and whose corals appeared tolerant against heat.

One of the mechanisms that could explain this phenomenon is local adaptation, but the underlying molecular mechanisms are poorly known.
In this work, we applied a seascape genomics approach to study heat stress adaptation in three coral species of New Caledonia (southwestern Pacific) and to uncover molecular actors potentially involved. We used remote sensing data to characterize the environmental trends across the reef system, and sampled corals living at the most contrasted sites. These samples underwent next generation sequencing to reveal single-nucleotide-polymorphisms (SNPs) of which frequencies associated with heat stress gradients. As these SNPs might underpin an adaptive role, we characterized the functional roles of the genes located in their genomic neighborhood.
In each of the studied species, we found heat stress associated SNPs notably located in proximity of genes coding for well-established actors of the cellular responses against heat. Among these, we can mention proteins involved in DNA damage-repair, protein folding, oxidative stress homeostasis, inflammatory and apoptotic pathways. In some cases, the same putative molecular targets of heat stress adaptation recurred among species.
Together, these results underscore the relevance and the power of the seascape genomics approach for the discovery of adaptive traits that could allow corals to persist across wider thermal ranges.

4.2. Introduction

One of the most dramatic consequences of climate change is the worldwide decline of coral reefs, which are the most biodiverse ecosystems in the marine environment (Hughes et al., 2017). Among the main drivers of this decline is coral bleaching, a stress response to anomalous heat waves that eventually causes the death of hard corals (Bellwood et al., 2004; Hughes et al., 2017). In the most severe episodes, coral bleaching provoked local living coral cover loss of up to 50% (Hughes et al., 2017; Hughes, Kerry, et al., 2018), with climate change projections expecting for bleaching conditions to be persistent worldwide by 2050 (Van Hooidonk et al., 2013).
Despite these catastrophic perspectives, a glimpse of hope is brought by coral reefs that show resistance after recurrent heat waves (Dance, 2019; Hughes et al., 2019; Krueger et al., 2017; Penin et al., 2013; Thompson & van Woesik, 2009). One of the mechanisms that might promote heat tolerance in corals is genetic adaptation (Sully, Burkepile, Donovan, Hodgson, & van Woesik, 2019). In recent years, there has been a growing body of literature investigating how coral thermal adaptation might alter the predictions of reef persistence, and how conservation policies could be modified accordingly (Logan et al., 2014; Matz et al., 2018; van Oppen et al., 2015).

Given the crucial role adaptation will play in long-term reef persistence, there is an urgent need to characterize the adaptive potential of corals (Logan et al., 2014; van Oppen et al., 2015). For instance, there are still open questions concerning the spatial and temporal scales at which local adaptation operates (Matz et al., 2018; Roche, Williams, & Turner, 2018). Changes in adaptive potential have been observed along thermal gradients over hundreds of kilometres (*e.g.* (L. Thomas et al., 2017), but also at reefs with distinct thermal variations located only a few hundreds of meters apart (*e.g.* Bay & Palumbi 2014). Furthermore, different coral species are reported to show differential vulnerability against thermal stress, leading to the question of how different life-history traits (*e.g.* reproductive strategies) drive the pace of adaptation (Darling et al., 2012; Hughes, Kerry, et al., 2018; Loya et al., 2001).

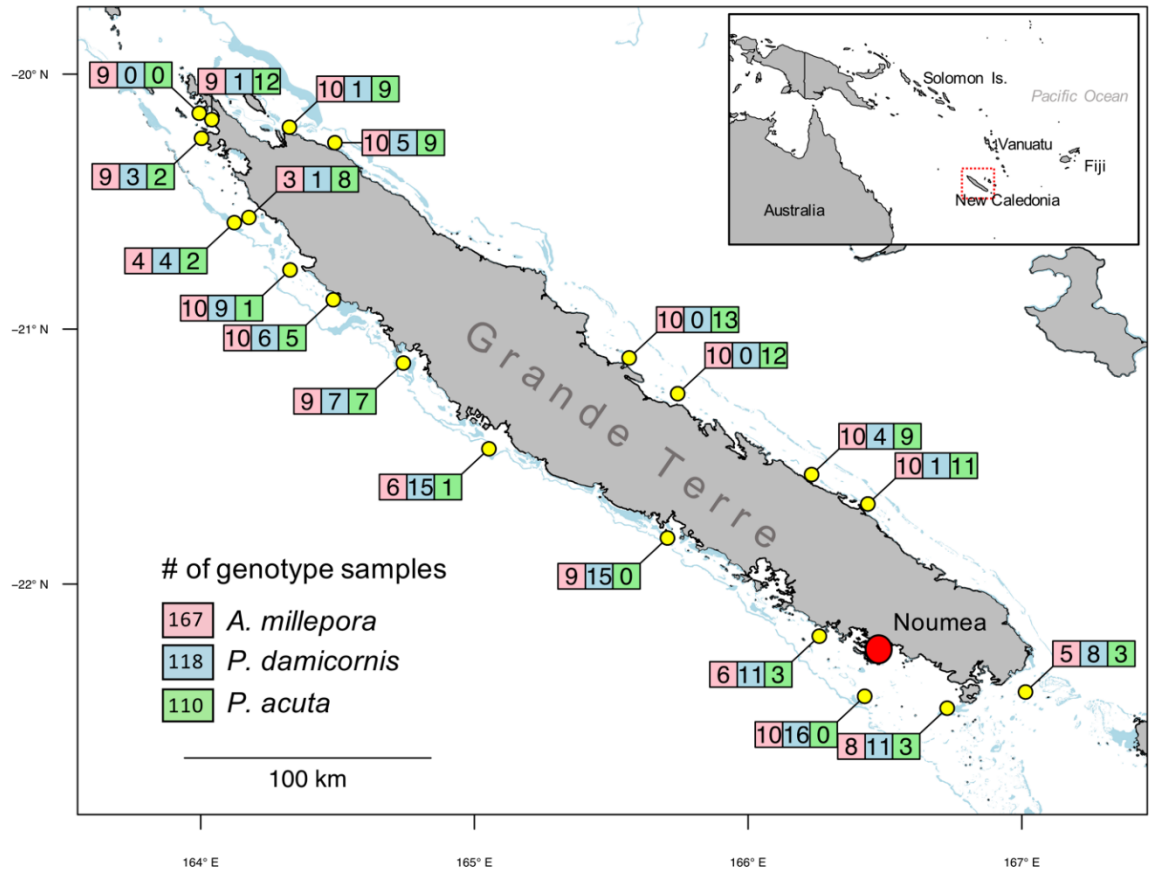
There are also open questions concerning the molecular mechanisms that might be targeted by heat stress adaptation in corals (Mydlarz et al., 2010; Palumbi et al., 2014; van Oppen & Lough, 2009). Some cellular responses to heat stress are now well characterized, such as DNA repair mechanisms, the activation of the protein folding machinery in the endoplasmic reticulum (ER) or the accumulation of reactive oxygen species (ROS, either endogenous or produced by the symbiont) that progressively elicits inflammatory and apoptotic responses (Maor-Landaw & Levy, 2016; Mydlarz et al., 2010; Oakley et al., 2017; Patel et al., 2018; van Oppen & Lough, 2009). However, little is known about which of the many molecular actors participating to these cascades could be hijacked by evolutionary processes to increment thermal tolerance.

Seascape genomics could contribute to filling these gaps. Seascape genomics is a budding field of population genomics that allows the study of local adaptation in wild populations (Riginos et al., 2016). This method combines the environmental characterization of the seascape with a genomic analysis of its population (Rellstab et al., 2015). The goal is to identify genetic variants that correlate with environmental gradients that might underpin an adaptive role (Rellstab et al., 2015). Seascape genomics could enhance the characterization of coral adaptive potential because: (1) it requires an extensive sampling strategy that allows for studying adaptation at different geographic scales, and against different types of environmental constraints simultaneously (*e.g.* mean temperatures, standard deviations, accumulated heat stress; Leempoel et al., 2017; Selmoni, Vajana, et al., 2020); (2) its experimental protocol is less laborious in comparison to traditional approaches used for studying coral adaptation (*e.g.* aquarium experiments, transplantations), and therefore facilitates scaling-up to a multiple species analysis; (3) it is based on genomic data and thus reports candidate molecular targets of adaptation (Rellstab et al., 2015; Riginos et al., 2016). Moreover, recent work described how the results of seascape genomics studies on corals can be directly transposed to a conservation perspective and support reef prioritization (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, & Joost, 2020).

Here we applied the seascape genomics approach to study the adaptive potential against heat stress in three bleaching-prone coral species of New Caledonia, in the southwestern Pacific (Fig. 4-1). We first used publicly available satellite data to characterize the seascape

conditions for over 1,000 km of the reef system. A sampling campaign was then organized to collect colonies at the 20 sites exposed to the most contrasted environmental conditions. The collected samples underwent a genotype-by-sequencing (DArT-seq) genomic characterization, followed by a seascape genomics analysis accounting for the confounding role of demographic structure. This allowed us to uncover single nucleotide polymorphisms (SNPs) associated with heat stress. We then performed the functional annotations of genes surrounding these SNPs and found molecular targets that notably recurred among species and that referred to well established heat stress responses in coral cells. Our study lays the foundations for the discovery of adaptive traits that could allow corals to persist across wider thermal ranges.

Figure 4-1. Study area and sampling sites. The 20 sampling sites around Grande Terre, the main island of New Caledonia (South Western Pacific), are shown in yellow. For every sampling site, the number of genotyped individuals per species (*Acropora millepora*: red, *Pocillopora damicornis*: blue, *Pocillopora acuta*: green) are given in the corresponding boxes.



4.3. Material and Methods

4.3.1. Environmental data

The seascape genomics approach requires an exhaustive description of the environmental conditions in order to prevent the misleading effect of collinear gradients (Leempoel et al., 2017; Riginos et al., 2016). For this reason, the seascape characterization we used encompassed seven environmental variables: sea water temperature (SST), chlorophyll concentration, sea surface salinity, sea current velocity, suspended particulate matter, alkalinity and bleaching alert frequencies (BAF; Tab. S4-1). The environmental characterization was performed in the R environment (R Core Team, 2016) using the *raster* package (Hijmans, 2016) and following the method described in previous work on coral seascape genomics (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020) with some modifications outlined hereafter.

For the description of SST we used two different georeferenced datasets covering the extent of New Caledonia: (1) daily records of SST since 1981 at a spatial resolution of 5 km ($SST_{5km;}$ EU Copernicus Marine Service 2017); (2) daily records of SST since 2002 at resolution of 1 km (SST_{1km}; Group for High Resolution Sea Surface Temperature; Chao et al. 2009; Chin et al. 2017). The first dataset covers a wider temporal range, therefore providing a more reliable characterization of historical trends. The second dataset covers a smaller temporal window, but the higher geographic resolution allows to portray fine scale thermal patterns with a higher degree of confidence. Both datasets were used to compute, for each pixel of the study area, averages and standard deviations of the warmest month, the coldest month, and the entire observational period. Furthermore, both datasets were used to compute three indices of bleaching alert frequencies (BAF), representing the frequency of days (over the whole period of remote sensing) during which the bi-weekly accumulated heat stress (*i.e*. SST above the average maximum) exceeded 0°C ($BAF_{0°C}$), 4°C ($BAF_{4°C}$) and 8°C ($BAF_{8°C}$) (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). Similarly, we computed the frequency of the bleaching warning conditions as defined by the Coral Reef Watch (BAF_{CRW}), corresponding to the accumulation of heat over the three previous months (Liu et al., 2003).

For the other datasets (chlorophyll concentration, sea surface salinity, sea current velocity and suspended particulate matter; EU Copernicus Marine Service 2017), the spatial resolution ranged between 4 and 9 km (Tab. S4-1). All the datasets covered a temporal extent of at least 20 years before 2018 (the year of sampling) and were processed to compute: (1) highest monthly average, (2) lowest monthly average and (3) overall average. For all the datasets captured at daily resolution (*i.e.* all except suspended particulate matter), we also computed the standard deviation associated with the three means. Seawater alkalinity was estimated by combining SST_{5km} and salinity in a polynomial equation as described by Lee and colleagues (Lee et al., 2006).

In total, 47 environmental descriptors (Tab. S4-1) were computed and assigned to the shapes of the reefs of New Caledonia (UNEP-WCMC et al., 2010), reported for a regular grid (~3,000 cells of size: 2x2 km) using QGIS (QGIS development team, 2009).

4.3.2. Sampling

Twenty sampling sites were selected out of the ~3,000 reef cells surrounding Grande Terre, the main island of New Caledonia (Fig. 4-1). Sampling sites were chosen following an approach that simultaneously maximized environmental contrasts and replicated them at distant sites (the "hybrid approach" described in Selmoni, Vajana, et al., 2020). The method consists of applying Principal Component Analysis (PCA) and hierarchical clustering to the 47 environmental descriptors in order to separate the ~3,000 reef cells into distinct environmental regions. Next, the algorithm selects the same number of sampling sites within each region in order to maximize physical distance between sites. Increasing environmental variation is expected to raise the sensitivity of seascape genomics analysis, while the replication of environmental gradients is expected to reduce false discovery rates (Selmoni, Vajana, Guillaume, Rochat, et al., 2020). Here the number of environmental clusters was five (Fig. S4-1) and we established at four sampling locations per cluster. When this was not possible (*e.g.* because of logistic constraints during the sampling campaign), additional sampling sites were added to the neighbouring clusters.

The sampling campaign was performed from February to May 2018 (under the permits N°609011-/2018/DEPART/JJC and N°783-2018/ARR/DENV) and targeted three flagship species of the Indo-Pacific: *Acropora millepora, Pocillopora damicornis sensu* Schmidt-Roach et al. (2013) [corresponding to PSH04 *sensu* Gélin et al. (2017)] and *Pocillopora acuta sensu* Schmidt-Roach et al. (2013) [corresponding to PSH05 *sensu* Gélin et al. (2017)]. Of note, *P. acuta* and *P. damicornis* belong to the complex of species formerly named *P. damicornis* (Gélin, Fauvelot, Bigot, Baly, & Magalon, 2018; Johnston et al., 2017; Schmidt-Roach, Miller, Lundgren, & Andreakis, 2014). At every sampling site, we collected up to 20 samples of *A. millepora* and 20 of *Pocillopora* aff. *damicornis* (we did not discriminate between species while sampling as it can be difficult to distinguish them in the field). All the samples were collected in a 1 km area and at a depth ranging between 2-4 m. The centre of this area was used for georeferencing the sampling site. Before sampling, each colony was imaged underwater, then a portion of a branch was sampled with hammer and chisel. Each sample consisted of a 1-2 cm branch that was immediately transferred to 80% ethanol and stored at -20°C. DNA from the 730 samples (370 *A. millepora* and 360 *Pocillopora;* Tab. S4-2) were extracted using the DNeasy 96 Tissue kit (Qiagen) following manufacturer instructions.

4.3.3. Species identification

The 360 *Pocillopora* samples were identified molecularly *a posteriori* of sampling to be assigned to one species or the other (*P. damicornis or P. acuta*). Samples were thus genotyped using 13 microsatellite loci, as in Gélin et al. (2017; Online Resource 1). Then, colonies belonging to *P. damicornis* and to *P. acuta* were identified using assignment tests performed with STRUCTURE (v. 2.3.4 ;Pritchard et al. 2000), as in Gélin et al. (2018). Colonies assigned to *P. damicornis* or *acuta* with a probability of at least 0.70 were retained in the final dataset for this study. The *Pocillopora* sampling was composed of 148 *P. damicornis* (more precisely to SSH04b *sensu* Gélin et al. 2017), 159 *P. acuta* colonies (more precisely, a mix of SSH05a and SSH05b *sensu* Gélin, Pirog, et al. 2018) and 53 unassigned colonies (excluded from further analysis; Tab. S4-2).
Acropora species are genetically and morphologically notoriously challenging in terms of identification and species boundaries detection. However, *A. millepora* can be recognised in the field based on its typical axial and radial corallite shape (Wallace, 1999).Back from the field, *in situ* images of each specimen were examined to look for the species diagnostic morphological characters and initial identifications validated.

4.3.4. Screening and SNP genotyping

All DNA samples from *A. millepora*, *P. damicornis* and *P. acuta* were sent to the Diversity Arrays Technology (Canberra, Australia) for quality check screening and genotype-by-sequencing using the DArT-sequencing method (DArT-seq; Kilian et al. 2012). The restriction enzymes used for library preparation for *A. millepora* and *Pocillopora* samples were PstI and HpaII. Prior to sequencing, all the DNA samples underwent a one-hour incubation with the digestion buffer, followed by a step of quality check for integrity, purity and concentration running 1 µL per sample on a 0.8% agarose gel. Samples from each site were then ranked based on their quality (degree of smearing on the agarose gel). We then selected the samples with the best scores from each site and defined a list of 188 *A. millepora,* 128 *P. damicornis* and 150 *P. acuta* samples that proceeded to the sequencing step in four and five lanes of an Illumina Hiseq2500, respectively. During each step of the workflow (DNA purification, library preparation and sequencing), *A. millepora* and *Pocillopora* samples were kept separated and randomly distributed across the respective batches (*e.g.* 96-well plates, sequencing lanes) to minimize the risks of technical bias. SNPs were called using the DArT-seq analytical pipeline (DArTsoft14).

4.3.5. SNP filtering

The DArT-loci (*i.e.* the DNA sequences surrounding each SNP) initially underwent a sequence similarity search (BLASTn; v. 2.7.1; Madden & Coulouris 2008) against a reference genome to

retain only those associated with the coral host. For *A. millepora,* the reference genome was the *A. millepora* chromosome-level assembly from Fuller and colleagues (v. 2; Fuller et al. 2019, unpublished data, available on arXiv), while for *P. damicornis* and *P. acuta* we used the only *Pocillopora* reference assembled to date (*P. damicornis sensu lato;* v. 1; Cunning et al. 2018). Only DArT-loci scoring an E-value below 10^{-6} were retained.

The processing of the SNPs data followed a pipeline from previous work on coral seascape genomics (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). For each species' dataset, we removed SNPs and individuals with high missing rates (> 50%) by using custom functions in the R environment. Next, we proceeded with imputation of missing genotypes using the *linkimpute* algorithm (based on k-nearest-neighbours imputation; Money et al. 2015) implemented in Tassel 5 (Bradbury et al., 2007) using the default settings. Afterwards, we repeated the filtering of SNPs and individuals for missing rates, but this time using a more stringent threshold (5%). We also applied a filter to exclude rare alleles (minor allele frequency < 5%) and highly frequent genotypes (major genotype frequency > 95%). SNPs were then filtered for linkage disequilibrium using the R package *SNPrelate* (function *snpsgdsLDpruning*, LDthreshold=0.3, v.1.16; Zheng et al. 2012). Finally, we applied a filter for clonality: when groups of colonies shared highly correlated genotypes (Pearson correlation > 0.9) only one colony pert group was kept.

4.3.6. Neutral genetic structure analysis

Prior to the seascape genomics analysis, the neutral genetic structure of the studied populations was investigated by running a PCA on the genotype matrix of each species using the R *stats* package (*prcomp* function). Firstly, we visually inspected the percentage of variance of the genotype matrix explained by each principal component (eigenanalysis); in highly structured populations the first principal components (PCs) are expected to explain a larger proportion of the variance, when compared to the subsequent PCs (Johnstone, 2001; Novembre et al., 2008). We ran a Tracy-Widom test as implemented in R package *AssocTests* (L. Wang et al., 2017) to determine the number of PCs underlying a non-random genetic structure ($P < 0.05$; Patterson et al. 2006).

We also visualized the spatial distribution of the main axis of variation (PC1), in order to investigate the presence of geographical structures (Novembre et al., 2008). Finally, we focused on the SNP-specific loadings on PC1 and their distributions across the genome. In fact, groups of genetically isolated individuals (*e.g.* hybrids, cryptic species) are expected to display genomic islands of low-recombination (*i.e.* groups of physically close SNPs contributing to high loading on the main axis of variation; Nosil et al. 2009; Li & Ralph 2019). We therefore visualized the distribution of average PC1-loadings by genomic windows of 50 and 100 kb. In these calculations, only genomic windows containing at least 5 SNPs were retained.

4.3.7. Seascape genomics

The seascape genomics analyses were performed separately on the three species using the LFMM method implemented in the *LEA* R package (v. 2.4.0; Frichot et al. 2013; Frichot & François 2015). This method associates single environmental gradients and individual SNPs variations in mixed models, where the confounding effect of neutral genetic variation is accounted for through latent factors (Frichot et al., 2013).

Briefly, the first step of the LFMM pipeline is to estimate the number of latent factors (K; Frichot & François 2015). This parameter corresponds to the number of ancestral populations and can be estimated by using the *snmf* function of the LEA package. The method processes a genotype matrix to estimate individual admixture coefficients under different K's, and then evaluates the quality of fit for each K via cross validation (Frichot & François, 2015). We ran ten replicates of this analysis for all the studied species, and found that the optimal number of K (according to the lowest entropy criterion) ranged from 2-4 for *A. millepora,* 6-8 for *P. damicornis,* and 10-12 for *P. acuta* (Fig. S4-2).

We then proceeded to the genotype-environment association analysis with LFMM. Since this method does not accept missing genotypes, we first ran the *impute* function of the LEA package. For each species, this function inferred the missing genotypes out of the ancestral genotype frequencies previously calculated with the *snmf* function. Finally, we ran the association analysis between the SNPs of each species and the environmental condition descriptors. This was done by using the *lfmm* function, setting K to the ranges previously estimated for each species and running five replicates of each analysis. Since *lfmm* calculations can be computationally intensive, when two or more environmental descriptors were highly collinear (absolute value of Pearson correlation > 0.9), only one was used in the analysis.

LFMM returns *P*-values describing the statistical significance of every genotype-environment association under different values of K. For each association model related to the same environmental variable, *P*-values were corrected for multiple testing using the q-value method (R package *q-value*, v. 2.14, Storey 2003) and deemed significant if $q < 0.01$ under at least one level of K. When a SNP was significantly associated to multiple environmental variables, or with an environmental variable highly correlated with others (*i.e.* those excluded from the LFMM calculations), we defined a main environmental descriptor as the variable most strongly correlated (Pearson) with the SNP.

4.3.8. Annotation analysis of heat stress associated SNPs

For each of the three studied species, we annotated the genomic neighbourhood of every SNP in order to characterize the possible functional implications of a genetic variant. Firstly, we uniformed the annotations of genes from different species in order to facilitate comparisons. We did this by retrieving the positions of genes in the two reference genomes

(Cunning et al., 2018; Fuller et al., 2019) and the corresponding predicted protein sequences. These sequences were used to perform a similarity search (blastP; Madden & Coulouris 2008) against the Uniprot/swissprot database (*metazoa* entries, release 2020_01; Boeckmann et al. 2003). Each gene was annotated with the best significant hit (E-value < 0.01) and inherited protein name and gene ontology (GO) terms describing molecular functions when existing (Ashburner et al., 2000).

Afterwards, we focused on the annotation of genes surrounding SNPs associated with heat stress descriptors. Significant SNPs were deemed "heat stress associated" if they were best correlated with an environmental descriptor relating to average temperature, standard deviation of temperature, or bleaching alert frequency. We mapped every SNP associated with heat stress as the genes located within a ± 50 kb window. We selected this window size because genes associated with a SNP can be located hundreds of kilobases away (Brodie et al., 2016; Visel et al., 2009) , with 50 kb being roughly the median contig size in the *P. damicornis* reference genome (Cunning et al., 2018).

We then computed the observed occurrence of each GO term among the genomic neighbourhoods of significant SNPs. As a comparison, we split the reference genome into 100 kb windows and computed the expected occurrence for each term. This procedure was performed separately for each of the three studied species, using the respective reference genome. The statistical analysis of differences between the expected and observed GO term occurrences (enrichment analysis) was performed using the Fisher exact test method as implemented in the R *topGO* package (v. 2.34; Alexa et al. 2006). As suggested in the *topGO* guidelines, we ranked GO terms according to the *P-value* of the Fisher test and discarded those occurring fewer than 10 times throughout the genome.

4.4. Results

Three coral species were sampled at 20 sites across the reef system of New Caledonia: *Acropora millepora* (Ehrenberg, 1834; n=360), *Pocillopora damicornis* (Linnaeus, 1758; (n=128) and *Pocillopora acuta* (Lamarck, 1816 *;* n=150; Tab. S4-2). The DArT-seq analytical pipeline resulted in the genotyping of 188 samples by 57,374 bi-allelic single nucleotide polymorphisms (SNPs) for *A. millepora*, and 128 and 150 samples by 70,640 SNPs for *P. damicornis* and *P. acuta*, respectively (Tab. 4-1). After filtering for rare variants, missing values and clonality, we obtained a final genotype matrix of 167 individuals by 11,935 SNPs for *A. millepora*, of 118 individuals by 7,895 SNPs for *P. damicornis* and of 110 individuals by 8,343 SNPs for *P. acuta* (Tab. 4-1). The *A. millepora* genotyped samples distributed across all the 20 sampling sites (18 of which counted five samples or more), while genotyped samples of *P. damicornis* and *P. acuta* were distributed across 17 sites each (both with 10 sites counting five samples or more), with 15 sites where both species were found in sympatry (Fig. 4-1; Tab. S4-2).

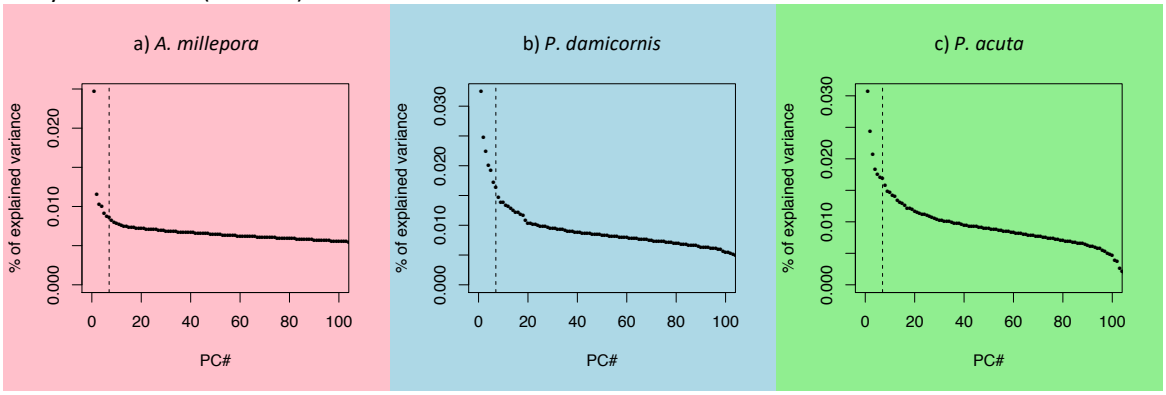
Table 4-1. Workflow of the analysis. For each of the species of interest (Acropora millepora, Pocillopora damicornis, Pocillopora acuta), we report the number of individuals (ind.) and single nucleotide polymorphism (SNPs) obtained or retained after each the various step of the workflow.

	A. millepora	*Pocillopora*	
		P. damicornis	*P. acuta*
Sampling	370 ind.	360 ind.	
Microsatellite	-	128 ind.	150 ind.
DArT-seq	188 ind. x 57,374 SNPs	128 ind. x 70,640 SNPs	150 ind. x 70,640 SNPs
BLAST against reference	188 ind. x 47,529 SNPs	127 ind. x 48,049 SNPs	145 ind. x 48,049 SNPs
Filtering (Missing values, MAF, MGF, LD, clonality)	167 ind. x 11,935 SNPs	118 ind. x 7,895 SNPs	110 ind. x 8,343 SNPs

4.4.1. Neutral genetic structure

We ran a principal component analysis (PCA) of the genotype matrix of each species to anticipate possible confounding role of neutral genomic variation on the adaptation study (Fig. 4-2, Fig. S4-3, Fig. S4-4). The Tracy-Widom test (P<0.05) revealed that the number of PCs underlying a non-random structure were seven for all the species, accounting for 8% of the total variance in *A. millepora*, 15% in both *P. damicornis* and *P. acuta* (Fig. 4-2). In *P. damicornis,* the spatial distribution of PC1 values appeared to be spatially structured following a north-south separation along the west coast of Grande Terre (Fig. S4-3B). In *P. acuta,* colonies in the north-west of Grande Terre displayed lower PC1 values, compared to those on the eastern coast (Fig. S4-3C). In *A. millepora,* no clear geographical patterns emerged as individuals with different values on PC1 were often located on the same reef. Finally, we analysed the presence of genomic widows clustering SNPs with high PC1-loadings,

Figure 4-2. Principal component analysis (PCA) of the genotype matrices for three species studied, *Acropora millepora*, *Pocillopora damicornis* and *Pocillopora acuta*. The three graphs display the percentage of variance explained by the first 100 principal components (PC) of the genotype matrix for the three studied species. The vertical dotted lines represent the number of PCs deemed as underlying a non-random structure by the Tracy-Widom test (P < 0.05).



expected to be frequent in genetically isolated groups (cryptic species, hybrids). These genomics windows were rare, as we observed no more than three per species (Fig. S4-4).

4.4.2. Local adaptation

We investigated the presence of SNPs that were associated with 47 environmental gradients describing the seascape conditions in New Caledonia (Fig. 4-3). When a SNP was found associated with multiple environmental descriptors, only the most significant association was kept. In total, 120 significant (q<0.01) genotype-environment associations were found for *A. millepora*, 90 for *P. damicornis* and 100 for *P. acuta* (Tab. 4-2a; Tab. S4-3, Tab. S4-4). In all of the three species, we investigate the environmental descriptors that most frequently associated with significant SNPs were those related to sea surface temperature (SST; 63 in *A. millepora,* 47 in *P. damicornis*, 43 in *P. acuta*). Among these, we found that putative adaptive signals related to bleaching alert frequencies (73 genotype-environment

Figure 4-3. Example of significant genotype-environment association. The map displays the superposition between environmental gradient (here highest monthly SST average) and the distribution of an associated (q<0.01) SNP of *Acropora millepora*. Every circle corresponds to the SNP genotype for an individual colony. For illustrative reasons, genotypes are radially distributed around the sampling locations. The boxplot in the top-right corner shows how the environmental variable distributes within each genotype. The SNP represented here is located on the contig xpSc0000535 (position 118526) of *A. millepora* genome, and the closest annotated gene codes for ATP-dependent DNA helicase Q5.

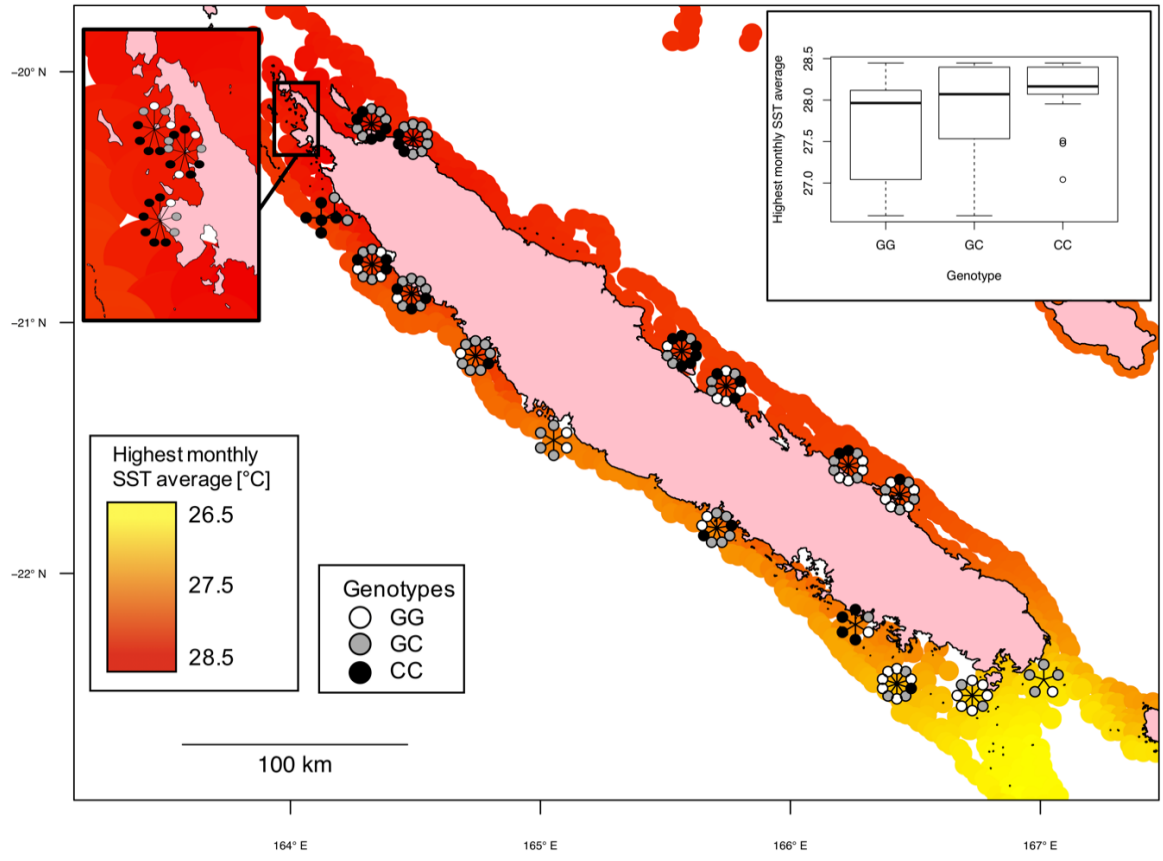


Table 4-2. Significant genotype-environment associations in *Acropora millepora* (*AM*), *Pocillopora damicornis* (*PD*) and *Pocillopora acuta* (*PA*). Table a displays the number of SNPs significantly associated (q<0.01) with environmental descriptors; Table b, the complete list of environmental descriptors related to sea surface temperature (averages and standard deviations at two different spatial resolutions and indices of bleaching alert frequencies - BAF). Note that when a SNP was significantly associate to multiple environmental descriptors, only the best association was kept. The detailed list of the SNP-environment associations is available in the supplementary Table 4.

a) All environmental descriptors

Environmental descriptor	*AM*	*PD*	*PA*
Sea surface temperature	63	47	43
Alkalinity	11	7	14
Chlorophyll concentration	15	22	27
Sea current velocity	4	4	2
Suspended particulate matter	8	1	2
Salinity	19	9	12
Total	**120**	**90**	**100**

b) *Sea surface temperature only*

Environmental descriptor			*AM*	*PD*	*PA*
Sea surface temperature	5 km	Overall average	1	0	0
		Average warmest month	1	1	1
		Average coldest month	1	1	0
		Overall standard deviation	1	2	1
		Standard deviation hottest month	0	0	0
		Standard deviation coldest month	11	1	7
		$BAF_{0°C}$	2	1	6
		$BAF_{4°C}$	3	9	1
		$BAF_{8°C}$	1	1	0
		BAF_{CRW}	2	3	0
	1 km	Overall average	0	1	1
		Average warmest month	5	2	4
		Average coldest month	12	4	3
		Overall standard deviation	4	3	2
		Standard deviation hottest month	2	1	0
		Standard deviation coldest month	2	4	1
		$BAF_{0°C}$	10	8	8
		$BAF_{4°C}$	3	3	3
		$BAF_{8°C}$	2	1	1
		BAF_{CRW}	0	1	4
Total			**63**	**47**	**43**

associations) were more frequent than those relating to the standard deviation (42) and average temperatures (38; Tab. 4-2b).

When we focused on the environmental descriptors not relating to temperature, we observed that those describing chlorophyll concentration were those associated to more SNPs (64 across the three species), followed by salinity (40; Tab. 4-2a). In contrast, current velocity variables were the best environmental descriptors associated with fewer genetic variants (10 across the three species, Tab. 4-2a).

4.4.3. Functional annotations of heat stress associated SNPs

Genes around SNPs that were associated with heat stress were annotated with gene ontology (GO) terms to investigate the molecular functions potentially altered by a genetic variant. GO terms were ranked for over-representation according to the Fisher exact test *P*-value (Tab. 4-3, S4-5). Among the top 50 ranks we found GO terms describing molecular functions such as "mismatched DNA binding", "heat shock protein binding", "chaperone binding", "unfolded protein binding", "cytoskeletal protein binding" and "actin binding" in *A. millepora* (Tab. 4-3a, S4-5a); "actin filament binding receptor activity", "exonuclease activity", "endonuclease activity" and "death effector domain binding" in *P. damicornis* (Tab. 4-3b, S4-5b); "NAD binding", "nucleotide binding" and "mitogen-activated protein kinase binding" in *P. acuta* (Tab. 4-3c, S4-5c). Four terms ("signaling receptor binding", "receptor regulator

activity", "organic cation transmembrane transporter activity", "enzyme activator activity") recurred among top ranked GO terms in at least two different species (Tab. 4-3). Each of the

Table 4-3. **Functional annotations of heat stress associated SNPs.** For each of the studied species, *Acropora millepora*, *Pocillopora damicornis* and *Pocillopora acuta,* the tables display the list of GO terms describing molecular functions that are overrepresented in the genomic neighborhoods (± 50 kb) of heat stress associated SNPs. For each GO term, the tables display the id (GO.ID), the term description (Term), the occurrence in the genome split in 100 kb windows (#Ann.), the observed (#Obs.) and expected (#Exp.) occurrence in the neighborhoods of heat stress associated SNPs and the *P*-value associated to the Fisher exact test comparing the expected and observed occurrences. For every species, the tables show a subset of the top 50 GO terms. The complete list of top ranked GO terms is available in the supplementary material (Tab. S4-5). GO terms in bold are those appearing in the top GO list of two different species.

a) *A. millepora*						
Rank	GO.ID	Term	#Ann.	#Obs.	#Exp.	*P*-value
3	GO:0003779	actin binding	278	11	4.25	<0.01
10	GO:0016670	oxidoreductase activity, acting on a sulfur group of donors, oxygen as acceptor	10	2	0.15	0.01
12	GO:0098772	molecular function regulator	724	19	11.07	0.01
15	GO:0008047	**enzyme activator activity**	244	9	3.73	0.01
16	GO:0030547	receptor inhibitor activity	11	2	0.17	0.01
25	GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	38	3	0.58	0.02
26	GO:0019209	kinase activator activity	39	3	0.6	0.02
32	GO:0030983	mismatched DNA binding	16	2	0.24	0.02
33	GO:0043028	cysteine-type endopeptidase regulator activity involved in apoptotic process	16	2	0.24	0.02
34	GO:0008092	cytoskeletal protein binding	644	16	9.85	0.03
39	GO:0015101	**organic cation transmembrane transporter activity**	19	2	0.29	0.03
41	GO:0051082	unfolded protein binding	83	4	1.27	0.04
43	GO:0060090	molecular adaptor activity	167	6	2.55	0.04
44	GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	22	2	0.34	0.04
47	GO:0031072	heat shock protein binding	88	4	1.35	0.04
50	GO:0030545	**receptor regulator activity**	131	5	2	0.05
b) *P. damicornis*						
2	GO:0005506	iron ion binding	126	6	1.78	0.01
7	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	167	6	2.36	0.03
8	GO:0051015	actin filament binding	127	5	1.79	0.03
11	GO:0048487	beta-tubulin binding	24	2	0.34	0.04
15	GO:0008047	**enzyme activator activity**	251	7	3.55	0.06
17	GO:0004888	transmembrane signaling receptor activity	1045	20	14.76	0.06
19	GO:0008528	G protein-coupled peptide receptor activity	324	8	4.58	0.08
20	GO:0005102	**signaling receptor binding**	749	15	10.58	0.08
30	GO:0005096	GTPase activator activity	139	4	1.96	0.13
40	GO:0038023	signaling receptor activity	1142	20	16.13	0.14
42	GO:0016796	exonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters	47	2	0.66	0.14
46	GO:0016894	endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 3'-phosphomonoesters	11	1	0.16	0.15
47	GO:0030169	low-density lipoprotein particle binding	11	1	0.16	0.15
49	GO:0035877	death effector domain binding	11	1	0.16	0.15
c) *P. acuta*						
5	GO:0022857	transmembrane transporter activity	702	18	10.17	0.01
9	GO:0015368	calcium:cation antiporter activity	11	2	0.16	0.01
10	GO:0005102	**signaling receptor binding**	749	18	10.85	0.01
21	GO:0015101	**organic cation transmembrane transporter activity**	16	2	0.23	0.02
35	GO:0000166	nucleotide binding	1297	25	18.79	0.03
41	GO:0031435	mitogen-activated protein kinase binding	23	2	0.33	0.04
42	GO:0030545	**receptor regulator activity**	137	5	1.98	0.05
45	GO:0051287	NAD binding	57	3	0.83	0.05
47	GO:0016651	oxidoreductase activity, acting on NAD(P)H	59	3	0.85	0.05
49	GO:0090722	receptor-receptor interaction	26	2	0.38	0.05

three species displayed top ranked GO terms referring to “oxidoreductase activity” acting on different molecules (Tab. 4-3).

4.5. Discussion

4.5.1. Different types of heat stress adaptation

In each of the studied species, we detected genotype-environment associations that might underpin local adaptation (Tab. 4-2a). Approximately half of these associations concerned descriptors of sea surface temperature (SST; Tab. 4-2a), partially because these are the most numerous types of descriptors employed in the analysis (20 out of 47).
As we focused on SST-related associations, we found that those involving bleaching alert frequencies (BAFs) were more frequent (73) than those related to temperature variations (42) and temperature averages (38; Tab. 4-2). Coral bleaching is a major threat for coral survival, and bleaching conditions emerge when SST variation exceeds seasonal averages (Hughes et al., 2017; Liu et al., 2003). BAFs descriptors account precisely for this selective constraint (SST variation over average), and this might explain why genotype-environment associations with BAFs were more frequent. Previous work on coral seascape genomics also reported a predominance of adaptive signals related to BAF (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). Coral adaptation appeared to be also driven by SST averages (regardless of variations) or by SST variations (regardless of the averages; Tab. 4-2b). This kind of adaptation might relate to bleaching (*e.g.* being adapted to high thermal variability promotes bleaching resistance; Safaie et al. 2018), or to other types of heat stress responses (*e.g.* impaired injury recovery at elevated average SST; Bonesso et al. 2017).

4.5.2. Candidate molecular targets for heat stress adaptation

Previous research reported that reefs exposed to high frequency of daily thermal variability showed reduced bleaching prevalence (Safaie et al., 2018). One of the reasons might be that corals at these sites manage to rapidly readjust their cellular homeostasis (Ruiz-Jones & Palumbi, 2017). This view is supported by the numerous GO terms describing activity regulators (for instance “signaling receptor binding”, “receptor regulator activity”, “enzyme activator activity”, “molecular function regulator”; Tab. 4-3, S4-5) found surrounding heat stress associated SNPs. The fact that these terms are not heat stress specific, however, invites to a cautious interpretation.
In contrast, we also detected several genes coding for well-established molecular actors of corals thermal stress responses in the neighborhood of heat stress associated SNPs (Tab. 4-4, S4-4). In some rare cases, we found that SNPs fell directly in the coding sequence of genes, but more frequently the SNPs were located several kb distance from genes (Tab. 4-4).

However, this does not exclude causative effects, as (1) the SNP detected could be physically linked to an adaptive SNP in the coding sequence; (2) the adaptive SNP could be located several kb from the target gene, as it is often the case (Brodie et al., 2016). Hereunder, we highlight the different molecular functions and the related proteins that were found as potential targets for thermal adaptation in corals.

- *DNA repair.* Heat stress impacts the integrity of nucleic acids and elicits mechanisms that promote DNA damage-repair and RNA stability (Henry, Yancey, & Kushner, 1992; Sottile & Nadin, 2018). A previous seascape genomics study on *Acropora digitifera* found five SNPs associated with heat stress to be proximal to genes coding for Helicase Q (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). Here we found Helicase Q5 in the genomic neighbourhood of a SNP associated with heat stress in *A. millepora* (Tab. 4-4). Helicases Q are required for efficient DNA repair during the initiation of the replication fork (Sharma et al., 2006). Another family of helicases participating to this process are Helicases MCMs (Daniel, Dagdanova, & Johnson, 2013) and one was found next to an heat stress associated SNP in *A. millepora* (Fig. 4-3; Tab. 4-4).
In addition, we found proteins involved in DNA damage-repair that are known to be differentially expressed in corals under heat stress. For instance, claspin (Palumbi et al., 2014; Smits, Cabrera, Freire, & Gillespie, 2019) and RAD51/54 homologs (Maor-Landaw & Levy, 2016) were found here surrounding heat stress associated SNPs in *A. millepora*, while DNA damage-binding protein 1 (J. Li et al., 2006) and DNA polymerase delta catalytic subunit (Prindle & Loeb, 2012) in *P. damicornis* (Tab. 4-4). Of note, GO terms describing molecular functions related to DNA repair ("mismatched DNA binding", "exonuclease activity", "endonuclease activity", "nucleotide binding") were found as over-represented in genes surrounding heat stress associated SNPs in the three species (Tab. 4-3).

- *Protein folding.* One of the main groups of gene annotations surrounding heat stress associated SNPs concerned molecular chaperones (Tab. 4-4). These are proteins that intervene in cellular responses to heat stress, where they assist the folding or unfolding of proteins in the endoplasmic reticulum notably (ER; Oakley et al. 2017). In corals, the role of these proteins in heat response, as well as their up-regulation under thermal stress, have been reported in several studies (Desalvo et al., 2010, 2008; Ishikawa et al., 2009; Maor-Landaw & Levy, 2016; Oakley et al., 2017; Rosic et al., 2011; Ruiz-Jones & Palumbi, 2017; van Oppen & Lough, 2009). The annotation analysis for the related GO terms (*e.g.* "heat shock protein binding", "unfolded protein binding"), revealed that this function was over-represented in genes close to the heat stress associated SNPs of *A. millepora* (Tab. 4-3a). In the three species, the genomic neighborhood of heat stress associated SNP contained several classes of chaperones: DnaJ homologs (four in *A. millepora,* one in *P. damicornis*), Tubulin-specific chaperone A and NudC domain-containing protein (*A. millepora*; Zheng et al. 2011), prolyl 3-hydroxylase 1 (*P. damicornis*; Ishikawa et al. 2009) and selenoprotein-F (*P. acuta;* Ren

et al. 2018). Another important class of chaperones are "Protein disulfide-isomerase", which catalyze the formation or breakage of disulfide bonds in proteins and produce reactive oxygen species (ROS) as byproducts (Oakley et al., 2017; van Oppen & Lough, 2009). For example, the disulfide-isomerase gene expression has been shown to be upregulated in *Stylophora pistillata* under experimental heat stress (Maor-Landaw et al., 2014). Disulfide-isomerase 2 genes were found next to heat stress associated SNPs in each of the studied species (Tab. 4-4).

- ***Oxidative stress response.*** In parallel to the protein folding and recycling response, coral cells under heat stress accumulate ROS (Nielsen et al., 2018; Oakley et al., 2017). This accumulation can derive from the leakage of ROS from the damaged photosynthetic machinery of the endosymbiont, as well as from the endogenous production of the host mitochondria elicited under heat stress (Nielsen et al., 2018; Oakley et al., 2017). ROS accumulation causes oxidative stress, and the GO terms describing the inherent responses ("oxidoreductase activity") were found as over-represented in genes next to heat stress associated SNPs in the three studied species (Tab. 4-3). For instance, we found genes coding for Peroxidasin homolog and Isocitrate dehydrogenase subunit beta next to heat stress associated SNPs in *A. millepora* (Tab. 4-4). Peroxidasin homolog is in the first line of defence against ROS accumulation, displays high rates of evolution in *A. millepora* and was found highly up-regulated under heat stress in *Monstastraea faveolata* (= *Orbicella faveolata*; Voolstra et al. 2009; Voolstra et al. 2011; Louis et al. 2017). Isocitrate dehydrogenase is one of the few sources of NADPH in the animal cell and was found upregulated in *E. pallida* under heat stress (Kültz, 2005; Oakley et al., 2017). NADPH is an essential substrate to contrast ROS accumulation (Oakley et al., 2017; Patel et al., 2018) and another gene implicated in its metabolism, Quinone oxidoreductase PIG3, was found close to heat stress associated SNPs in *P. acuta* (Tab. 4-4; Zangar et al. 2004). In *P. damicornis,* we found the Glutathione peroxidase 5 gene, belonging to a family of well-characterized antioxidants contrasting ROS accumulation in corals (Nielsen et al., 2018).
In the host mitochondria, ROS production occurs in a series of redox reactions across the inner membrane (the electron transport chain; Lutz et al. 2015). One of the main components of this chain is the respiratory complex I (NADH-ubiquinone oxidoreductase), and we found genes coding for two of its subunits in the genomic neighborhood of heat stress associated SNPs in both *Pocillopora* species (Tab. 4-4). The mechanism leading to ROS leakage from host mitochondria into coral cell cytoplasm is poorly known (Dunn, Pernice, Green, Hoegh-Guldberg, & Dove, 2012; Nielsen et al., 2018; Oakley et al., 2017). However, it is noteworthy to mention that in *A. millepora* the SNP most strongly associated with heat stress was close to the MIC60 gene (Tab. 4-4, S4-4a). MIC60 is a subunit of the MICOS complex, a key protein in the maintenance of the mitochondrial inner membrane architecture, through which ROS are produced, and the outer membrane, through which ROS diffuse into cytoplasm (Muñoz-Gómez, Slamovits, Dacks, & Wideman, 2015; Zhao, Jiang, Zhang, & Yu, 2019).

- *Inflammatory response and apoptosis.* The effects of ROS depends on the level of accumulation: medium levels elicit an inflammatory response, while excessive levels lead to cell apoptosis (Patel et al., 2018). Mitogen-activated protein kinases (MAPK) are key actors in the inflammatory response (Courtial et al., 2017; Patel et al., 2018; Son, Kim, Chung, & Pae, 2013). In corals, MAPKs were shown to repress ROS accumulation in *S. pistillata* (Courtial et al., 2017) and were found in proximity of a SNP associated to heat stress in another seascape genomics study on Japanese *A. digitifera* (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). Here we found a MAPK coding gene around heat stress associated SNPs in *A. millepora* (MAPK1), and genes coding for a MAPK activating protein *P. damicornis* (Putative MAPK-activating protein FM08) and *P. acuta* (TNF receptor-associated factor 6; Tab. 4-4; Mason et al. 2004).
ROS excess eventually results in cell apoptosis (Patel et al., 2018). In each of the three species, we found apoptosis-related genes near heat stress associated SNPs: Programmed cell death protein 6 and Death-associated protein kinase 2 (*A. millepora);* Death effector domain-containing protein (*P. damicornis*); and apoptosis regulator Bcl-2 and Death-associated protein 4 (*P. acuta*; Tab. 4-4). These proteins participate in the caspase-mediated apoptotic cascade involved in the coral bleaching process (Ahmad et al., 1997; Dunn, Schnitzler, & Weis, 2007; Oakley et al., 2017; Tchernov et al., 2011; Valmiki & Ramos, 2009; Yuasa et al., 2015).

- *Cell structure*. Heat stress has been shown to lead to cytoskeleton reorganization (C. Wilson, Terman, González-Billault, & Ahmed, 2016). In Cnidaria, cytoskeletal proteins displayed changes in abundance under experimental heat stress in *Exaiptasia pallida* and *A. palmata* (Oakley et al., 2017; Ricaurte et al., 2016). Here we found several genes implicated in the cytoskeletal architecture in proximity of SNPs associated to heat stress: myosin III (twice), unconventional myosin VIIb and actin (*A. millepora*), unconventional myosin-Id (*P. damicornis*), Myosin heavy chain and actin-1 (*P. acuta*; Tab. 4-4). Moreover, the GO terms "actin binding" and "cytoskeletal protein binding" were over-represented in the set of genes neighbouring heat stress associated SNPs in *A. millepora*, and "actin filament binding" in *P. damicornis* (Tab. 4-3a-b). Of note, in three of the myosin genes the heat stress associate SNPs were found inside the coding sequence (Tab. 4-4, S4-4).

4.5.3. Limitations and future directions

Seascape genomics studies are exploratory analyses that come with the drawback of being subjected to high false discovery rates (Rellstab et al., 2015; Riginos et al., 2016). This bias is stressed when the confounding role of neutral genetic variation is not accounted for (Selmoni, Vajana, Guillaume, Rochat, et al., 2020). The preliminary analysis of population structure, however, did not reveal any cryptic speciation nor isolated reefs among the studied populations (Fig. 4-2, S4-3, S4-4). Furthermore, we used a statistical method (LFMM) and a

Table 4-4. Candidate molecular targets for coral adaptation to heat stress. Annotations of genes surrounding (± 50 kb) heat stress associated SNPs were sorted by molecular function. For six specific types of molecular functions, the table displays the genes potentially involved in heat stress response (Putative molecular target) and the position of the corresponding SNP associated with heat stress (SNP position, in format contig/chromosome: base position). The CDS tag indicates SNP falling inside the coding sequence of the putative molecular target. The background colours correspond to the species where the candidate molecular target was found (pink: *Acropora millepora,* blue: *Pocillopora damicornis,* green: *Pocillopora acuta*). The role that each molecular function has in coral heat response is described in the last column.

Molecular function	SNP position	Putative molecular target	Role in coral heat response
DNA damage repair	xpSc0000535:118526	ATP-dependent DNA helicase Q5	Heat stress and UV radiation can provoke DNA damage. Proteins in this list participate to the DNA damage-repair mechanism during replication.
	chr7:20210292	DNA helicase MCM9	
	chr13:7351849	DNA repair protein RAD51 homolog 2	
	chr13:20841309	Claspin	
	NW_020844825.1:56001	DNA polymerase delta catalytic subunit	
	NW_020847490.1:185714	DNA damage-binding protein 1	
Protein folding / chaperone / heat stress responses	chr7:2529319	DnaJ homolog subfamily B member 6	Molecular chaperones are activated in early response to thermal stress to assist protein folding in the endoplasmic reticulum.
	Sc0000122:685084	Sacsin (DnaJ homolog subfamily C member 29)	
	chr3:18355053		
	chr5:20803650	DnaJ homolog subfamily A member 3	
	chr5:10318559	NudC domain-containing protein 2	
	chr13: 18019180	Tubulin-specific chaperone A	
	chr12:16446416	Protein disulfide-isomerase 2	
	NW_020844635.1:120545	prolyl 3-hydroxylase 1	
	NW_020844967.1:239041	DnaJ homolog subfamily C member 9	
	NW_020845264.1:254233	Protein disulfide-isomerase 2	
	NW_020846699.1:47695	Protein disulfide-isomerase 2	
	NW_020847700.1: 75539	selenoprotein-F	
Oxidative stress	chr2:15023513	MIC60	Heat stress leads to the accumulation of ROS. All the proteins in this list participates to the metabolism of ROS. MIC60 is a key structural protein of the inner membrane of mitochondria were endogenous ROS is produced.
	xfSc0000142:60614	Peroxidasin homolog	
	chr7:20210292	Isocitrate dehydrogenase [NAD] subunit beta	
	NW_020843829.1:322689	NADH-ubiquinone oxidoreductase 23 kDa subunit	
	NW_020844825.1:56001	Glutathione peroxidase 5	
	NW_020846699.1:47695	NADH-ubiquinone oxidoreductase B16.6 subunit	
	NW_020845243.1:143874	Quinone oxidoreductase PIG3	
Inflammatory and apoptotic response	chr13:19271503	Mitogen-activated protein kinase-binding protein 1 (MAPK1)	MAPKs are activated in the inflammatory response caused by ROS accumulation. Excessive ROS accumulation elicits the caspases-mediated apoptotic response observed in coral bleaching.
	chr11:11687264	Death-associated protein kinase 2	
	xfSc0000077:45906	Programmed cell death protein 6	
	NW_020844212.1: 38069 (CDS)	Partial similarity with DED domain-containing protein	
	NW_020845264.1:254233	Putative MAPK-activating protein FM08	
	NW_020846901.1:188731	apoptosis regulator Bcl-2	
	NW_020846154.1:13424	Death-associated protein 4	
	NW_020846522.1:185468	TNF receptor-associated factor 6	
Cytoskeleton	chr1:19795515	Myosin IIIA	ROS accumulation degrade structural proteins and lead to a pronounced reorganization of the cytoskeleton.
	chr12:22835441 (CDS)	Unconventional myosin-VIIb	
	chr2:15023513 (CDS)	Partial similarity with Myosin III	
	chr3:17298197	Actin, cytoplasmic	
	NW_020844967.1:239041	Unconventional myosin-Id	
	NW_020847027.1:161971	Actin-1	
	NW_020846942.1:1165830 (CDS)	Myosin heavy chain	

sampling design allowing to mitigate such confounding effects (Frichot et al., 2013; Selmoni, Vajana, Guillaume, Rochat, et al., 2020).

There are, nevertheless, some points that could have increased the statistical power of the analysis, as for instance the use of a larger sample size (over 200 individuals per species; Selmoni, Vajana, et al., 2020). In addition, the assignment of heat stress associated SNPs to candidate molecular targets for adaptation would have been further facilitated with a higher genome resolution in the sequencing strategy. Higher genome resolution would also allow to infer the structural modification that SNPs falling inside the coding sequence might cause. In

the years to come, whole-genome-sequencing on corals is likely to become more affordable and can then be applied to the large sample sizes required for seascape genomics studies.
The next step in the characterization of corals' adaptive potential is experimental validation. Our work found several genetic variants that might confer selective advantages against thermal stress (Tab. 4-4). For each of the studied species, we can now define multiple-loci genotypes of heat stress resistant colonies and test their fitness under experimental heat stress conducted in aquaria (Krueger et al., 2017). As a result, this analysis will allow to 1) further investigate the role of different heat stress associated genotypes and molecular pathways and 2) provide a concrete measure of the thermal ranges that these coral populations might sustain in the years to come. This information is of paramount importance, as it will allow to predict the reefs that are expected to already carry heat tolerant colonies and to define conservation strategies accordingly (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). For instance, marine protected areas could be established to preserve reefs with higher adaptive potential against heat stress, where such reefs could provide the breeding stock to restore the damaged ones (Baums, 2008; van Oppen et al., 2017, 2015).

4.5.4. Conclusions

In this study, seascape genomics allowed to uncover genetic variants potentially implicated in adaptive processes against different types of heat stress in three coral species of New Caledonia. These variants were located next to genes coding for molecular actors that participate in well-understood cellular reactions against thermal stress. In addition, the approach pointed out new candidate genes (*e.g.* Helicase Q) or processes (*e.g.* signalling receptor binding) that might be implied in such responses. Of note, some of these potential targets for adaptation recurred in the analyses of different species, supporting the robustness and the power of the seascape genomics. Future studies will focus on performing experimental assays to validate the implication of potentially adaptive genotypes and newly identified genes in the heat stress response and to measure the thermal ranges tolerated by the diverse adaptive genotypes.

4.6. Acknowledgments

We thank Gerard Mou-Tham, Joseph Baly and Miguel Clarque, for their support during the field campaign, Annie Guillaume for the comments and suggestions provided during the redaction of this paper and Andrew Baird for helpful discussion on field identifications.
This work was supported by the United Nations Environment Programme (UNEP) and International Coral Reef Initiative (ICRI) coral reefs small grants programme (grant number: SSFA/18/MCE/005). We also thank the Government of France and the Government of the Principality of Monaco who provided the funding for the small grants.

5. Article D

Coral cover surveys corroborate predictions on reef adaptive potential to thermal stress

Oliver Selmoni[1,2], Gaël Lecellier[2,3], Laurent Vigliola[2], Véronique Berteaux-Lecellier[2] and Stéphane Joost[1]

[1]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

[2]UMR250/9220 ENTROPIE IRD-CNRS-UR, Labex CORAIL, Noumea, New Caledonia.

[3]UVSQ, Université de Paris-Saclay, Versailles, France.

Preprint version of the article submitted to *Scientific Reports:*
Selmoni, O., Lecellier, G., Vigliola, L., Berteaux-Lecellier, V. and Joost, S. (2020). Coral cover surveys corroborate predictions on reef adaptive potential to thermal stress. *Available on bioRxiv: https://doi.org/10.1101/2020.06.05.136523*

Contribution of the candidate:
As first author of this article, I contributed, in collaboration with the co-authors to initiate the research. I retrieved the data from field surveys, computed the conservation indices and ran the statistical analyses and wrote the first version of the manuscript. The co-authors provided advice concerning the methods used, interpretation of the results and critically revised the manuscript before submission.

Article D is the second study focusing on coral reefs of New Caledonia in the framework of the SABLE project. Unlike article C, which emphasized the molecular side of adaptation against heat stress, here the focus is brought to the implications of adaptation to heat stress and reef connectivity under a conservation perspective. To do so, we applied the prioritization indices developed in article B to the New Caledonian case study.
Methods estimating adaptation from remote sensing data are usually supported by field surveys observations (Maina et al., 2008; Matz et al., 2019)*. Field surveys provide information on coral bleaching occurrence, or more generally on changes in living coral cover (Bruno & Selig, 2007; Donner et al., 2017). In the Japanese case study, a correlation was observed between the predictions of adaptive potential against heat stress and the mortality rates recorded during the 2016 bleaching event (Tab. 3-2). However, this correlation might have been fortuitous, as it was based on a single year of observation. In this article we further*

investigated this correlation by employing publicly available data from a long-term survey campaign across the New Caledonian reef system (Job, 2018).
The main challenge in this work was the definition of a rigorous statistical framework to evaluate the association of living coral cover with heat-stress adaptation and connectivity. There are three reasons underpinning this difficulty. First, coral cover changes can be driven by numerous factors (e.g. climatic, anthropogenic), some of which are not measurable (Maina et al., 2011, 2008). Second, the field surveys we used are irregular between sites and years, with some sites being followed for up to 15 years, and others for less than five (Job, 2018). This irregularity obstructs the statistical inference in longitudinal data analysis (Garcia & Marder, 2017). Third, heat-stress adaptation in corals might imply substantial trade-offs (e.g. reduced growth rate; Kenkel, Almanza, & Matz, 2015). Consequently, heat-stress adaptation is not expected to have positive effects on coral cover persistently, but only in response to heat-stress exposure (Matz et al., 2019). We overcame these obstacles by using general linearized mixed models (GLMMs), which accounted for the hidden effects driving the baseline of coral cover at each survey site. In addition, remote sensing data was used to describe the degree of heat stress experienced at each site before the survey. Including this variable via interaction models allowed to contextualize the association of coral cover with adaptive potential in response to specific heat stress.
While predictions on adaptation to heat stress are corroborated by field observation, this work does not replace the need for further validation using experimental approaches. However, it reinforces two views proposed by this thesis. The first is that seascape genomics is confirmed to be a valuable tool to synthesize information to support the conservation of the adaptive potential of corals. The second is that local adaptation against heat stress will play a decisive role for the persistence of reef-building corals.

5.1. Abstract

As anomalous heat waves are causing the widespread decline of coral reefs worldwide, there is an urgent need to identify coral populations tolerant to thermal stress. Heat stress adaptive potential is the degree of tolerance expected from evolutionary processes and, for a given reef, depends on the arrival of propagules from reefs exposed to recurrent thermal stress. For this reason, assessing spatial patterns of thermal adaptation and reef connectivity is of paramount importance to inform conservation strategies.
In this work, we applied a seascape genomics framework to characterize the spatial patterns of thermal adaptation and connectivity for coral reefs of New Caledonia (Southern Pacific). In this approach, remote sensing of seascape conditions was combined with genomic data from three coral species. For every reef of the region, we computed a probability of heat stress adaptation, and two indices forecasting inbound and outbound connectivity. We then compared our indicators to field survey data, and observed that decrease of coral cover after

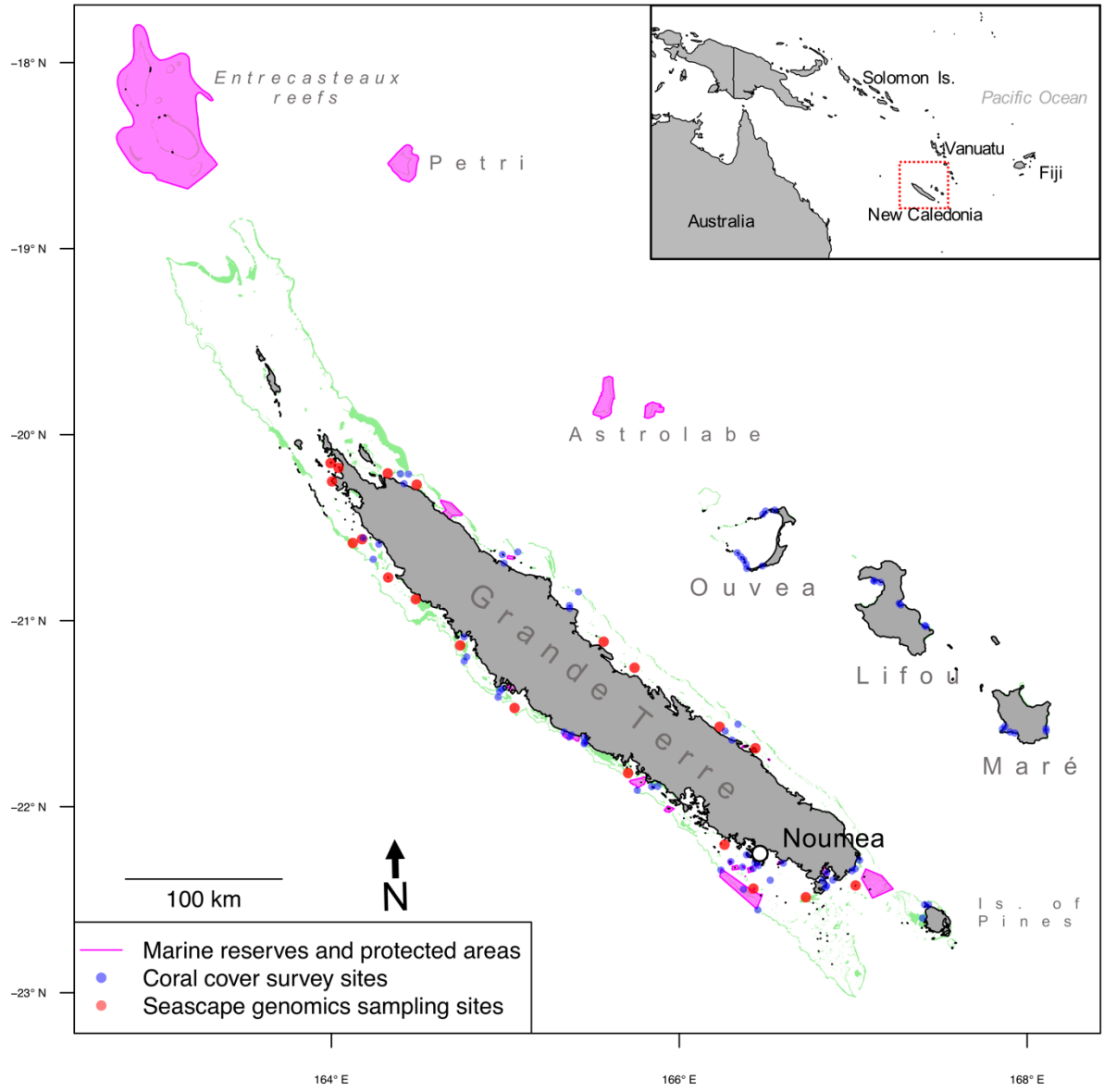
heat stress was lower at reefs predicted with high probability of adaptation and inbound connectivity. Last, we discussed how these indicators can be used to inform local conservation strategies and preserve the adaptive potential of New Caledonian reefs.

5.2. Introduction

Coral bleaching is one of the main causes of severe declines of coral reefs around the world (Bellwood et al., 2004; Hughes, Anderson, et al., 2018; Hughes et al., 2017). This phenomenon is mainly caused by anomalous heat waves leading to the death of hard-skeleton corals, which are the cornerstone of reefs (Bellwood et al., 2004). Over the last 30 years mass coral bleaching events repeatedly struck worldwide, causing losses of local coral cover up to 50% (Hughes, Anderson, et al., 2018; Hughes et al., 2017). In the coming years, bleaching conditions are expected to occur more frequently and to become persistent by 2050 (Van Hooidonk et al., 2013). As up to one third of marine wildlife depends on coral reef for survival and at least 500 million people livelihoods worldwide (Costanza et al., 2014), there is an urgent need to define new strategies to improve the preservation of these ecosystems (Moberg & Folke, 1999).

Recent research reported reefs that rebounded from repeated heat stress and showed an increased thermal resistance (Hughes et al., 2019; Krueger et al., 2017; Penin et al., 2013; Sully et al., 2019; Thompson & van Woesik, 2009). Adaptation of corals against heat stress might explain such observations (Bay & Palumbi, 2014; L. Thomas et al., 2018). Under this view, identifying adapted coral populations is of paramount importance, as conservation strategies might be established to protect reefs hosting these corals from local stressors (*e.g.* via marine protected areas, MPAs; Wilson, Tittensor, Worm, & Lotze, 2020). Furthermore, adapted corals could be of use in reef restoration plans and repopulate damaged reefs (Baums et al., 2019). The adaptive potential of corals at a given reef depends on the arrival of propagules from reefs exposed to recurrent thermal stress (Matz et al., 2019; Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). This is why characterizing spatial patterns of thermal adaptation and reef connectivity is crucial to empower the conservation of the adaptive potential of corals (Matz et al., 2019; Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). Seascape genomics is a powerful method to evaluate spatial patterns of environmental variation and connectivity (Riginos et al., 2016; Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). This method relies on a thorough analysis of environmental conditions around reefs using satellite data. Daily records of surface temperature are remotely sensed using satellites, and processed to compute indicators of thermal patterns associated with bleaching events (Liu et al., 2003; Maina et al., 2008; Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). Corals exposed to different thermal patterns are then sampled and genotyped to identify genetic variants correlated with these indicators (Riginos et al., 2016; Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). The

Figure 5-1. Reef system of New Caledonia. Coral reefs are highlighted in green. The blue dots correspond to sites of coral cover survey of the New Caledonian observational network of coral reef (Job, 2018). The red dots correspond to the sampling locations of coral specimen for the seascape genomics study that provide genetic data in the present study (Selmoni, Lecellier, Magalon, et al., 2020). Sea regions highlighted in purple correspond to the marine reserves and protected areas as catalogued by the French agency for MPAs (http://www.aires-marines.fr/).



association between genetic variants and a given indicator defines a model of adaptation that can be used to predict the probability of adaptation, based on the value of the indicator itself (Rochat & Joost, 2019; Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). In addition, by remote sensing sea current movements, it is possible to draw a connectivity map between every reef within an area of interest. This can be done using spatial graphs that resume multi-generational dispersal matching spatial patterns of genetic diversity in a given species (Boulanger, Dalongeville, Andrello, Mouillot, & Manel, 2020). This approach results in indices of connectivity defining, for a reef of interest, the predisposition in sending (outbound

connectivity) and receiving (inbound connectivity) propagules to/from neighboring reefs (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020).

In this study, we predicted spatial patterns of heat stress adaptation and connectivity for over 1000 km of coral reefs of New Caledonia, in the Southern Pacific (Fig. 5-1). The study area encompassed the barrier reef surrounding Grande Terre, the main islands of the Archipelago, as well as the intermediary and fringing enclosed in the lagoon. We also considered reefs surrounding the Loyalty Islands (Ouvéa, Lifou and Maré) and the Astrolabe (east of Grande Terre) and those in the Entrecasteaux and Petri atolls (north of Grande Terre). We first used remote sensing data to (1) evaluate the thermal variability of the study area and (2) estimate patterns of sea current connectivity between reefs. Next, we employed genomic data from a seascape genomics study on three coral species of the region (Selmoni, Lecellier, Magalon, et al., 2020) in order to (1) compute the probability of adaptation to heat stress across the whole region, and (2) verify whether predicted reef connectivity matched genetic correlation between corals. Last, we compared our predictions with field surveys of living coral cover recorded by the New Caledonian observational network of coral reef (RORC; Job, 2018). Our results suggest that negative effects of recent heat stress on coral cover are mitigated at reefs predicted with high probability of heat stress adaptation and inbound connectivity. We then discuss the conservation status of reefs around New Caledonia, and assess how conservation indices of probability of adaptation and connectivity can be used to protect the adaptive potential of corals of the region.

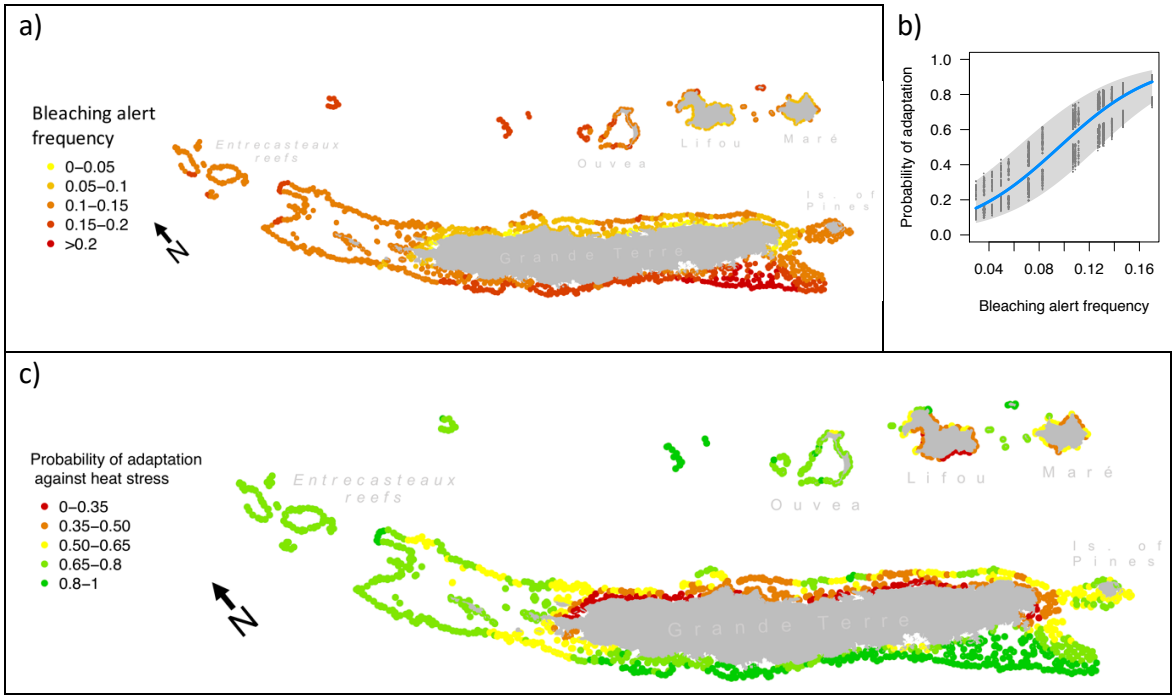
5.3. Results

5.3.1. Heat stress and probability of adaptation

The remote sensing data of sea surface temperature since 1985 were processed to calculate the frequency of bleaching alert conditions ($BAF_{overall}$) across the reef system of New Caledonia (Fig. 5-2a). $BAF_{overall}$ was higher in reefs on the western coast of Grande Terre (average BAF: 0.16±0.04) than in those on the eastern coast (0.08±0.03). Reefs in Lifou, Maré and Isle of Pines displayed $BAF_{overall}$ values comparable to those on the eastern coast of Grande Terre (0.09±0.03, 0.10±0.02 and 0.11±0.01, respectively), while in Ouvéa and Entrecasteaux reefs the $BAF_{overall}$ values (0.15±0.01 and 0.12±0.01, respectively) were closer to the values observed on the western coast.

Previous seascape genomics analyses on three corals of the region (*Acropora millepora, Pocillopora damicornis and Pocillopora acuta*) revealed the presence of multiple genetic variants (32 in total) potentially implicated in heat stress resistance (Selmoni, Lecellier, Magalon, et al., 2020). We employed these data to construct logistic genotype-environment association models defining the expected frequency of potentially adaptive genetic variants as a function of $BAF_{overall}$. We then used a "leave-one-population-out" cross-validation

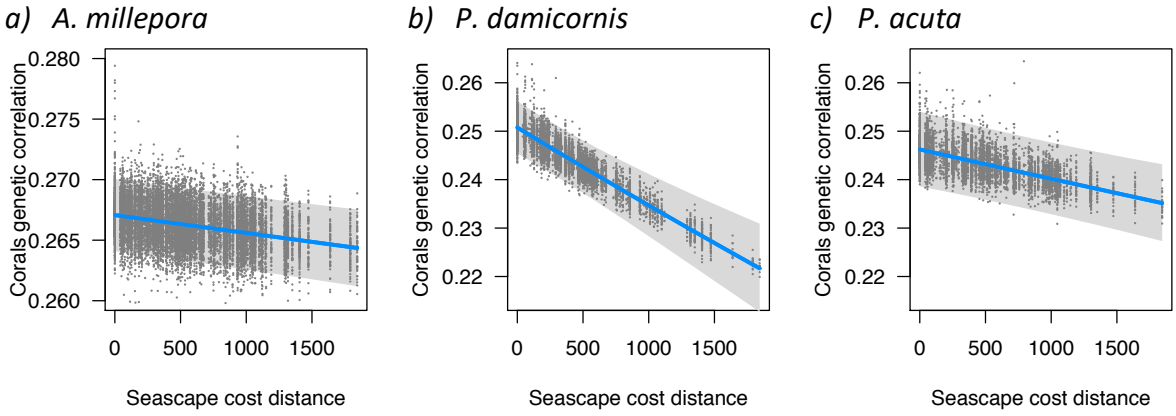
Figure 5-2. Bleaching alert frequency and probability of heat stress adaptation. In (a), bleaching alert frequency ($BAF_{overall}$) is displayed for each reef of New Caledonia. This value is derived from remote sensing data of sea surface temperature, and describes the frequency of cumulated heat stress conditions that can lead to bleaching. In (b), a logistic model of heat stress adaptation is shown. This model is based on the frequencies of potentially adaptive genotypes of three coral species of New Caledonia (Selmoni, Lecellier, Magalon, et al., 2020). The plot displays the probability of adaptation to heat stress as a logistic function of $BAF_{overall}$ (blue line, with the grey band showing the 95% interval of confidence). The model shown in (b) was used to translate $BAF_{overall}$ displayed in (a) in the probability of adaptation (PA_{HEAT}) against heat stress. The map in (c) displays PA_{HEAT} for every reef of New Caledonia.



method and found that, in all the three species, the expected frequencies of adaptive genotypes were correlated with the observed ones (*A. millepora:* r=0.52±0.09, *P. damicornis*: r= 0.55±0.16*, P. acuta*: r=0.6±0.08; Fig. S4-1). As a comparison, the same cross-validation method applied to 1000 randomly selected genetic variants resulted weak correlations between expected and observed frequencies (*A. millepora:* r=-0.2±0.32*, P. damicornis*: r=-0.07±0.36, *P. acuta*: r=-0.09±0.37).

We then constructed a unique model of heat stress adaptation combining all the genotype-environment association models across the three species and defining the overall probability of presence of potentially adaptive variants (PA_{HEAT}) as a function of $BAF_{overall}$ (Fig. 2b). This model was then used to produce a map of predicted PA_{HEAT} values for the whole region (Fig. 2c). It revealed accentuated differences compared with $BAF_{overall}$ patterns, with PA_{HEAT} generally above 0.65 in reefs on the western coast of Grande Terre, Isle of Pines, Entrecasteaux and Ouvéa. In contrast, values below 0.35 were observed at reefs located along the east coast of Grande Terre, in Lifou and Maré.

Figure 5-3. Seascape cost distance and genetic correlation between corals. The three plots display genetic correlations between pairs of corals sampled in New Caledonia as a function of the cost distance separating reefs where corals were sampled (blue line, with the grey band showing the 95% interval of confidence). Genetic correlations were computed as the correlation of single-nucleotide-polymorphisms, while seascape cost distance was predicted through seascape connectivity graphs. Each plot displays this association for a different species (a: *Acropora millepora*, b: *Pocillopora damicornis*, c: *Pocillopora acuta*).



5.3.2. Reef connectivity and genetic correlation between corals

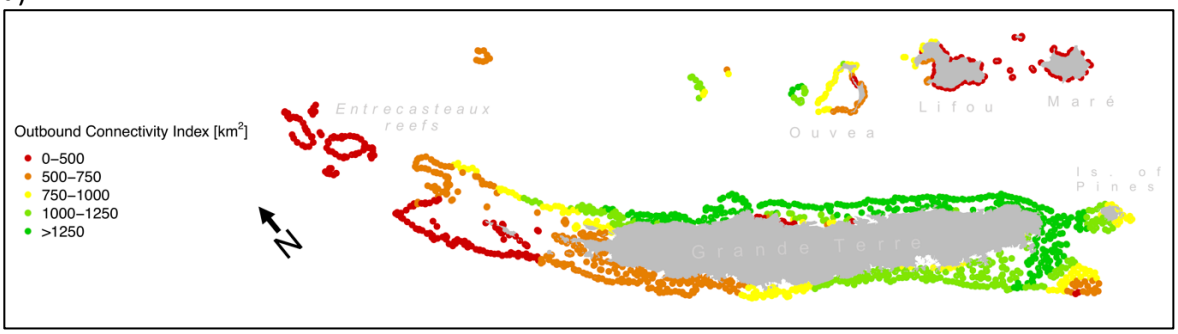
Remote sensing of sea current was used to compute a spatial graph of seascape connectivity predicting cost distances between reefs of New Caledonia. By using generalized linear mixed models (GLMMs) regression, we investigated whether such predictions on reef connectivity were representative proxies of the population structures of corals of the region. In three studied species (*A. millepora, P. damicornis and P. acuta*), we found that the genetic correlations between corals were significantly associated with the seascape cost distances separating the reefs where corals were sampled (*A. millepora*: p=1.46e-06; *P. damicornis*: p=1.63e-10 and *P.* acuta: p=6.7e-05; Fig. 5-3). This relationship was more stressed in the two *Pocillopora* species (regression coefficient for *P. damicornis*: β=-8.7E-05+ 1.4E-05; for *P. acuta*: β=-3.1E-05 ± 7.8E-06) than in *A. millepora* (β=-7.9E-06 ± 1.6E-06). The GLMMs accounted for the ancestral distance between pairs of individuals (*i.e.* the difference in admixture from ancestral populations) which was found significantly associated to genetic correlations between corals in all the three species (p~0; Fig. S5-2).

5.3.3. Reef connectivity indices

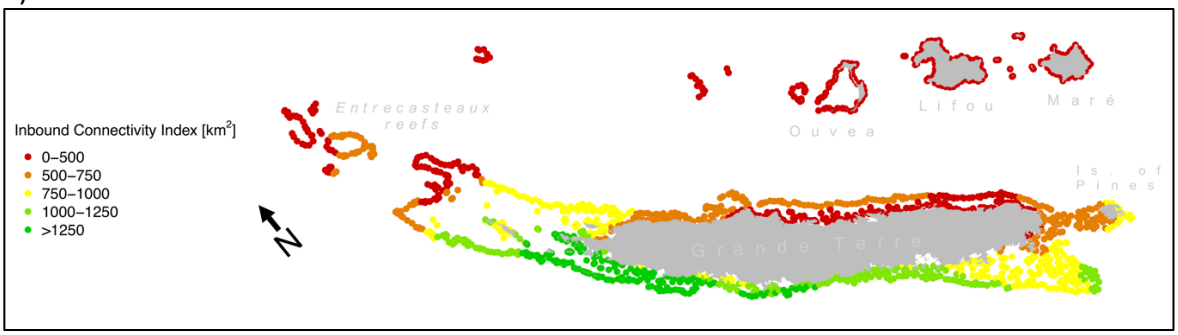
The seascape connectivity graph was used in the calculation of two indices describing the dispersal characteristics of every reef of New Caledonia (Outbound Connectivity Index, OCI, Fig. 5-4a; Inbound Connectivity Index, ICI, Fig. 5-4b). Both indices are expressed in km^2, as they represent the area of the reefs neighboring a reef of interest. In OCI, neighboring reefs

Figure 5-4. Connectivitiy indices. Two connectvitiy indices based on sea current data are shown for every reef of New Caledonia. In a), the Outbound Connectivity Index (OCI) describes the predisposition in sending dispersal to neighboring reefs. In b), the Inbound Connectivity Index (ICI) summarizes the predisposition in receiving propagules from neighboring reefs. Both indices are given in km^2, as this represents the total surface of neighboring reefs.

a)



b)



are those potentially receiving propagules from the reef of interest, while in ICI neighboring reefs are those potentially sending propagules towards the reef of interest.

Reefs that are more distant to Grande Terre (Entrecasteaux, Lifou, Maré and Ouvéa) had lower OCI (average OCI: 202±35 km^2, 410±270 km^2, 210±66 km^2, 864±254 km^2, respectively; Fig. 5-4a) than reefs surrounding Grande Terre. Reefs surrounding Grande Terre showed highest values on the southern reefs of the eastern coast (1929±300 km^2), while lower values were predicted for the rest of the eastern coast (1377±435 km^2) and the southern part of the western coast (1119±82 km^2). OCI was lower at reefs located at the northern extremity of Grande Terre (632±244 km^2).

Like with OCI, ICI was lower at reefs furthest from Grande Terre (Entrecasteaux, Ouvéa, Lifou, Maré; average ICI of 460±93 km^2, 177±7 km^2, 97±30 km^2, 111±6 km^2, respectively; Fig. 5-4b). ICI at reefs surrounding Grande Terre displayed a net contrast between the east and west coasts, where ICI was lower on the east (498±113 km^2) than the west (1287±407 km^2).

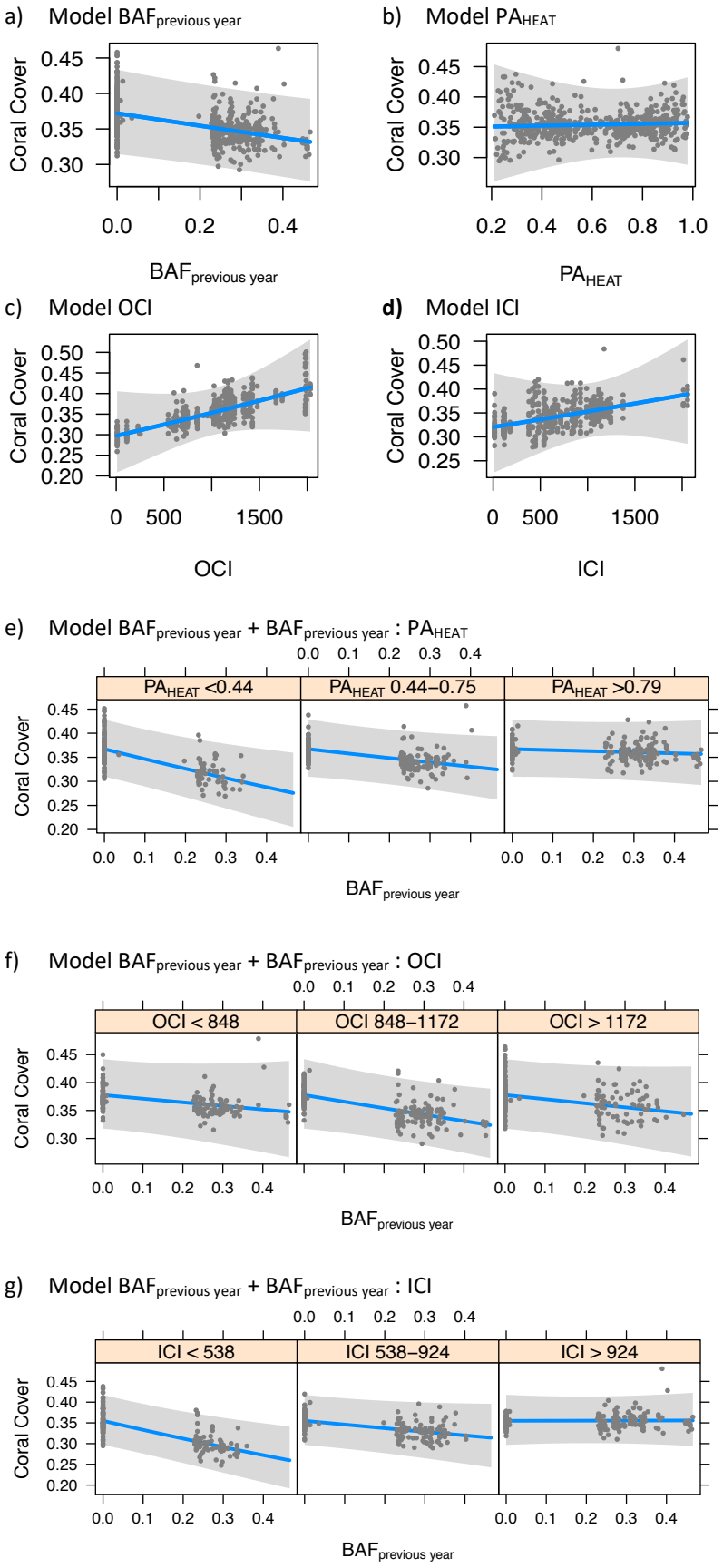


Figure 5-5. Coral cover association analysis. The plots display the association of coral cover rates (blue line, with the grey band showing the 95% interval of confidence) with recent thermal stress ($BAF_{previous\ year}$), probability of heat stress adaptation (PA_{HEAT}) and connectivity indices (inbound connectivity index, ICI, and outbound connectivity index, OCI). In plots (a) to (d), the association with coral cover rates is shown for each explanatory variable alone (a: $BAF_{previous\ year}$, b: PA_{HEAT}, c: OCI, d: ICI). In the remaining plots, the association between coral cover and $BAF_{previous\ year}$ and is showed across different ranges PA_{HEAT} (e), OCI (f) and ICI (g).

5.3.4. Coral cover analysis

Underwater surveys of New Caledonian reefs were analyzed to characterize the association of living coral cover with recent thermal stress ($BAF_{previous\ year}$), probability of heat stress adaptation (PA_{HEAT}) and connectivity indices (ICI and OCI; Fig. 5-5). We first investigated the association between coral cover and individual explanatory

variables using single fixed effect GLMMs (Fig. 5-5a-d). We found that coral cover was significantly associated with $BAF_{previous\ year}$ (p=0.02), and that this association was of negative sign (β=-0.06±0.03; Fig. 5-5a). In contrast, none of the other univariate models resulted in a significant association with coral cover (PA_{HEAT}: p=0.93, Fig. 5-5b; OCI: p=0.46, Fig. 5-5c; ICI: p=0.41, Fig. 5-5d). The Akaike Information Criterion (AIC) suggested a higher quality-of-fit for the model employing $BAF_{previous\ year}$ as explanatory variable (AIC=-883), compared with the other univariate models (PA_{HEAT}: AIC=-878; OCI: AIC=-879, ICI: AIC=-879).

We then investigated whether the negative association between coral cover and $BAF_{previous\ year}$ varied under different values of PA_{HEAT}, OCI or ICI. This analysis employed three bivariate GLMMs setting as fixed effects $BAF_{previous\ year}$ and the interaction between $BAF_{previous\ year}$ and each of the three other explanatory variables (PA_{HEAT}, OCI, ICI; Fig. 5-5e-g). In comparison to all the univariate models, those accounting for the interaction of $BAF_{previous\ year}$ with PA_{HEAT} and ICI resulted in a higher quality-of-fit (AIC=-886 and AIC=-888, respectively). In both cases, the effect of $BAF_{previous\ year}$ was significant ($p<0.01$) and of negative sign, whereas the effect of the interaction was also significant but of positive sign (for the interaction with PA_{HEAT}: β=+0.05±0.02, p=0.03; with ICI: β=+0.07±0.03, p=0.01; Fig. 5-5e-f). In contrast, the bivariate model incorporating OCI had a quality-of-fit comparable to univariate models (AIC=-883), and showed no significant association in interaction with $BAF_{previous\ year}$ (Fig. 5-5g).

5.4. Discussion

5.4.1. Local divergences in conservation indices

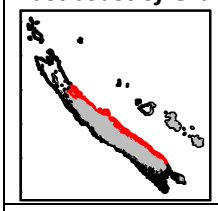
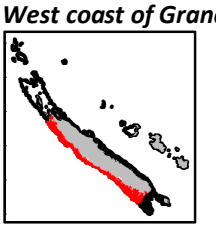
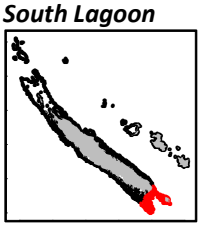
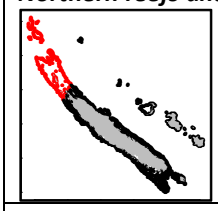
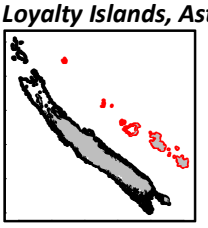
The metrics computed in this study stressed the strong asymmetry, in terms of both probability of heat stress adaptation (PA_{HEAT}) and connectivity (inbound connectivity index, ICI; outbound connectivity index; OCI), between reefs on the two coasts of Grande Terre (Fig. 5-2a, Fig. 5-4). The climatic differences between the two coasts are modulated by the mountain range covering Grande Terre, and water conditions inside the lagoon reflect the combination of these differences coupled with oceanic influences (Lefèvre, Marchesiello, Jourdain, Menkes, & Leroy, 2010). For example, the southern part of the west coast of Grande Terre is subjected to coastal upwelling, a seasonal phenomenon bringing cold water to the surface (Marchesiello, Lefèvre, Vega, Couvelard, & Menkes, 2010). While logic would suggest that cold water alleviates heat stress, research on the Great Barrier Reef in Australia showed that intense upwelling is followed by severe heat stress, and consequent coral bleaching (Berkelmans, Weeks, & Steinberga, 2010). While it is unknown whether this same effect occurs on the south-western coast of Grande Terre, this region does enclose the reefs that are predicted to experience the highest frequency of bleaching conditions across New Caledonia, and consequently to host corals with the highest PA_{HEAT} (Fig. 5-2).

Asymmetrical spatial patterns between the coasts of Grande Terre were also predicted for connectivity (Fig. 5-4), and this matched the genetic population structure of corals of the region (Fig. 5-3). In this work, we estimated connectivity using a straightforward approach, conceived to be reproduceable on any reef system around the world but that might lead to local inaccuracies(Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020). However, our predictions were generally consistent with previous work that characterized the regional water circulation around New Caledonia using more sophisticated methods (*i.e.* combining oceanographic models, in situ measurements and shipboard detectors of sea currents; Cravatte et al., 2015). For instance, we observed a higher inbound connectivity index (ICI) on the west coast of Grande Terre (Fig. 5-4b), and a higher outbound connectivity index (OCI) on the east coast (Fig. 5-4a). This west-oriented connectivity was expected because of the South Equatorial Current crossing the archipelago in this direction (Cravatte et al., 2015). This current bifurcates at the encounter of the New Caledonian shelf into 1) a weak and transient south-east oriented current between the Loyalty Islands and Grande Terre, and 2) a strong north-west oriented current flowing north of the Loyalty Islands (Cravatte et al., 2015; Hénin, Guillerm, & Chabert, 1984; Marchesiello et al., 2010). This bifurcation explains the lower OCI observed in Lifou and Maré, compared with Ouvéa and the Astrolabe atolls. Last, the water circulation inside the lagoon follows the north-west orientation of trade winds (Marchesiello et al., 2010), resulting in higher OCI in the south and higher ICI in the north.
Predictions of reef connectivity and PA_{HEAT} varied considerably across the different regions of the study area (Fig. 5-2, 5-4), and conservation planning should account for these regional peculiarities (Magris et al., 2014; K. L. Wilson et al., 2020). In table 5-1, we interpret the local divergences in values of PA_{HEAT}, ICI and OCI under a conservation perspective.

5.4.2. Predictions on adaptive potential match coral cover

Heat exposure is considered to be one of the main drivers of coral mortality worldwide (Hughes, Kerry, et al., 2018; Sully et al., 2019; Welle et al., 2017). Our results were consistent with this view, as we found a significant negative association of coral cover with $BAF_{previous\ year}$ (Fig. 5-5a). Adaptation might contribute to increase thermal tolerance in corals, but its potential depends on two elements: the existence of adapted corals and the presence of reef connectivity patterns facilitating their dispersal. In this study, we found both of these elements (PA_{HEAT} and ICI) as associated with reduced loss of coral cover after thermal stress.
Previous studies have reported reefs that display increased thermal tolerance after recurrent exposure to heat stress (Hughes et al., 2019; Krueger et al., 2017; Penin et al., 2013; Sully et al., 2019; Thompson & van Woesik, 2009), and recent research suggested that the thermal contrasts of New Caledonia might have driven adaptive processes in corals of the region (Selmoni, Lecellier, Magalon, et al., 2020). Our results supported this view: while recent thermal stress ($BAF_{previous\ year}$) was associated with a reduction in coral cover, this reduction

Table 5-1. Implications for reef conservation in New Caledonia. The table describes the implications for reef conservation of the probability of heat stress adaptation (PA_{HEAT}), the outbound and inbound connectivity indices (OCI, ICI) predicted for different regions of the New Caledonia reef system. Information on the existing marine protected areas were retrieved from the French agency for MPAs (http://www.aires-marines.fr/).

East coast of Grande Terre	
	The east coast of Grande Terre hosts reefs predicted with low ICI and PA_{HEAT}. In contrast, OCI was generally higher than in the rest of the Archipelago. Reefs of strategic importance might be those located in the southern part as they had the highest OCI of the Archipelago, and also moderate levels of PA_{HEAT}. To date, only 4 km^2 of reefs in this area are protected. In addition, the establishment of nurseries with heat stress adapted corals might increase the adaptive potential of these reefs.
West coast of Grande Terre	
	Reefs on the west coast of Grande Terre generally displayed higher levels of ICI and PA_{HEAT}, compared with the rest of the Archipelago. Under an adaptive potential perspective, reefs in the northern part are of paramount importance as they receive the propagules from all the south-western reefs that experienced frequent heat stress. No MPA is established in this area. Another strategic region are the reefs in front of Noumea, in the southern part of the west coast, since they were predicted with high PA_{HEAT} and OCI. Here, more than 200 km^2 of protected areas are already established.
South Lagoon	
	The South Lagoon displayed heterogenous patterns of PA_{HEAT} and connectivity. The highest PA_{HEAT} were observed in the south-western extremity, which in turn was a region predicted with low OCI. The eastern part might be more interesting under a conservation perspective, as it was predicted with moderate PA_{HEAT} and high OCI. These reefs are located upstream of the trade winds, and can simultaneously send propagules to both coasts of Grande Terre. A large marine reserve (180 km^2) is already established to protect these reefs. As for the southern part of the east coast, coral nurseries of heat stress adapted colonies might increase the adaptive potential of this region.
Northern reefs and Entrecasteaux reefs	
	Northern reefs and Entrecasteaux reefs were predicted with moderate to high levels of PA_{HEAT}, and low values of OCI and ICI, compared with the reefs around Grande Terre. The critical region under an adaptive potential perspective might be the eastern part of Northern reefs. This is because these reefs depend on the incoming propagules from the east coast of Grande Terre, which are predicted with low PA_{HEAT}.
Loyalty Islands, Astrolabe and Petri atolls	
	The main conservation issue for all the reefs in this region is the low ICI. It is likely that arrival of propagules substantially depends on the reefs from Vanuatu, located ~200 km upstream on the South Equatorial Current. Reefs in Ouvéa and Astrolabe atolls (already protected) might be of strategic importance, as they were predicted with moderate to high values of PA_{HEAT} and OCI. Since reefs in Maré and Lifou showed low PA_{HEAT}, establishment of nurseries with heat stress adapted coral might be useful under an adaptive potential perspective.

was mitigated at reefs that have experienced past thermal stress and were therefore predicted with high PA_{HEAT} (Fig. 5-5e).

In addition, PA_{HEAT} alone did not result in a significant association with coral cover rates (Fig. 5-5b), and this might be due to the fact that thermal adaptation is advantageous only in response to heat stress. Indeed, previous research reported trade-offs in traits involved in local adaptation and acclimatization to heat stress in corals (Kenkel et al., 2015). These trade-

offs might explain why the highest rates of coral cover (>0.4) in absence of heat stress ($BAF_{previous\ year}$=0) were mainly observed at reefs with low PA_{HEAT} (Fig. 5-5e).
Outbound connectivity was not found to be associated with changes in coral cover (Fig. 5-5c,f). This is not surprising, because beneficial effects of dispersal are expected at reefs receiving incoming propagules, rather than the opposite (Matz et al., 2019; Palumbi, 2003). Indeed, inbound connectivity was found to mitigate the negative association between $BAF_{previous\ year}$ and coral cover (Fig. 5-5g). Two non-mutually exclusive reasons might explain this observation. First, high levels of incoming propagules might facilitate the turnover of dead colonies caused by heat stress (Hock et al., 2017), although it has to be noted that this kind of recovery usually requires several years (Robinson et al., 2019). Second, incoming dispersal facilitates the arrival of adapted propagules, and therefore promotes an adaptive response even at reefs that did not experience thermal stress before (Kawecki, 2008). Indeed, we observed that the frequency of adaptive genotypes in *A. millepora* and *P. acuta* was generally higher at reefs predicted with low PA_{HEAT} and high ICI, than in those predicted with both low PA_{HEAT} and low ICI (Fig. S5-3). This view on genetic rescue via incoming migration is supported by the fact that every reef depends, to some extent, on its neighbors for larval recruitment (Treml et al., 2012).

5.4.3. Limitations and future directions

The associations found between changes in coral cover and the descriptors of thermal stress, probability of heat stress adaptation and connectivity do not necessarily imply causative relationships. Despite evidence of effects of thermal patterns on coral cover reported by previous studies, there might be other environmental constraints that are asymmetrical between the two coasts of Grande Terre and modulate coral cover changes. Further validation remains necessary and could be achieved via experimental assays of heat stress resistance (Krueger et al., 2017) in colonies sampled at reefs with different PA_{HEAT}. This approach would also enable disentangling of the possible confounding role of acclimatization in heat stress adaptive responses (Kenkel et al., 2015; L. Thomas et al., 2018).
Another important aspect to consider in future studies is the resolution of remote sensing datasets used for predictions. Here, we worked at a resolution of ~5 km for thermal variables and ~8.5 km for sea current data. While the overall environmental patterns appeared consistent with those characterized in previous studies, it is likely that small scale phenomena were neglected. For instance, reef heat stress exposure can vary substantially under the fine-scale (<1km) of a seascape (Bay & Palumbi, 2014). The same applies to connectivity, since the use of high resolution (≤1 km) hydrodynamic models could improve the characterization of coral larvae fine-scale dispersal (Colberg, Brassington, Sandery, Sakov, & Aijaz, 2020; Storlazzi, van Ormondt, Chen, & Elias, 2017).
A third limitation of our approach concerns the generalization of the biological and ecological characteristics of a reef. Here we assumed that the reef system of New Caledonia was a single

homogenous ecological niche, hosting an "average" species with an "average" heat stress adaptive response. This simplification is useful to portray an overall prediction, but might lead to local inaccuracies. This is because the reef types of New Caledonia are variegated and species distributions varies accordingly (Andréfouët, Cabioch, Flamand, & Pelletier, 2009; Dalleau et al., 2010)**.** Furthermore, different species have different levels of bleaching sensitivity (Loya et al., 2001) and reproduce under different strategies (Darling et al., 2012). For instance, the propagules of a broadcast spawning coral as *A. millepora* travel over longer distances, compared with those of brooding species as *P. damicornis* and *P. acuta* (Ayre & Hughes, 2000). Consequently, *A. millepora* showed a lower rate of decrease of genetic correlation between corals per unit of seascape cost distance, in comparison with the *Pocillopora* species (Fig. 5-3). Differences in the dispersal range can also modulate adaptive processes, since limited dispersal capabilities magnify the strength of natural selection(Kawecki & Ebert, 2004). The result are sharper gradients of adaptive genotype frequencies that in our study were not observed. Indeed, the accuracy in predicting the expected frequencies of adaptive genotypes did not significantly differ between species (Fig. S4-1), even though this observation might be biased by the unbalanced sample size between species (Selmoni, Vajana, Guillaume, Rochat, et al., 2020).
In future studies, PA_{HEAT} and connectivity predictions should be calibrated to match these biological differences. It is for this reason that seascape genomics studies will become of paramount importance into the future, as they provide species-specific indications on 1) how thermal stress might be translated in probability adaptation, and 2) the biological meaning (*e.g.* degree of genetic separation) of a cost distance by sea currents (Riginos et al., 2016; Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020).

5.4.4. Conclusions

In this study, we combined remote sensing of environmental conditions with genomic data to predict spatial patterns of heat stress adaptation and connectivity for the coral reefs of New Caledonia. We then retrieved field survey data and showed that recent heat stress was associated with a decrease in living coral cover, but also that such association appeared to be mitigated at reefs predicted with 1) high probability of heat stress adaptation and 2) high levels of incoming dispersal. The metrics computed in this work resumes the adaptive potential of corals against heat stress, and therefore represents valuable indices to support spatial planning of reef conservation.

5.5. Methods

5.5.1. Remote sensing of sea surface temperature

Satellite data characterizing sea surface temperature (SST) were retrieved from a publicly available database (dataset: ESA SST CCI reprocessed sea surface temperature analyses; EU Copernicus Marine Service, 2017; Merchant et al., 2019). This dataset provides daily records of SST at a ~5 km resolution from the years 1981 to 2017 across the whole study area (Fig. 5-1). The shapes of the reef of the region(UNEP-WCMC et al., 2010) were transformed into a regular grid (1,284 cells with maximal size of 5x5 km), and for each reef cell we extracted the average temperature for every day of the observational period using QGIS software(QGIS development team, 2009).

We performed calculations of heat stress patterns in the R environment using the *raster* package (v. 3.0; Hijmans, 2016; R Core Team, 2016). For each reef cell, patterns of heat stress were computed using the bleaching alert definition developed by the Coral Reef Watch briefly described hereafter (Liu et al., 2003). For every day, we calculated the "hotspot value" as the difference between SST and the maximal monthly mean (MMM, usually the monthly average of February in New Caledonia). The hotspot value was retained only when SST exceeded the MMM by at least 1 °C. Next, for each day, we calculated the cumulated hotspot values over the previous 84 days (3 months), and if this sum is > 0, the day is flagged as being 'under bleaching alert'. Finally, we computed the frequency of days under bleaching alert for every year (BAF_{year}) from 1985 to 2017. For the preceding years (1981-1984), BAF_{year} was not calculated such to avoid bias caused by estimating MMM over a limited number of years. An overall measure of BAF ($BAF_{overall}$) was calculated as the average of all the BAF_{year} from 1985 to 2017.

5.5.2. Seascape connectivity graph

For the estimation of connectivity we applied a method based on spatial graphs previously employed to study coral reef connectivity (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020) and briefly outlined hereafter. We retrieved a publicly available dataset describing the eastward and northward surface water velocity (Global Ocean Physics Reanalysis; EU Copernicus Marine Service, 2017). This dataset provided daily records at ~8.5 km resolution from 1993 to 2017. Since this resolution can be inaccurate close to coastlines, we increased the resolution to 1 km using the "resample" function ("bilinear" method) of the *raster* R package, and used high resolution bathymetry data (100 m resolution; Ryan et al., 2009)) to remove the sea velocity value from pixels located on land. We then used the R package *gdistance* (v. 1.2; van Etten, 2018) to create a matrix describing the transition costs between each adjacent pixel in the study area. These costs were inversely proportional to the frequency of transition based on sea currents. This seascape connectivity graph was

calculated as the shortest cost distances across this matrix between for each pair of the 1,284 reef cells. Of note, two least-cost-paths were calculated for each pair of reef cells, one for each direction of the transition.

5.5.3. Connectivity indices

The seascape connectivity graph was used to compute two indices connectivity for every reef cell of the study area: inbound connectivity and outbound connectivity. These indices had been defined in previous work on corals (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020) and were calculated in the R environment.

- *Outbound connectivity index (OCI)*: represents the predisposition of a reef to send coral propagules to its neighbors. For a given reef cell, it is calculated by defining all the neighboring reef cells that can be reached under a determined cost distance threshold (CDt). OCI is the total area (in km^2) of the destination reef cells.
- *Inbound connectivity index (ICI)*: represents the predisposition of a reef to receive coral recruits from its neighbors. For a given reef cell, it is calculated by defining all the neighboring reef cells that can reach this target reef cell under a determined CDt. ICI is the total area (in km^2) of these departure reef cells.

We set the value of CDt to 800 units in order to maximize the neighborhood extent without causing border effects. This value was calculated based on the reef cells' cost distance to and from the borders of the study area (located ~250 km around the most peripheral reef cells; Fig. S4-4), where the minimal cost distances to and from the border were 836 and 801 units, respectively.

5.5.4. SNPs dataset

We retrieved genomic data employed in previous seascape genomics analyses on three coral species of New Caledonia: *Acropora millepora, Pocillopora damicornis* and *Pocillopora acuta* (Selmoni, Lecellier, Magalon, et al., 2020). This dataset encompassed more than one hundred individuals per population (167 in *A. millepora,* 118 in *P. damicornis*, 110 in *P. acuta*), collected at multiple sampling sites around Grande Terre (20 sites for *A. millepora,* 17 for *P. damicornis*, 17 for *P. acuta*) and genotyped using a Genotype-By-Sequencing approach (Kilian et al., 2012) characterizing thousands of single-nucleotide-polymorphisms (SNPs; 11,935 in *A. millepora*, 7,895 in *P. damicornis* and 8,343 in *P. acuta*). Of note, SNPs in this dataset were already filtered for rare allelic variants (minor allele frequency<0.05%) and linkage disequilibrium (LD-pruning threshold=0.3 ;Zheng et al., 2012).

5.5.5. Probability of heat stress adaptation

The previous seascape genomics study investigated the genotype-environment associations between SNPs and 47 environmental descriptors (among which is $BAF_{overall}$) using LFMM software (Frichot et al., 2013; Selmoni, Lecellier, Magalon, et al., 2020). In each of the three species, the analysis reported significant associations ($q<0.01$) of $BAF_{overall}$ with potentially adaptive SNPs (10 in *A. millepora*, 18 in *P. damicornis,* and 4 in *P. acuta*). We employed these genotype-environment associations to predict the probability of heat stress adaptation (PA_{HEAT}) from $BAF_{overall}$ values. We used a method based on logistic regressions (Joost et al., 2007; Rochat & Joost, 2019) that was previously applied to corals (Selmoni, Rochat, Lecellier, Berteaux-Lecellier, et al., 2020), with some modifications outlined hereafter.

First, we evaluated the accuracy of the approach. For each individual used in the analysis, we retrieved the $BAF_{overall}$ value at the sampling location. Next, we encoded the presence/absence of the every putatively adaptive genotype as a binary variable using a custom function in the R environment. For every putatively adaptive genotype, we constructed a logistic genotype-environment association model with $BAF_{overall}$. These models define the expected frequency of a genotype of interest for a given $BAF_{overall}$ value. We evaluated the predictive accuracy of every model by running a cross-validation using a "leave-one-population-out" approach. This approach consisted in excluding all samples from one sampling site during model training, and then in using the model to predict the expected genotype frequency at that site. This procedure was reiterated for every sampling site, and the correlation (Pearson) between the observed and expected genotype frequencies was calculated for every putatively adaptive genotype. As a comparison, we applied the same cross-validation method for 1000 genotypes randomly selected in each of the three species.

Next, we employed a generalize linear mixed model (GLMM) to build an overall genotype-environment association model combining all the putatively adaptive genotypes across the three species. This was done through the R package glmmTMB (v 1.0; Brooks et al., 2017), using a logistic regression model where genotype identifier, sample identifier and species were introduced as random factors. The resulting model then was used to transform $BAF_{overall}$ values associated with each of the 2,284 reef cells of New Caledonia in PA_{HEAT}. The model was plotted using the visreg R package (v. 2.6.1; Breheny & Burchett, 2017).

5.5.6. Reef connectivity and genetic correlations between corals

The SNPs dataset was used to evaluate whether reef connectivity predictions were representative proxies of the structure of three coral populations. In the R environment, we applied the following framework to the genotype matrix of each of the three species. First, we evaluated the relatedness between each pair of individuals in the dataset (13,861 pairs in *A. millepora*, 6,903 pairs in *P. damicornis*, 5,995 pairs in *P. acuta*) by calculating the genetic correlation (Pearson) based on SNPs values (Novembre et al., 2008; Patterson et al., 2006).

We then computed the distribution of the genetic correlation values, and excluded pairs of individuals with anomalously high or low correlation (*i.e.* exceeding the boundaries defined by the median of the distribution ± three times the interquartile range).

Next, we investigated the drivers of genetic correlations by using GLMMs designed through the R package glmmTMB (Brooks et al., 2017). We set two fixed effects as possible drivers of genetic correlation between individuals: ancestral distance and reef connectivity. Accounting for ancestral genetic structure is particularly important as corals are prone to hybridization or cryptic speciation (Schmidt-Roach et al., 2013; van Oppen et al., 2002). The computation of ancestral distance featured the R package ALStructure (v. 0.1; Cabreros & Storey, 2019). For a given SNP matrix, ALStructure predicts the number of ancestral populations, and then estimates, for every individual, the admixture proportions to the ancestral populations. For every pair of individuals, we then calculated the ancestral distance as the Euclidean distance between the respective admixture proportions. For which concerns reef connectivity, the fixed effect corresponded to the least-cost-path (from the seascape graph) linking the sampling sites of every pair of individuals. Since genetic correlations could not exceed the 0-1 boundaries, GLMMs were built using a beta regression (Ferrari & Cribari-Neto, 2004). In order to account for the non-independence of different corals pairs, the random factors in the GLMM were the identifiers of the individuals in the pairs, as well as the identifier of the pair of sampling sites.

Finally, we evaluated the relationship between ancestral distance, reef connectivity and genetic correlations of corals by 1) reporting the estimate and its standard deviation, as well as the p-value associated with Wald statistic (Brooks et al., 2017); 2) plotting the association using the visreg R package (Breheny & Burchett, 2017).

5.5.7. Coral cover data

Living coral cover data was retrieved from the 2017-18 report of the New Caledonian observational network of coral reefs ('Réseau d'observation des récifs coralliens de Nouvelle Calédonie', RORC; Job, 2018). Overall, we used data from 74 survey stations distributed across the Archipelago of New Caledonia (Fig. 5-1). At each station, yearly coral cover surveys were performed along the same 100m transect using the "point intercept" technique. Surveys covered the period from 2003 to 2017, where 18 sites have been visited for less than five years, 27 for five to ten years, and 29 for more than ten years. The exact coordinates of survey stations were retrieved from the geographic information web-portal of New Caledonia (https://georep.nc/).

5.5.8. Environmental characterization of survey sites

The coordinates of survey stations were used to find the corresponding reef cells and the associated values of the connectivity indices (OCI and ICI). For each survey record (*i.e.* survey

at a given station in a specific year) we also calculated $BAF_{overall}$ as the average BAF since 1985 to the year proceeding the survey. Based on the values of $BAF_{overall}$ we computed PA_{HEAT} for each survey record. In addition, we calculated BAF values on a rolling temporal window describing average BAF for the year ($BAF_{previous\ year}$) that preceded the year of survey.

5.5.9. Analysis of coral cover change

We investigated the association of $BAF_{previous\ year}$, PA_{HEAT} and connectivity indices (ICI and OCI; in total 4 explanatory variables) with coral cover rates (response variable) using GLMMs. This analysis focused on the coral cover rates of every survey record (total of 574 records). The computation of GLMMs was performed using the R package glmmTMB (v 1.0; Brooks et al., 2017), which allowed us to model coral cover rates via beta regression (Ferrari & Cribari-Neto, 2004). We accounted for the non-independence of survey records originated at the same station but on different years by setting the station effect as random factor on the coral cover rate (Verbeke, Molenberghs, & Rizopoulos, 2010). This approach is recommended for studies of longitudinal data with irregular time points (Garcia & Marder, 2017). To avoid bias due to scale differences between explanatory variables, each variable was standardized to mean 0 and standard deviation 1 using the R "scale" function.
We built two types of GLMMs: univariate and bivariate. In univariate GLMMs, $BAF_{previous\ year}$, PA_{HEAT}, ICI and OCI were employed each as unique fixed effect. The goal was to determine whether the explanatory variables showed a standalone association (*i.e.* independent from other variables) with coral cover change. In bivariate models, GLMMs were constructed each with two fixed effects: 1) $BAF_{previous\ year}$ and 2) the interaction between $BAF_{previous\ year}$ and each of the remaining explanatory variables: PA_{HEAT}, ICI and OCI. The goal of bivariate models was to investigate whether the potential effect of recent thermal stress ($BAF_{previous\ year}$) on coral cover might be modulated by PA_{HEAT}, ICI or OCI.
For each GLMM, we reported the estimate and its standard deviation, as well as the p-value (deemed significant when <0.05) associated with Wald statistic (Brooks et al., 2017) of the fixed effects. In addition, we compared the quality-of-fit of models by calculating the Akaike Information Criterion (AIC; Bozdogan, 1987).

5.6. Acknowledgments

We thank the *Réseau d'observation des récifs coralliens* of New Caledonia for collecting and sharing the field survey data. We thank Annie Guillaume and the anonymous reviewers for the comments and suggestions provided during the redaction of this paper. This work was supported by the United Nations Environment Programme (UNEP) and International Coral Reef Initiative (ICRI) coral reefs small grants programme (grant number: SSFA/18/MCE/005).

We also thank the Government of France and the Government of the Principality of Monaco who provided the funding for the small grants.

6. General discussion

In this last chapter, I resume how the findings of this thesis allowed us to answer the research questions stated in the introduction. Then I discuss the perspectives for coral conservation and seascape genomics in light of what was uncovered in the context of this work. Finally, we summarize the main achievements of this thesis in a general conclusion.

6.1. Answers to research questions

1. How does the sampling strategy drive statistical power of landscape genomics analyses?

Our research confirmed that among the elements composing a sampling strategy, sample size is the main driver of statistical power and false discoveries. Article A suggests to sample at least 200 individuals to enable the detection of true adaptive signals. In SABLE, budget choices forced us to work below this threshold, but in turn we selected sampling sites with a design approach maximizing environmental variability and anticipating the confounding effect of population structure. In the simulations of article A, this design approach was shown to increase statistical power and reduce false discoveries. These findings therefore highlight the importance of accounting for environmental variability when designing the sampling plan, as this can provide benefits to the analysis at no cost.

2. Can seascape genomics characterize the adaptive potential of corals against heat stress?

The case studies in Japan and New Caledonia well showed that seascape genomics was able to disclose putative signals of adaptation to heat stress in four distinct coral populations. There are two observations suggesting that at least part of these adaptive signals are true: the first concerns the functional annotations of the genomic neighbourhood of candidate adaptive SNPs. In fact, we found annotations describing genes or molecular functions known to be implicated in coral heat stress responses, and these annotations were recurrent among different species. The second observation is that reefs predicted with a higher probability of adaptation suffered less loss of coral cover under heat stress, compared with reefs with a lower probability. However, as discussed in articles B, C and D, these observations cannot be considered conclusive validations. For this purpose, experimental validation as *ex-situ* aquarium conditioning, remains necessary.

3. What are the molecular mechanisms implicated in heat stress adaptation in coral?

The seascape genomics analyses performed revealed multiple potential molecular targets for heat stress adaptation in corals. Among these, we found molecular actors implicated in cellular pathways well-known to participate in coral cellular response against heat: the activation of the protein folding machinery in the endoplasmic reticulum, the homeostasis of reactive oxygen species, and the related inflammatory and apoptotic responses. Of note, adaptive signals targeting these pathways recurred between species. In addition, we found new pathways potentially implicated in heat stress responses and adaptation. The most interesting is the DNA repair mechanisms and Helicases Q, which recurred in the Japanese case study and appeared also in New Caledonian corals. These molecular candidates can have important implications in coral research. For instance, presumptive heat adapted vs. non-adapted colonies could undergo experimental conditioning coupled with transcriptome profiling, to provide a better characterization of the role of these pathways in heat stress responses. Future studies should also consider the adaptive signals in the Symbiodiniaceae genome, overlooked here because of the low amount of symbionts' sequences obtained with the DNA purification method we used. Considering the role of symbionts would provide a more comprehensive view of adaptation against bleaching.

4. How can information on coral adaptive potential support reef conservation?

We developed a framework to transpose results from seascape genomics analyses into indices tailored for conservation. These indices summarize two crucial aspects that are often overlooked in reef conservation strategies: connectivity between reefs and adaptation against climatic constraints. The format of these indices is appropriate to support conservation planning for two reasons: 1) they are calculated at the scale of an entire reef system; 2) the indices are objective and quantitative. Because of these features, it is possible to prioritize reefs of the same region, or to compare the conservation status between different regions. Furthermore, these indices are mappable and this facilitates the uptake among non-specialists.

6.2. Perspectives for coral reef conservation

Previous work suggests the presence of entire reef systems that can act as coral refugia against thermal stress in the decades to come (Van Hooidonk et al., 2013). These are reefs systems like in the Red Sea, where average temperatures are generally high, such that corals have an higher threshold of thermal tolerance compared with the rest of the world (Osman et al., 2018). The two case studies presented in this thesis showed that contrasted thermal

patterns driving local adaptation might exist also on relatively small scales (<100 km). This observation would imply that every reef system could host local thermal refugia against coral bleaching. This view has important implications on reef conservation at both local (reef systems of single countries) and regional (interconnected reef systems) scale.

On a local perspective, policy makers should empower reef conservation strategies based on information for adaptive potential. Our results suggest that the impact of heat stress on coral cover is mitigated at reefs predicted to show a higher adaptive potential, but also that a higher heat adaptive potential is not advantageous in absence of heat stress. On the contrary, the physiological costs of adaptation to heat stress might even cause adapted corals to be counter-selected when exposed to other type of stressors (*e.g.* pollution; Kenkel et al., 2015). This is why protecting heat-adapted corals using MPAs, which are effective against local disturbances, is essential (Lester et al., 2009).

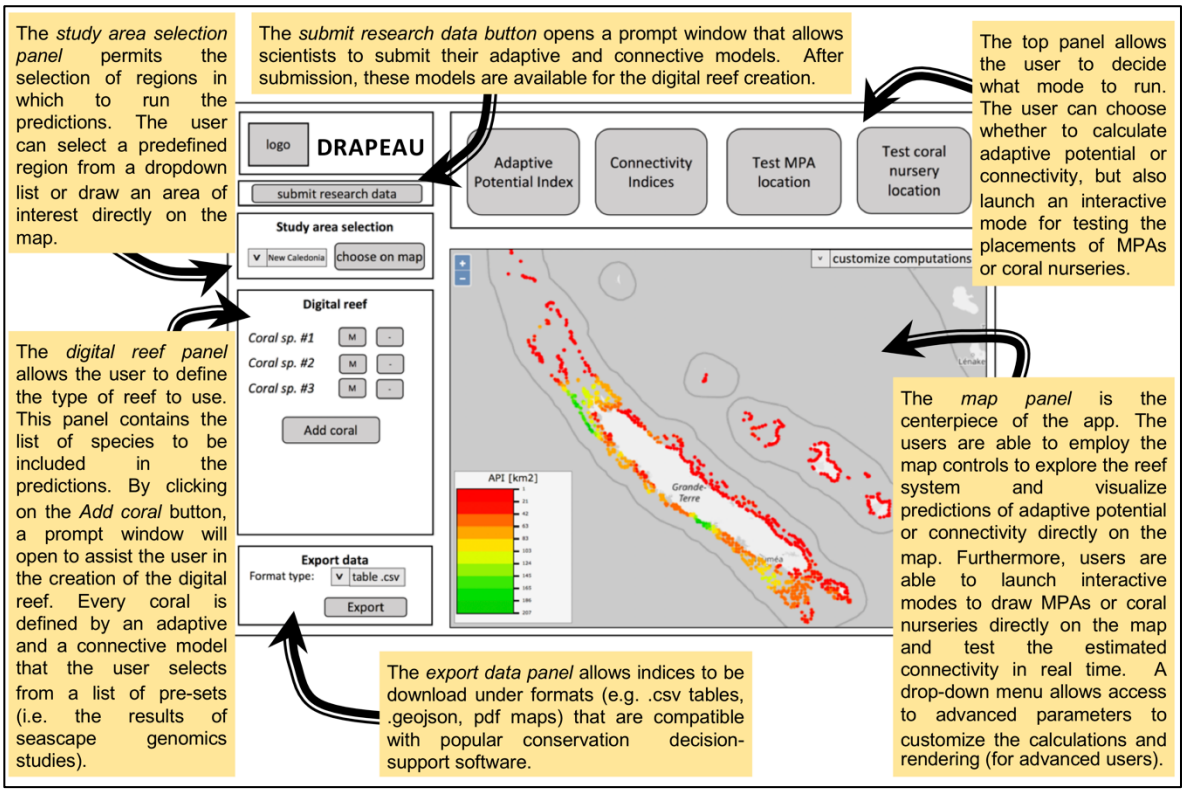
Another local conservation action that could benefit from information on adaptive potential is reef restoration. Restoration of damaged reefs could be performed using coral breeds originated from heat-stress exposed reefs (Baums et al., 2019). This practice could be supported by low-cost genetic analysis (*e.g.* allele specific PCR: Matsuda, 2017) to identify heat-stress adapted corals prior to transplantation. The development of a genetic tool of this kind requires supporting studies as the one presented in article C. Importantly, transplanting adapted corals within the same reef system will minimize the risks of introducing invasive breeds (Baums et al., 2019). Indeed, the consequences of importing corals from refugia located on distant reefs systems, or from artificial breeding in laboratory, are hard to anticipate and could even be disastrous (Timar & Phaneuf, 2009).

Regarding conservation at a regional scale, there is a growing need for collaboration between countries (Mazor, Possingham, & Kark, 2013). One crucial concern is the global network of MPAs, which does not guarantee sufficient connectivity between protected areas (Mora et al., 2006). The indices of reef connectivity could therefore provide a valuable support for planning of MPAs at both local and regional scales (Magris et al., 2014; Mora et al., 2006). The case studies presented in this thesis provide good examples for both situations. In the Ryukyu Archipelago, the northward-oriented Kuroshio current means that conservation actions on southern reefs have an impact on the northern ones (local connectivity). In New Caledonia, reefs in the Loyalty Islands are likely to receive few incoming propagules from reefs around Grande Terre, and therefore depend on propagules arriving form Vanuatu (regional connectivity). Optimizing MPA networks by accounting for connectivity and heat stress adaptive potential will create a win-win situation for all coral reef countries (Magris et al., 2014; Selig & Bruno, 2010).

To plan reef conservation strategies that are effective at both local and regional scales, it is essential to facilitate communication between all the coral reef stakeholders. First, it is necessary to promote exchanges between policy makers from neighbouring reef systems. In this regard, a good example is provided by the Transnational Red Sea Center (TRSC) which federates countries of the Red Sea in a common effort to protect the reef of the region

(Kleinhaus et al., 2020). Second, communication should be facilitated between reef researchers and conservation managers. The present research, for instance, shows how information on adaptive potential and connectivity can be transposed from an academic context into a practical reef conservation perspective. The same can be done for other innovative fields of coral reef research; to facilitate the practical transposition of such research into conservation, the ManaCo consortium was created (Selmoni, Lecellier, Ainley, et al., 2020; appendix 5). This is an international network that gathered coral reef scientists, conservation managers and local communities from 13 countries around the same table. The goal of ManaCo is to use innovative tools such as web-applications (*e.g.* Fig. 6-1) and online resources to orientate research toward the practical needs of conservation managers and consequently facilitate the application of research in preservation plans. International, multidisciplinary networks such as those described above will be decisive in coordinating effective conservation strategies to protect coral reefs in the years to come.

Figure 6-1. The Digital Reef Adaptive Potential EvAlUator (DRAPEAU). DRAPEAU is a web-application prototype conceived within the framework of the ManaCo consortium. The application features an intuitive click-on-map interface to compute predictions for reef connectivity and adaptive potentials. In addition, an interactive mode evaluates the predicted impact of management strategies such as the establishment of marine protected areas. The goal of DRAPEAU is to facilitate the transfer of knowledge from research (such as presented in this thesis) to practical management of reef conservation. This figure shows a mock-up of the DRAPEAU interface and overviews the main functionalities. DRAPEAU has be the subject of a training workshop during ManaCo kick-off meeting.



6.3. Perspectives in seascape genomics

The degradation of ocean conditions is already threatening the survival of a wide range of marine species (Webb & Mindel, 2015). Understanding how these organisms adapt to a changing environment and identifying the taxa that will persist in the seascape of the future will be of paramount importance (Halpern & Kappel, 2013). Despite this, landscape genomics experiments focusing on marine species remain underrepresented in the scientific literature to date (Riginos et al., 2016). With the present thesis, we advocate the application of seascape genomics to characterize the adaptive potential of endangered marine taxa. Based on the experience gained from two case studies on corals, we review hereunder some critical points to pay attention to in future seascape genomic studies.

First, sample size is the key for reliable statistical inference. In our case studies, we worked with sub-optimal (<200 units) sample sizes and we inevitably had to deal with the burden of higher false discoveries. On the other hand, we showed that rational choices accounting for environmental variability in sampling design can partially compensate this issue. Practical methods to apply these rational choices to any dataset are made available in article A. One example is the *hybrid* approach of sampling design, that consists in first determining regions with distinct environmental conditions, and then in spreading sampling sites across each region. Furthermore, care must be taken when increasing the number of sampling locations, as unnecessarily high numbers impact the costs of the study while bringing little benefit to statistical power.

Remote sensing derived datasets appeared to describe the seascape conditions at sufficient spatial resolutions in order to capture adaptive processes. In article C, however, we observed that increasing the resolution of sea surface temperature from five to one kilometre raised the number of potentially adaptive genotype-environment associations. This suggests that an even higher resolution might boost the statistical power of the analysis. Datasets with these characteristics already exist for SST (*e.g.* 100 m resolution; Vanhellemont, 2020) and might become available for other variables in the years to come.

As regards genomics, DArT-seq used in the New Caledonian case study appeared as a valuable solution for cost-effective genotyping. This is because the genomic neighbourhood of SNPs genotyped with DArT-seq was enriched in genes, and this facilitated the search for potentially adaptive variants (Lowry et al., 2017). Furthermore, our work in Japan and New Caledonia highlighted the importance of working on a species with an available reference genome, as this facilitates the interpretation of the potential molecular implications of genetic variants.

Furthermore, it is worth mentioning that the methods we used to assess the over-representation of molecular functions are borrowed from gene expression profiling experiments (Alexa et al., 2006; M. D. Young, Wakefield, Smyth, & Oshlack, 2010). These methods are not tailored for gene-sets derived from landscape genomics and therefore might fail to capture important molecular signatures in genes surrounding adaptive SNPs. Thus, we

advocate the creation of novel statistical methods to assess the over-representation of molecular functions in landscape genomic experiments.

Molecular annotations might also be beneficial to methods investigating significant genotype-environment associations. In fact, the two approaches we used (SamBada and LFMM) assume independence between SNPs and do not account for any information on the potential molecular implications of a variant (Frichot et al., 2013; Stucki et al., 2017). However, in articles B and C we observed the presence of genes participating in the same cellular pathway next to SNPs associated with heat stress. A method integrating functional annotations directly into genotype-environment association analysis would be completely new to landscape genomics, and might facilitate the discovery of adaptive genotypes composed of multiple loci.

The last point with room for improvement concerns experimental validation. The use of experimental conditioning as a follow-up of seascape genomics analyses should become praxis as the two methods are highly synergistic. On one hand, seascape genomics explores the variability of adaptive responses in a population, and this facilitates the choice of the genotypes when compared with an experimental set-up. On the other hand, experimental conditioning investigates adaptation without confounding factors, and can measure the degree of tolerance (*e.g.* how many degrees above seasonal maximal temperature) conferred by an adaptive genotype (Pardo-Diaz et al., 2015; Rellstab et al., 2015).

6.4. General conclusion

In this thesis, we used a seascape genomics approach to characterize the adaptive potential of reef building corals against heat stress. This approach was applied to two case studies: one coral species from the Ryukyu Archipelago (Japan) and three corals species from New Caledonia (Southern Pacific). We developed a framework to transpose the findings of seascape genomic analyses into practical indices of connectivity and heat-stress adaptive potential. These indices are objective, quantifiable and mappable across an entire reef system, and therefore constitute valuable indicators to support reef conservation. Field observations suggest that reefs predicted to show high values of heat-stress adaptive potential suffer less coral loss under heat stress, when compared with reefs with low values of adaptive potential. Furthermore, the seascape genomics approach allowed us to uncover the potential molecular targets of heat-stress adaptation in corals. Some of these candidates referred to well-known cellular response to heat in corals, while others suggest new pathways involved. Finally, this research contributed to the development of the landscape and seascape genomic approaches, by defining practical guidelines for sampling strategies in order to maximize statistical power and minimize false discoveries.

7. Appendices

7.1. Supplementary material from article A

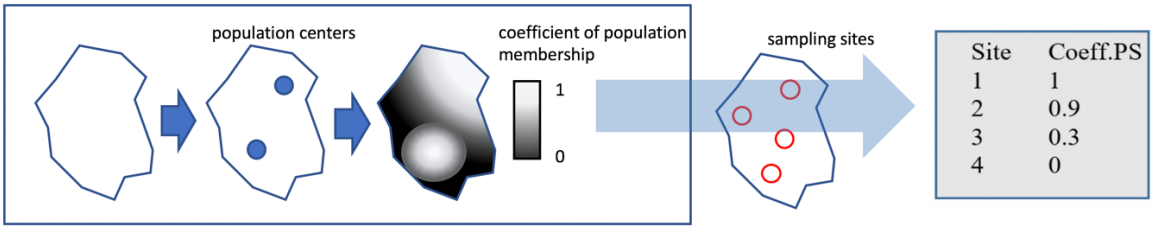
Supplementary Table 2-1. List of environmental variables employed.

Name	Geographic resolution	Source
Annual Mean Temperature	2.5 minutes	Bioclim (BIO1)
Mean Diurnal Range	2.5 minutes	Bioclim (BIO2)
Temperature Seasonality	2.5 minutes	Bioclim (BIO4)
Mean Temperature of Wettest Quarter	2.5 minutes	Bioclim (BIO8)
Annual Precipitation	2.5 minutes	Bioclim (BIO12)
Precipitation Seasonality	2.5 minutes	Bioclim (BIO15)
Precipitation of Warmest Quarter	2.5 minutes	Bioclim (BIO18)
Altitude	100 m	Marine Geoscience Data System

Supplementary Box 2-1. Computation of the genotype matrix. The vignettes describe how genotypes were computed during simulations. At each iteration, a new genotype matrix counting 1'000 loci was generated. Ten of them were set as adaptive and followed the respective pipeline, while the others were set as neutral and computed accordingly.

A) Neutral Locus

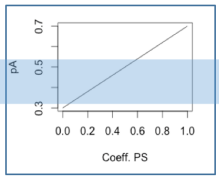
i. An artificial population membership coefficient is computed as the distance from randomly located population centers. The membership coefficient is extracted then at each sampling site.



Site	Coeff.PS
1	1
2	0.9
3	0.3
4	0

ii. A function translates the coefficient of population structure in the probability of carrying the allele characteristic of the population. Finally, alleles are sampled at each site using the probability associated. This step is reiterated if more than one individual is sampled at the same site and for all the loci related to a same population membership coefficient.

Site	Coeff.PS
1	1
2	0.9
3	0.3
4	0



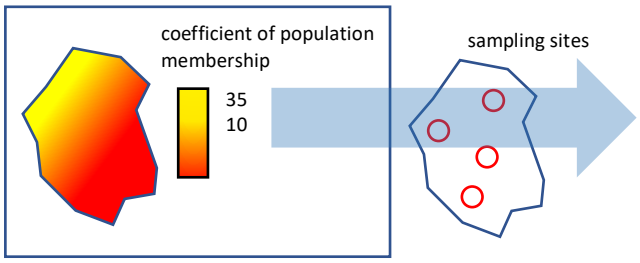
Site	pA |PS
1	0.7
2	0.65
3	0.4
4	0.3

Site	A1	A2
1	A	A
2	A	a
3	a	a
4	a	a

Site	GT
1	AA
2	Aa
3	aa
4	aa

B) Adaptive Locus

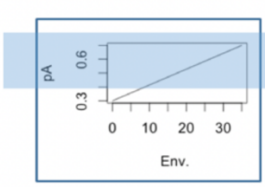
i. For each sampling site, the environmental values are extracted.



Site	Env
1	32
2	12
3	20
4	15

ii. A function computes the probability of carrying an allele conferring a selective advantage against the environmental condition. Alleles are sampled at each site using the probability associated. This step is reiterated if more than one individual are sampled at the same site.

Site	Env
1	32
2	12
3	20
4	15



Site	pA|Env
1	0.65
2	0.4
3	0.5
4	0.45

Site	A1	A2
1	A	A
2	A	a
3	A	a
4	a	a

Site	GT
1	AA
2	Aa
3	Aa
4	aa

Supplementary Box 2-2. Formulae and parameters for genotype computations

The probability function for the allele A depending on a population membership coefficient is calculated as follows:

$$p(A|PS) = \left(\frac{1-2c}{\max(PS) - \min(PS)}\right) PS + c - \left(\frac{1-2c}{\max(PS) - \min(PS)}\right) \min(PS)$$

where *PS* is a population membership coefficient and *c* a parameter representing the strength of the relationship. This parameter can range between 0 (strongest relation, *i.e.* maximal and minimal *PS* returns *p*=1 and *p*=0, respectively) and 0.5 (no relation, any level of *PS* returns *p*=0.5).

Similarly, probability for the allele A depending on environmental selection is calculated as follows:

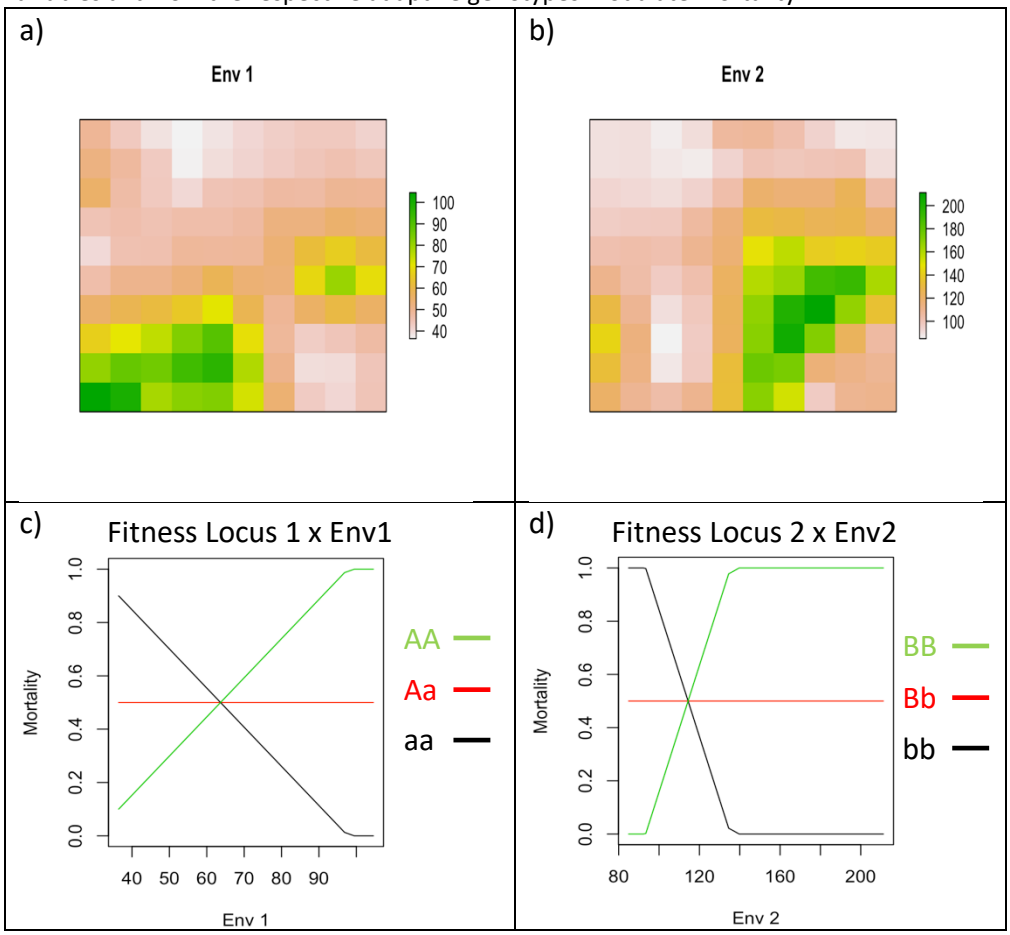
$$p(A|Env) = \left(\frac{1-2s_1}{\max(Env) - \min(Env)}\right) E + s_1 - \left(\frac{1-2s_1}{\max(Env) - \min(Env)}\right) \min(Env) + s_2$$

where *Env* are the values of the environmental variable and s_1 represents the strength of selection and behaves as the *c* in the previous equation. The additional parameter s_2 provides a baseline of allele frequency.

In our simulations, we set two scenarios employing the following parameters:

- *panmictic population scenario* (random neutral structure): *c=0.5,* s_1*=Unif(0.3, 0.4),* s_2*=Unif(-0.2,0.1)*
- *structured population scenario* (strong population structure): *c=Unif(0.2,0.4),* s_1*=0,* s_2*=Unif(-0.1,0.2)*

Supplementary Figure 2-1. Environmental gradients and fitness constraint employed in the CDPOP validation run. Panel a) and b) show the distribution of the two environmental variables across the 10-by-10 cells grid used for the CDPOP simulation. Plots in panels c) and d) show the fitness constraint set for the two environmental variables and how the respective adaptive genotypes modulate mortality.



Supplementary Table 2-2. CDPOP vs. our simulative approach comparison metrics. The tables show the rank of the simulative variants computed with our method (and defined by parameters *m, c* s_1 *and* $_{s2}$) that best matched the CDPOP replicates. In a) and b) are shown the metrics used to compare the neutral genetic structure with the CDPOP case of a panmictic population and a structured population, respectively. The three metrics employed are 1) the average random mean squared error (RMSE) when comparing the curves describing the differential of explained variation by the genetic principal components; 2) the Kullback-Leibler Divergence (KLD) used to compare the pairwise-Fst distributions; 3) the difference in the average mantel correlation (ΔmR), which describes the link between genetic and geographic distances. The ranking coefficient is the sum of the three scaled metrics. In c) and d) the comparison concerns the adaptive genotypes computed in panmictic structured scenario of CDPOP, respectively. Here the RMSE compares, for our simulation and CDPOP runs, the allelic frequency of the adaptive genotype as a function of the environmental variable causing adaptation

a) Panmictic Scenario: Neutral structure metrics

rank	*m*	*c*	RMSE (PCA)	KLD (Fst)	ΔmR	Ranking Coefficient
1	1	0.5	0.000780575	7.33E-06	0.003577	-4.35661
2	25	0.4-0.5	0.000771722	7.70E-06	0.022455	-4.25828
3	10	0.4-0.5	0.000771901	7.93E-06	0.023357	-4.24377
4	20	0.4-0.5	0.000780659	8.58E-06	0.022308	-4.21677
5	5	0.4-0.5	0.000770043	7.46E-06	0.034877	-4.21321
6	15	0.4-0.5	0.000766353	9.31E-06	0.025071	-4.17643
7	5	0.4-0.4	0.000796873	1.15E-05	0.067273	-3.88113
8	10	0.4-0.4	0.000763216	1.12E-05	0.074199	-3.87217
9	25	0.4-0.4	0.000771422	1.27E-05	0.072328	-3.81237
10	20	0.4-0.4	0.000761967	1.38E-05	0.073625	-3.7593

b) Structured Scenario: Neutral structure metrics

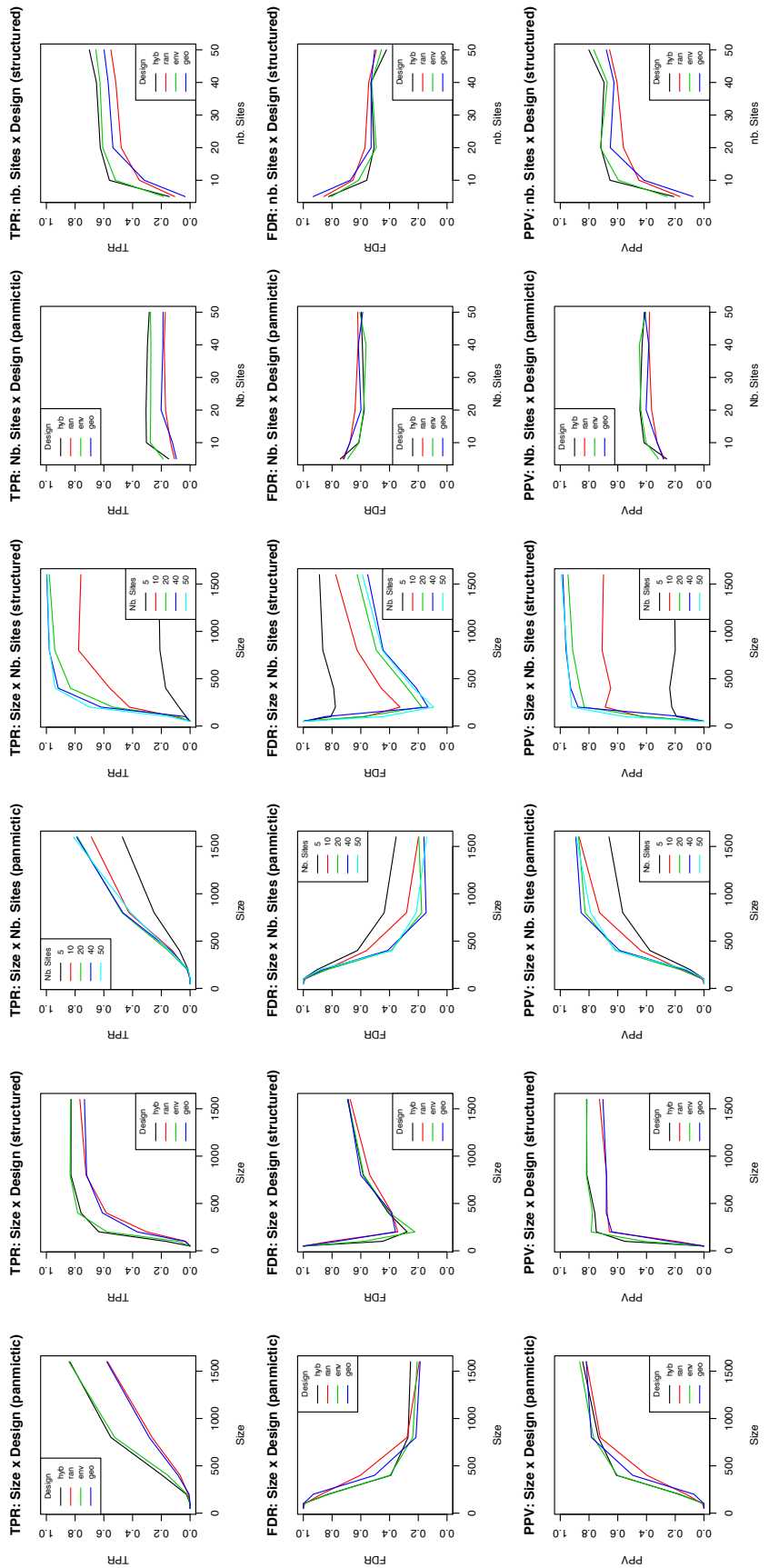
rank	*m*	*c*	RMSE (PCA)	KLD (Fst)	ΔmR	Ranking Coefficient
1	10	0.2-0.4	0.00290909	8.17E-06	0.320549	-3.63827
2	20	0.1-0.5	0.00266099	8.85E-06	0.339198	-3.63027
3	5	0.3	0.003023145	8.38E-06	0.312132	-3.45645
4	15	0.1-0.5	0.002793301	7.57E-06	0.37057	-3.43066
5	25	0.2-0.4	0.003250162	8.42E-06	0.314625	-3.31517
6	15	0.2-0.3	0.002468453	6.72E-06	0.422087	-3.31507
7	5	0.2-0.4	0.003092629	9.91E-06	0.329403	-3.27752
8	10	0.3	0.002819477	9.84E-06	0.295631	-3.26125
9	25	0.1-0.5	0.002947686	8.05E-06	0.373038	-3.23848
10	15	0.2-0.5	0.002799946	1.02E-05	0.280361	-3.09366

c) Panmictic Scenario: adaptive genotypes metrics

Moderate Selection			
rank	s_1	s_2	RMSE (AF)
1	0	-0.1	0.7417767
2	0.1	-0.1	0.75108
3	0.1	-0.2	0.7681983
4	0	-0.2	0.78917
5	0.2	-0.1	0.7946361
Strong Selection			
rank	s_1	s_2	RMSE (AF)
1	0	0.2	0.676855
2	0.1	0.2	0.683247
3	0.1	0	0.710474
4	0	0.1	0.715619
5	0.2	0.1	0.728321

d) Structured Scenario: adaptive genotypes metrics

Moderate Selection			
rank	s_1	s_2	RMSE (AF)
1	0.4	-0.2	0.6889893
2	0.3	-0.2	0.6895106
3	0.2	-0.2	0.7181186
4	0.3	-0.1	0.7319583
5	0.2	-0.1	0.7454251
Strong Selection			
rank	s_1	s_2	RMSE (AF)
1	0.3	0.1	0.624262
2	0.4	0.1	0.6417665
3	0.2	0.1	0.6484901
4	0.3	0	0.6709922
5	0.4	0	0.6831192



Supplementary Figure S 2-2. Interaction effects. The table displays different combination of the elements defining the sampling strategy and their effect on the average of the three diagnostic parameters (row 1: TPR, row 2: FDR, row 3: PPV). For every diagnostic parameter, the two demographic scenarios are represented (column 1,3 and 5: panmictic, 2,4 and 6: structured). In columns 1-2, the combined effects of sample size (x axis) and sampling design (colored lines) are observed. In columns 3-4, the combined effects of sample size (x axis) and number of sampling locations (colored lines) are observed. In columns 5-6, the combined effects of number of sampling locations (x axis) and sampling design (colored lines) are observed.

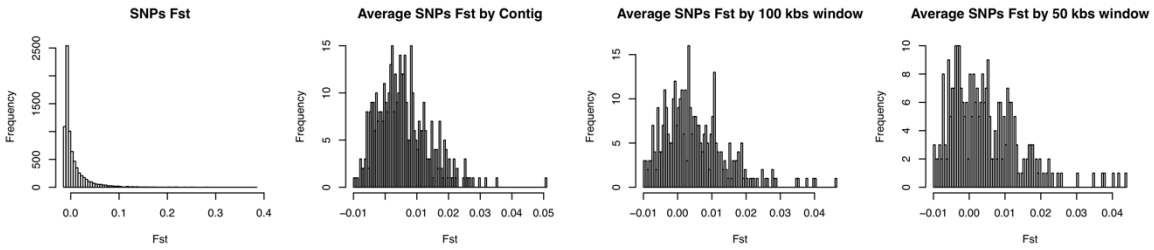
7.2. Supplementary material from article B

Supplementary Table 3-1. Environmental Variables included in the Seascape Genomics analysis. For each remote sensing product, the table shows the sources (CMEMS= Copernicus Marine Environment Monitoring System; NOAA= National Oceanic and Atmospheric Administration; IEDA= Interdisciplinary Earth Data Alliance) and the corresponding identifier (Product Name). Temporal range and resolution indicate the duration and the frequency of the remote sensing period, respectively. The variables calculated for the seascape genomics analysis are listed in the *Derived Variables* column with the following abbreviations: OM (overall mean), OMsd (overall standard deviation), HM (highest monthly mean), LM (lowest monthly mean), HMsd (standard deviation of the month with highest mean), LMsd (standard deviation of month with lowest mean).

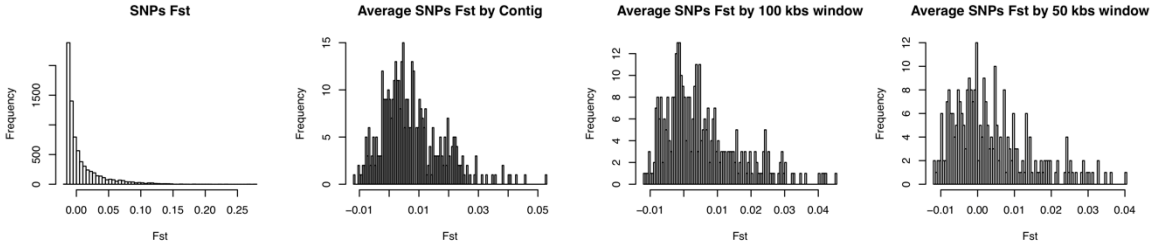
Variable Name	Product Name	Source	Temporal Range	Temporal Resolution	Spatial Resolution	Derived Variables
Sea Surface Temperature	SST_GLO_SST_L4_REP_OBSERVATIONS_010_011	CMEMS	1985-2007	Daily	0.05 °	OM, OMsd, HM, LM, HMsd, LMsd
Sea Surface Salinity	GLOBAL_REANALYSIS_PHY_001_030	CMEMS	1993-2010	Daily	0.083 °	OM, OMsd, HM, LM, HMsd, LMsd
Chlorophyll Concentration	OCEANCOLOUR_GLO_CHL_L4_REP_OBSERVATIONS_009_082	CMEMS	1998-2010	Daily	4 km	OM, OMsd, HM, LM, HMsd, LMsd
Sea Currents Velocity	global-reanalysis-phy-001-030	CMEMS	1993-2010	Daily	0.083 °	OM, OMsd, HM, LM, HMsd, LMsd
Suspended Particulate Matter	oc-glo-opt-multi-l4-spm	CMEMS	1997-2010	Monthly	4 km	OM, HM, LM
Photosynthetically Available Radiations	erdMH1par0mday	NOAA	2003-2010	Monthly	4 km	OM, HM, LM
Bathymetry	Global Multi-Resolution Topography	IEDA	-	-	down to 100 m	Depth
Population Density	SEDAC	Columbia University	2015	-	0.1 °	Population density on buffer 50 km
Bleaching-alert-temperature	Derived from SST		1985-2007	Daily	0.05 °	Frequency
Alkalinity	Derived from SST + pH		1993-2007	Daily	0.05 °	OM, OMsd, HM, LM, HMsd, LMsd

Supplementary Figure 3-1. Fst analysis by genomic position. The fixation index (Fst) between each pair of sub-populations (a: Okinawa and Kerama, b: Okinawa and Yaeyama, c: Yaeyama and Kerama) was calculated for every SNP in the filtered genetic dataset (graphs on the first column). Fst values were then averaged by genomic position: either by contig (graphs on the second column), non-overlapping 100 kbs window (third column) or non-overlapping 50 kbs window (fourth column).

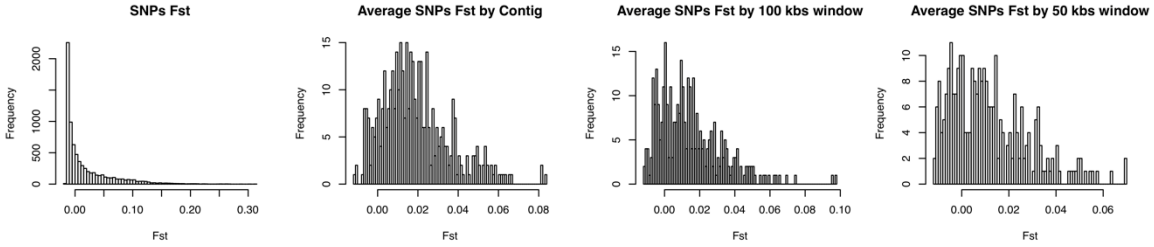
a) Fst Okinawa – Kerama



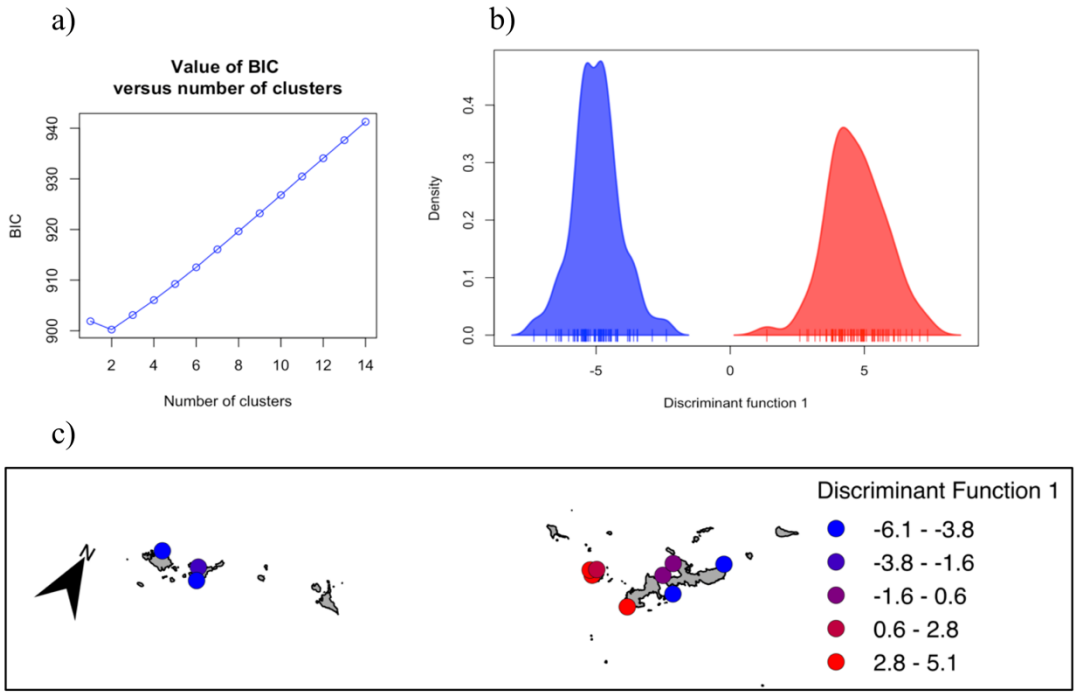
b) Fst Okinawa – Yaeyama



c) Fst Yaeyama – Kerama

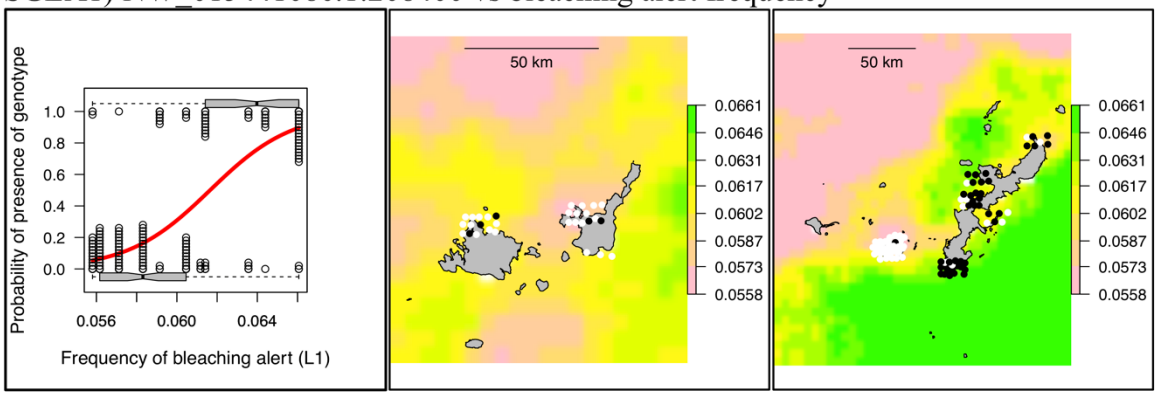


Supplementary Figure 3-2. Discriminant Analysis of Principal Components (daPCA) of the Genotype Matrix. A daPCA of the genotype table was performed to investigate the neutral structure of genetic variation in the population using the R adegenet package. Graph a) shows the (Bayesian Information Criterion) BIC value against the number of clusters, suggesting the presence of two groups. Graph b) displays that the first discriminant function allows to distinguish these groups and the map in c) shows the average of this value across sampling sites. We can see a strong contrast between sites reefs in the south and those in the North of Okinawa islands, together with those in the south of the Archipelago (Yaeyama).

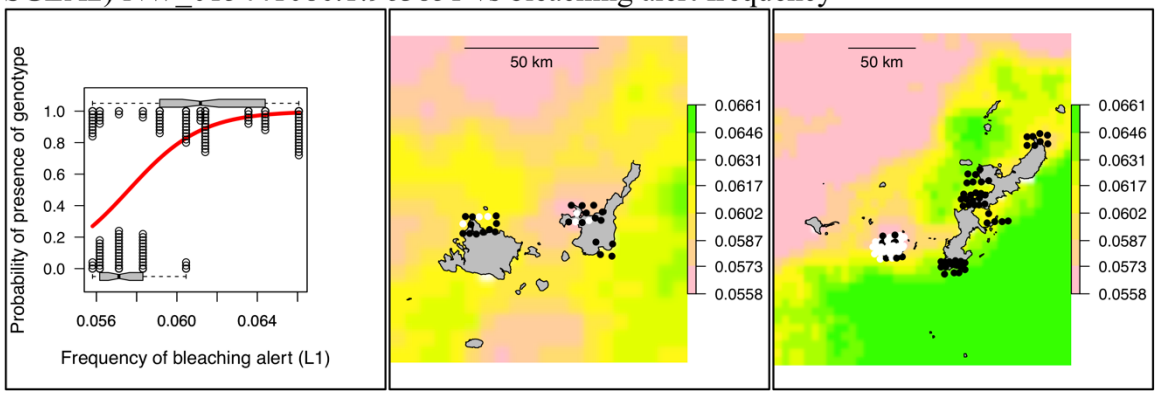


Supplementary Figure 3-3. Significant Genotype-Environment Associations (SGEA). For every SGEA, three display items are shown: the logistic model linking the genotype frequency with the environmental gradient of interest (left graph); the maps showing the genotype distribution (white points: colony without genotype, black points: colony with genotype) and the environmental gradient (background colors) in the Yaeyama area (central map) and Okinawa (right map). The points on the maps are scattered around the real sampling locations to facilitate the visualization.

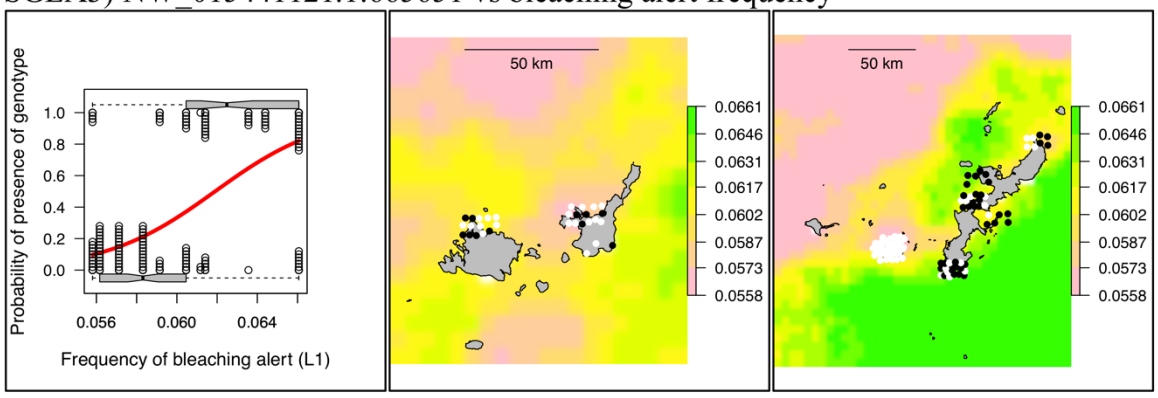
SGEA1) NW_015441080.1:208400 vs bleaching alert frequency



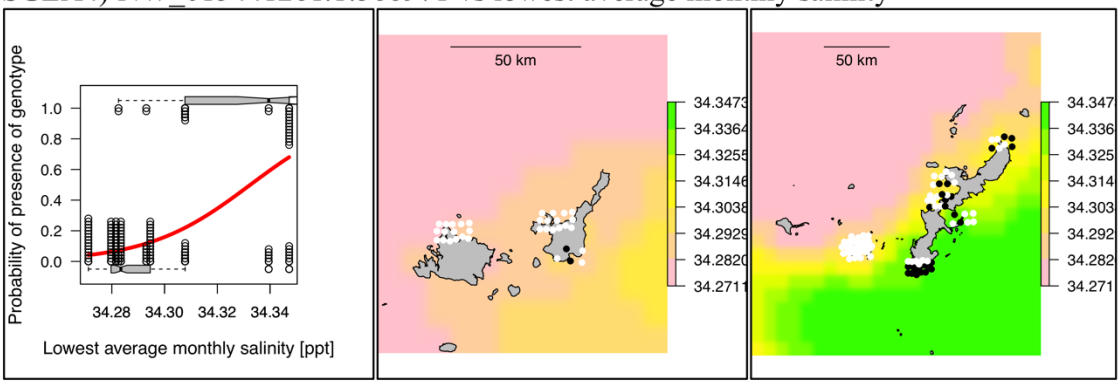
SGEA2) NW_015441080.1:963851 vs bleaching alert frequency



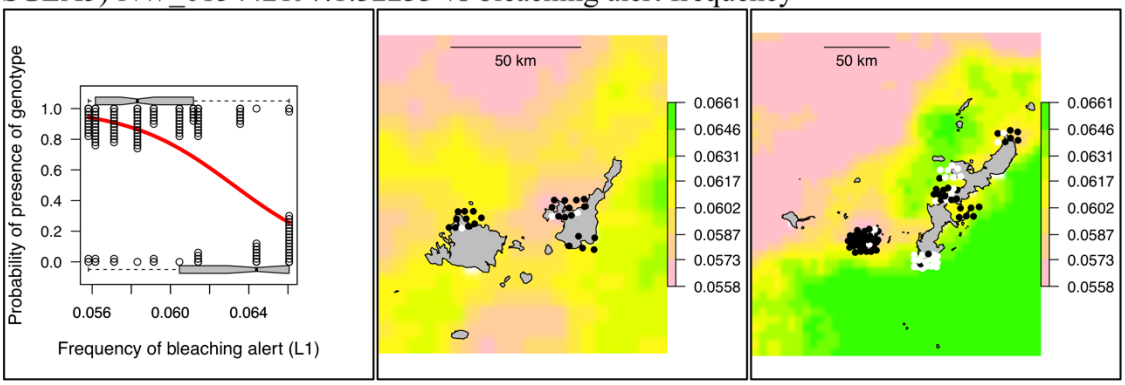
SGEA3) NW_015441121.1:665651 vs bleaching alert frequency



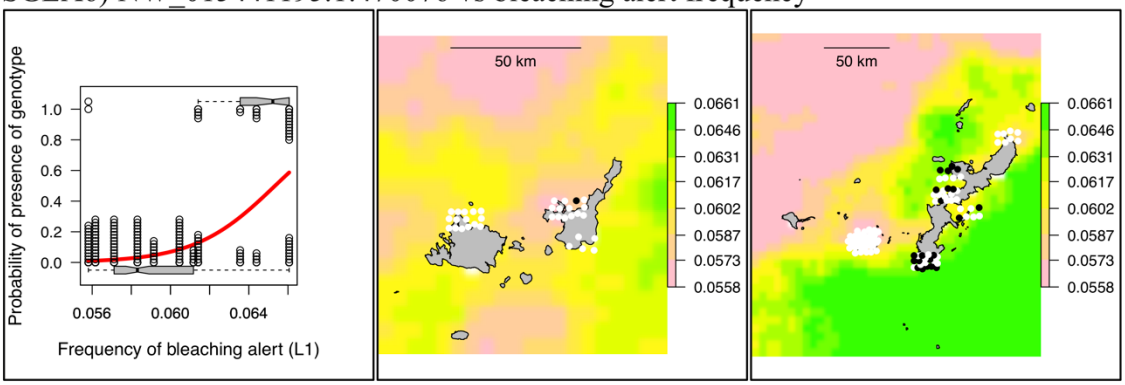
SGEA4) NW_015441261.1:566971 vs lowest average monthly salinity



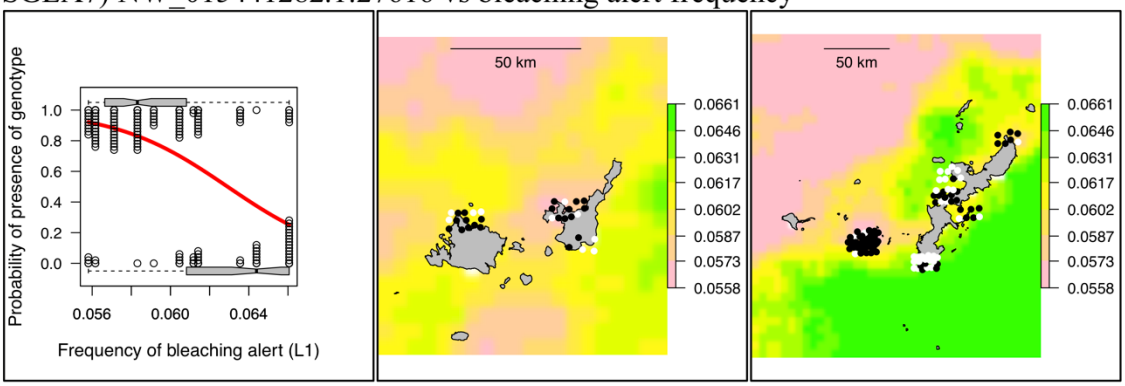
SGEA5) NW_015442197.1:32233 vs bleaching alert frequency



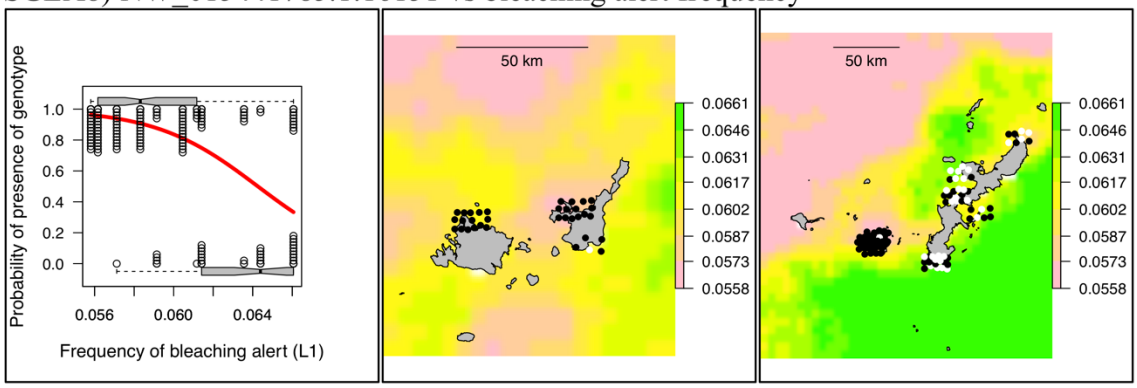
SGEA6) NW_015441195.1:470076 vs bleaching alert frequency



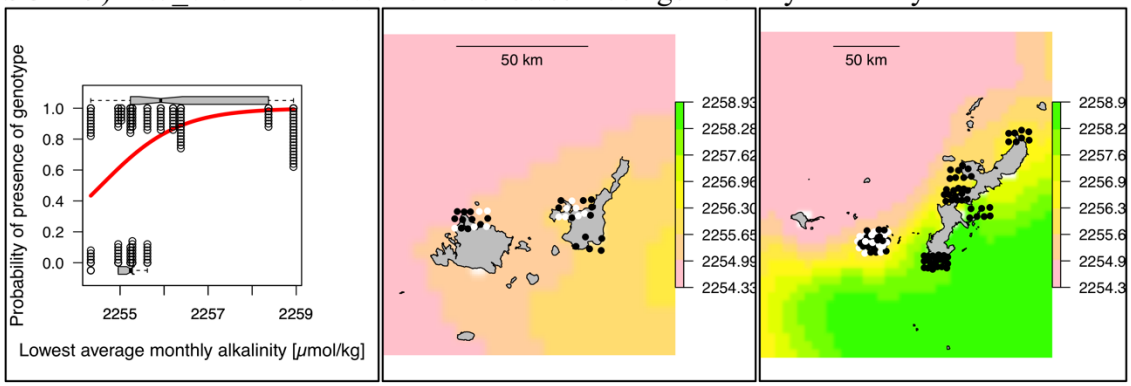
SGEA7) NW_015441282.1:27616 vs bleaching alert frequency



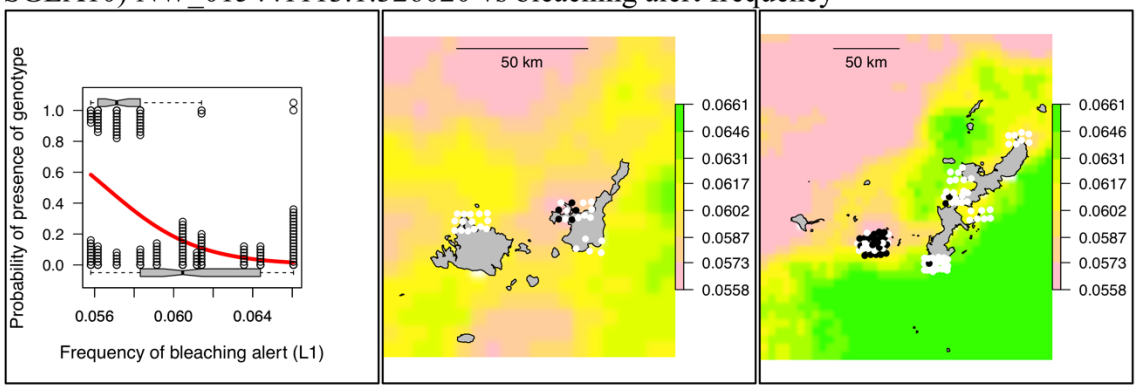
SGEA8) NW_015441785.1:16151 vs bleaching alert frequency



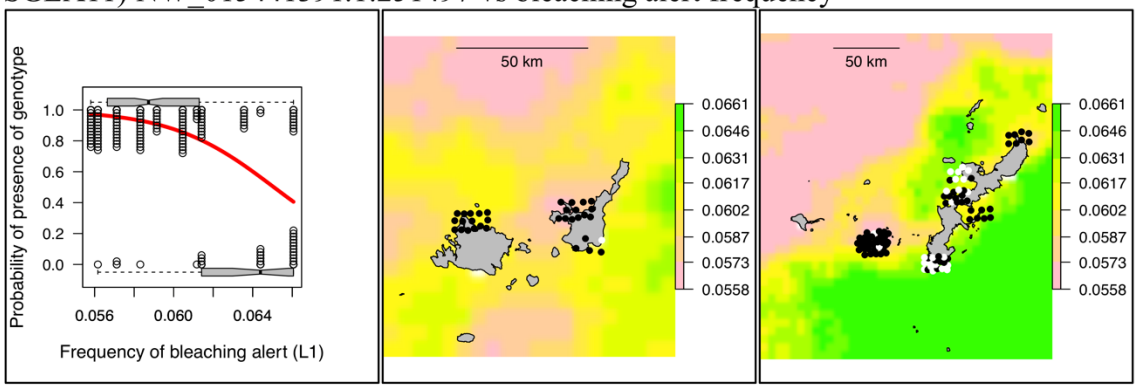
SGEA9) NW_015441192.1:602343 vs lowest average monthly alkalinity



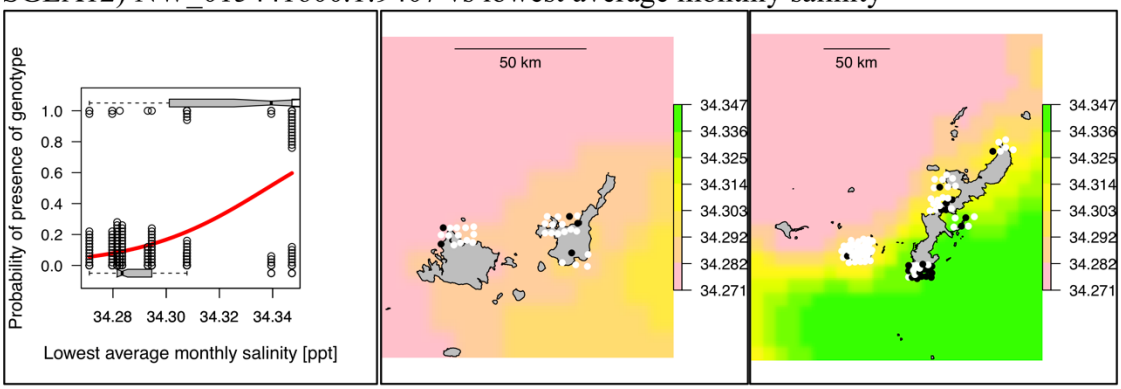
SGEA10) NW_015441113.1:326020 vs bleaching alert frequency



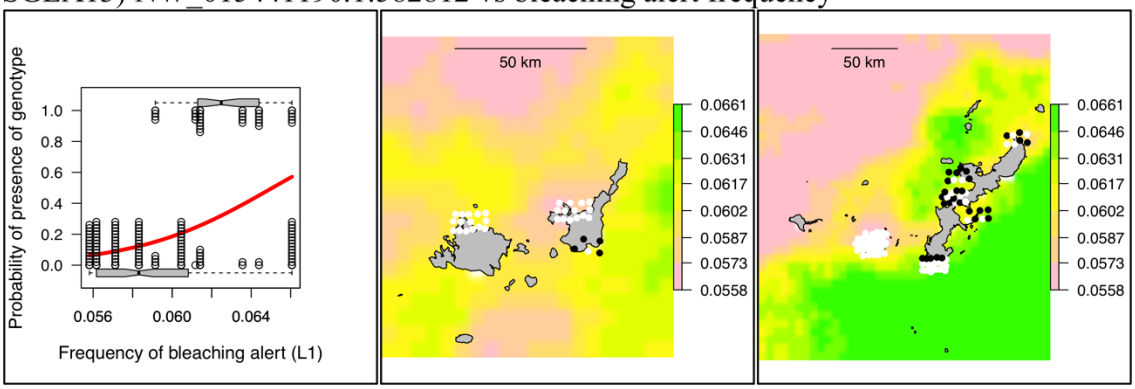
SGEA11) NW_015441391.1:251497 vs bleaching alert frequency



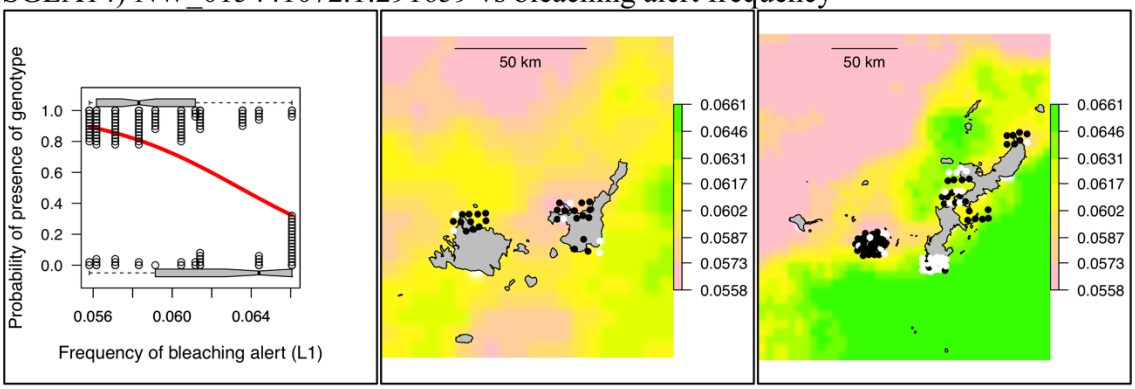
SGEA12) NW_015441600.1:9407 vs lowest average monthly salinity



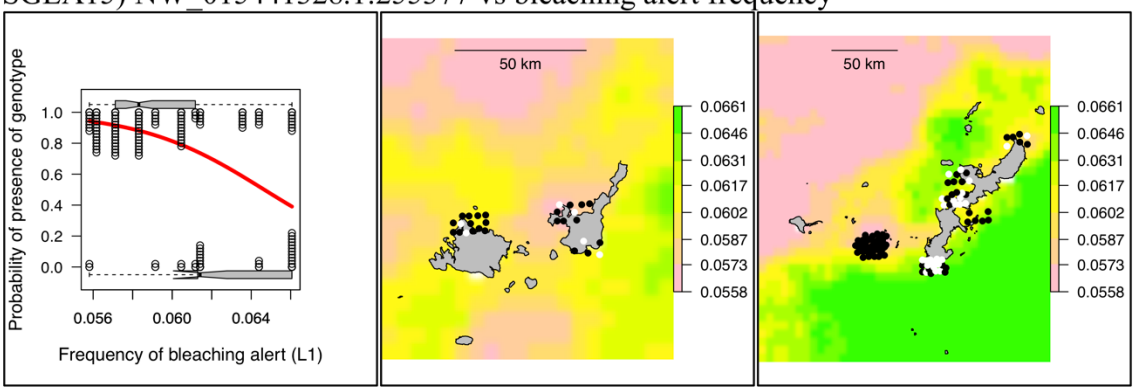
SGEA13) NW_015441190.1:582812 vs bleaching alert frequency



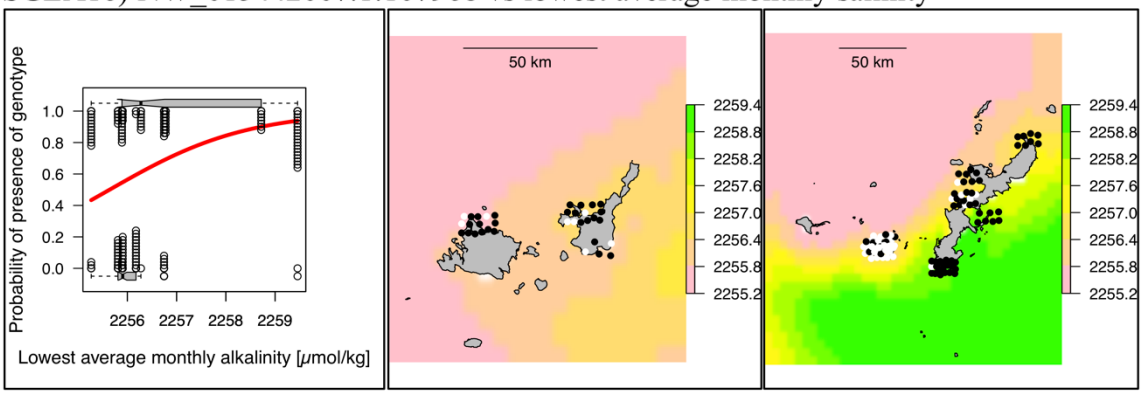
SGEA14) NW_015441072.1:291659 vs bleaching alert frequency



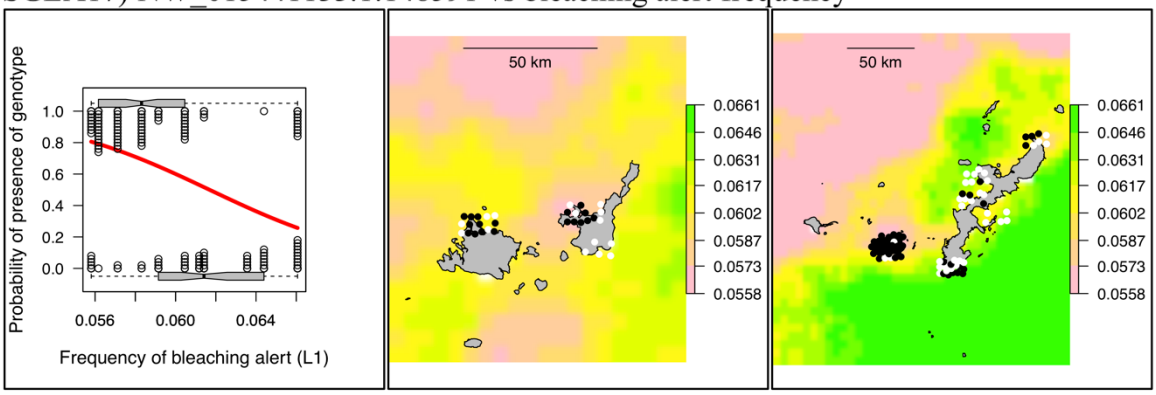
SGEA15) NW_015441328.1:255377 vs bleaching alert frequency



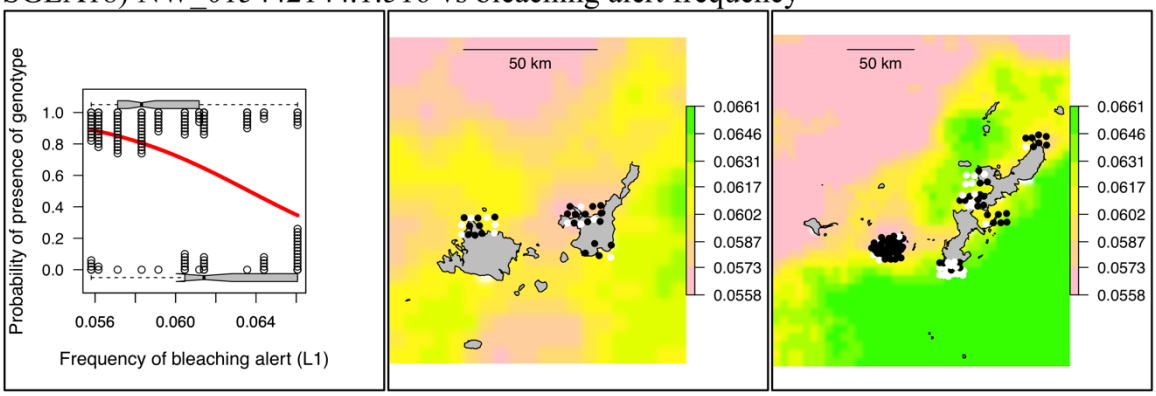
SGEA16) NW_015442007.1:107968 vs lowest average monthly salinity



SGEA17) NW_015441133.1:148591 vs bleaching alert frequency

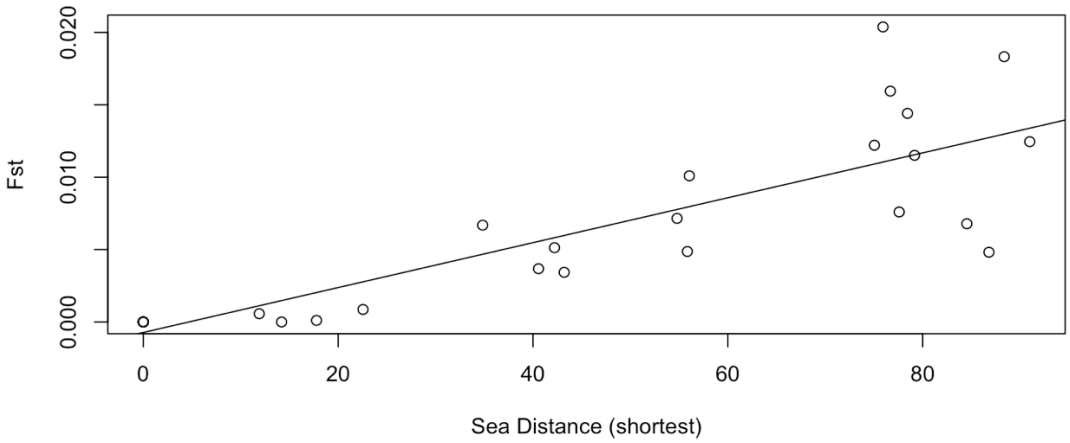


SGEA18) NW_015442144.1:516 vs bleaching alert frequency

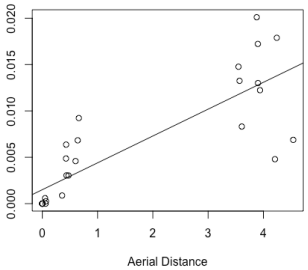


Supplementary Figure 3-4. Connectivity models. The two models describe the relationship between genetic distance (Fst) and geographic distance calculated in two different ways. In a), geographic distance is sea distance and represents the dispersal costs calculated out of sea current data from the whole study period (1993-2010). In b), geographic distance is the aerial distance between sampling sites.

a)
Fst = -7.253e-04+(SD*1.551e-04)
p = 1.45e-09
Multiple R-squared = 0.7155
AIC = -234

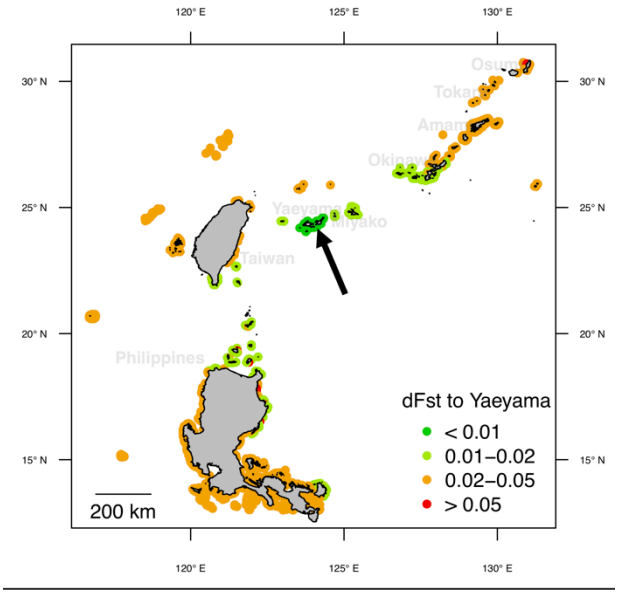


b)
Fst = 0.0016579+(AD*0.0027692)
p = 1.83e-07
Multiple R-squared = 0.66
AIC = -230

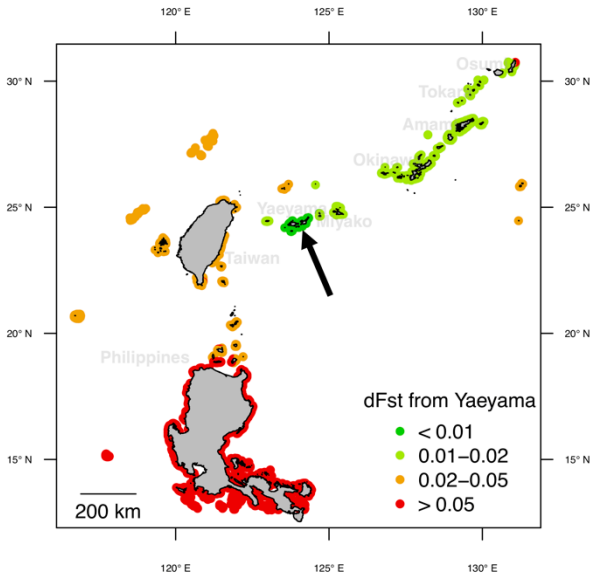


Supplementary Figure 3-5. Example of dFst variation across study area. The two maps show the dFst values connecting all the reefs of the study area to (a) and from (b) the same reef located in the center of Yaeyama islands (marked with an arrow).

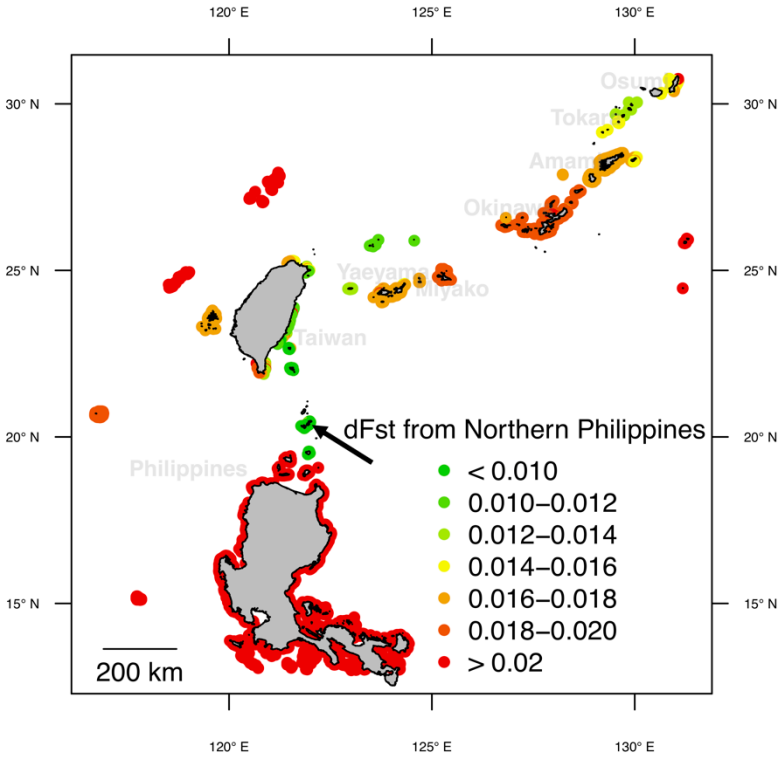
a) dFst to Yaeyama



b) dFst from Yaeyama



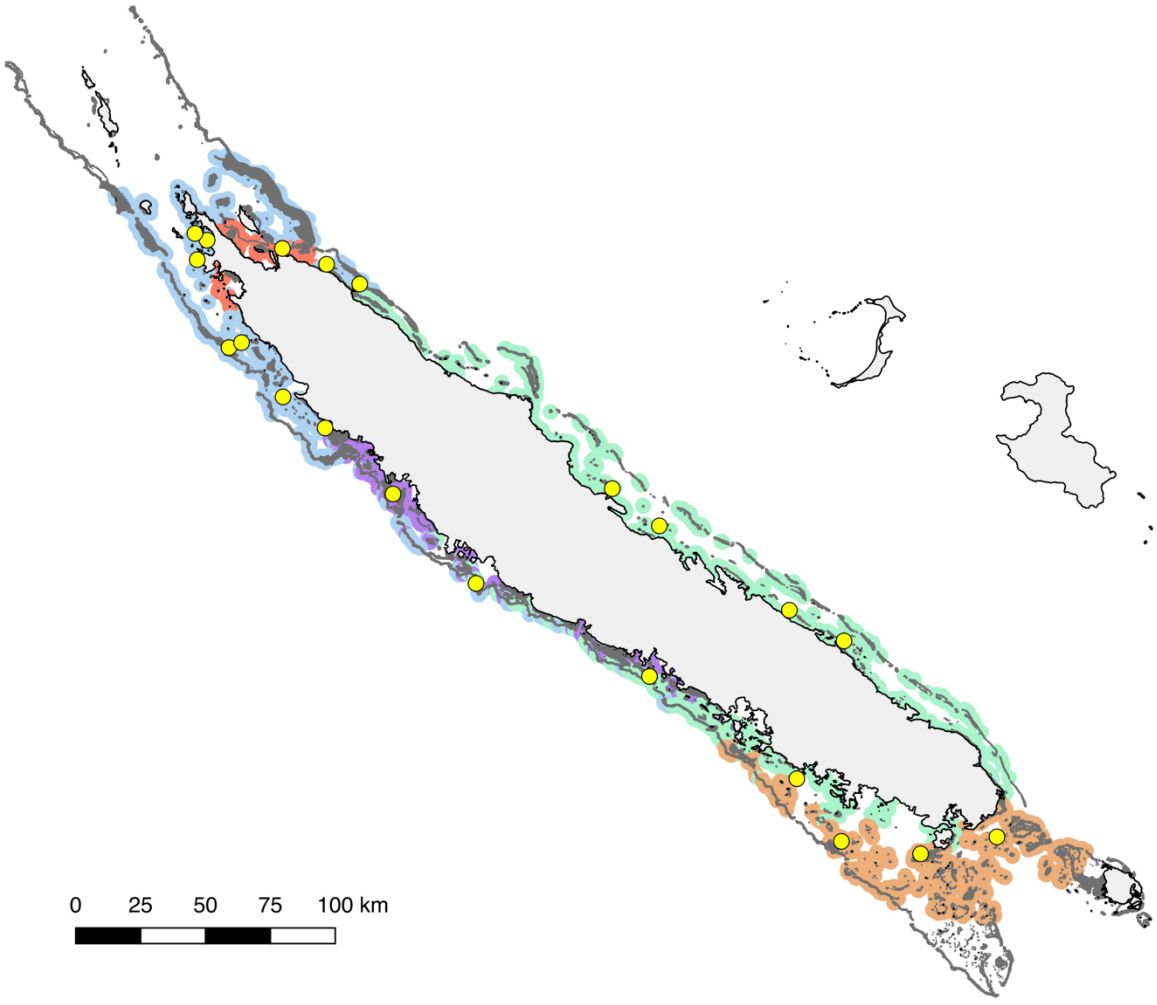
Supplementary Figure 3-6. dFst from Northern Philippines. The two maps show the dFst values connecting one reef in the north of Philippines (marked with an arrow) to all the reefs of the study area.



7.3. Supplementary material from article C

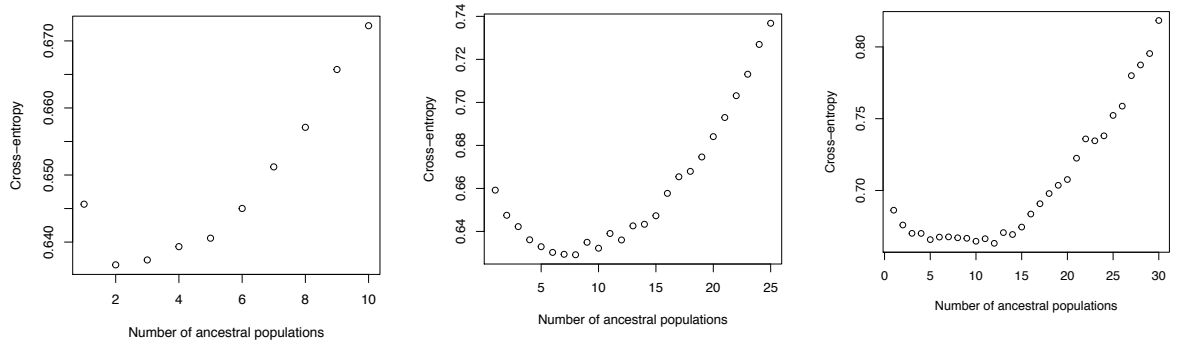
Note: Supplementary Tables 4-2, 4-4 and 4-5 are available as online supplementary material associated with the publication because of size limitations.

Supplementary Figure 4-1. Environmental clusters of reefs of New Caledonia. The reefs around Grande Terre are highlighted in five colors (blue, red, violet, orange and green) corresponding to the five clusters based on environmental variation. Where possible, we sampled corals at four sampling sites (yellow circles) per cluster.



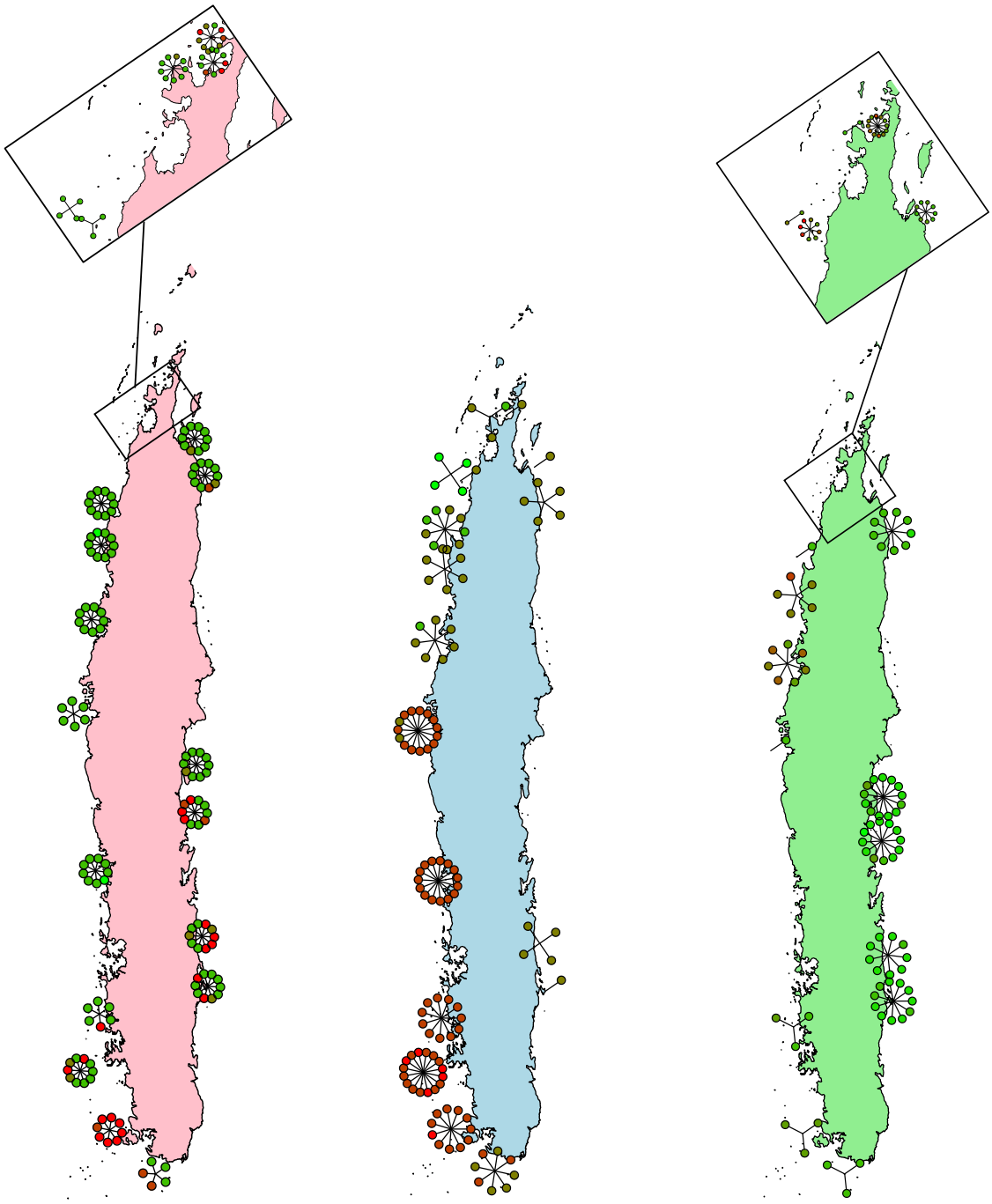
Supplementary Figure 4-2. Cross-entropy comparison for the estimation of number of ancestral populations. The graphs display the comparison of the quality of fit of admixture coefficients for different number of ancestral populations for the three species of interest. Lower cross-entropy criterion indicates a higher quality of fit.

a) *A. millepora* b) *P. damicornis* c) *P. acuta*

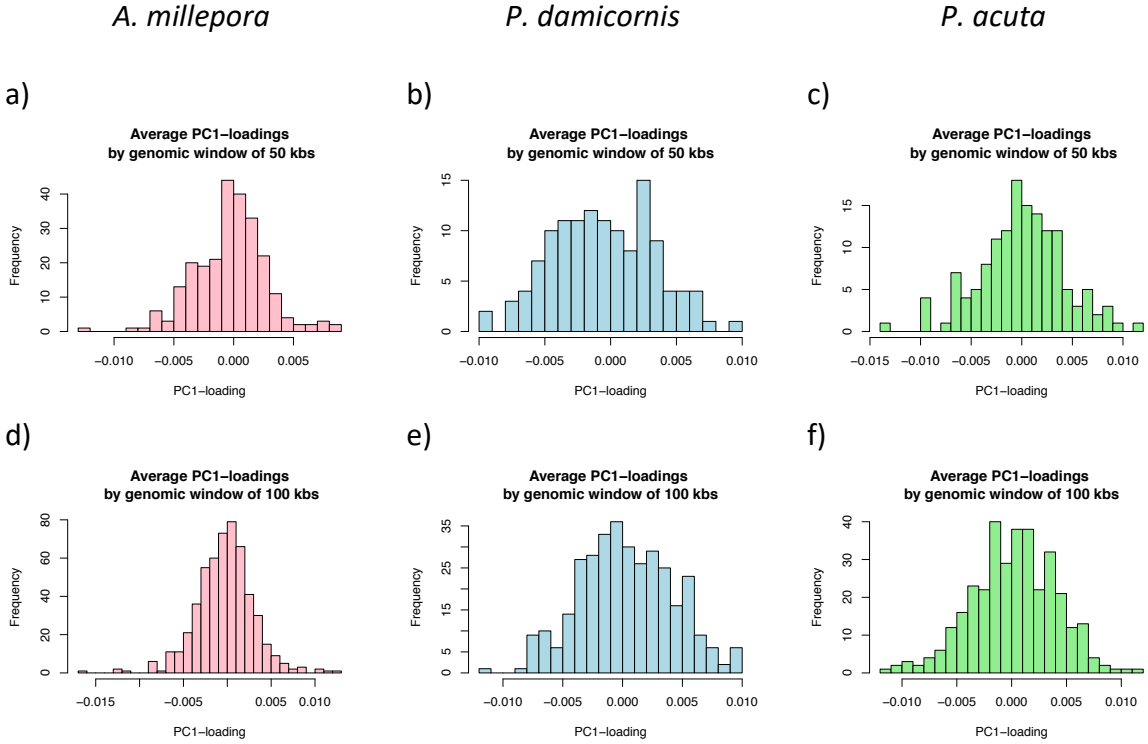


Supplementary Figure 4-3. Spatial distribution of first principal component (PC1) of genotype matrix. The PC1 values for the three studied species are represented on a scale from red (low value) to green (high value). Every point corresponds to an individual an its respective PC1 value. For illustrative reason, individuals are radially distributed around the sampling locations.

a) *A. millepora* b) *P. damicornis* c) *P. acuta*



Supplementary Figure 4-4. SNPs specific PC1-loadings averaged by genomic windows. PC1-loadings for every SNP were averaged across genomics windows of 50 kbs (a, b, c) and 100 kbs (d, e, f). The histograms display the distributions of these averages for the three studied species.



Supplementary Table 4-1. Environmental Variables included in the Seascape Genomics analysis. For each remote sensing product, the table shows the sources (CMEMS= Copernicus Marine Environment Monitoring System; OceanColor= OceanColor) and the corresponding identifier (Product Name). Temporal range and resolution indicate the duration and the frequency of the remote sensing period, respectively. The variables calculated for the seascape genomics analysis are listed in the Derived Variables column with the following abbreviations: OM (overall mean), OMsd (overall standard deviation), HM (highest monthly mean), LM (lowest monthly mean), HMsd (standard deviation of the month with highest mean), LMsd (standard deviation of month with lowest mean).

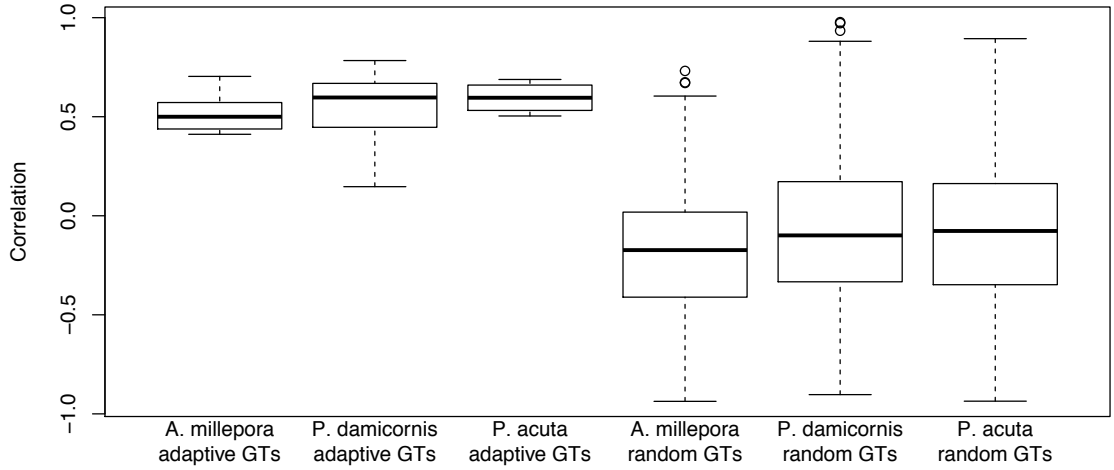
Variable Name	Product Name	Source	Temporal Range	Temporal Resolution	Spatial Resolution	Derived Variables
Sea Surface Temperature	ESACCI-GLO-SST-L4-REP-OBS-SST	CMEMS	1981-2016	Daily	5 km	OM, OMsd, HM, LM, HMsd, LMsd
Sea Surface Temperature (High spatial resolution)	GHRSST	NASA Earth Data	2002-2017	Daily	1 km	OM, OMsd, HM, LM, HMsd, LMsd
Sea Surface Salinity	GLOBAL_REANALYSIS_PHY_001_030	CMEMS	1993-2017	Daily	0.083 °	OM, OMsd, HM, LM, HMsd, LMsd
Chlorophyll Concentration	OCEANCOLOUR_GLO_CHL_L4_REP_OBSERVATIONS_009_082	CMEMS	1992-2017	Daily	4 km	OM, OMsd, HM, LM, HMsd, LMsd
Sea Currents Velocity	global-reanalysis-phy-001-030	CMEMS	1993-2017	Daily	0.083 °	OM, OMsd, HM, LM, HMsd, LMsd
Suspended Particulate Matter	oc-glo-opt-multi-l4-spm	CMEMS	1997-2016	Monthly	4 km	OM, HM, LM
Bleaching-alert-temperature (5 km)	Derived from SST (5 km)		1981-2016	Daily	5 km	Frequency, 4 versions
Bleaching-alert-temperature (1 km)	Derived from SST (1 km)		2002-2017	Daily	1 km	Frequency, 4 versions
Alkalinity	Derived from SST + pH		1993-2016	Daily	0.05 °	OM, OMsd, HM, LM, HMsd, LMsd

Supplementary Table 4-3. Significant genotype-environment associations. (detailed) The table displays the number of SNPs significantly associated to a given environmental variable for each of the three studied species. (BAF*= bleaching alert frequency)

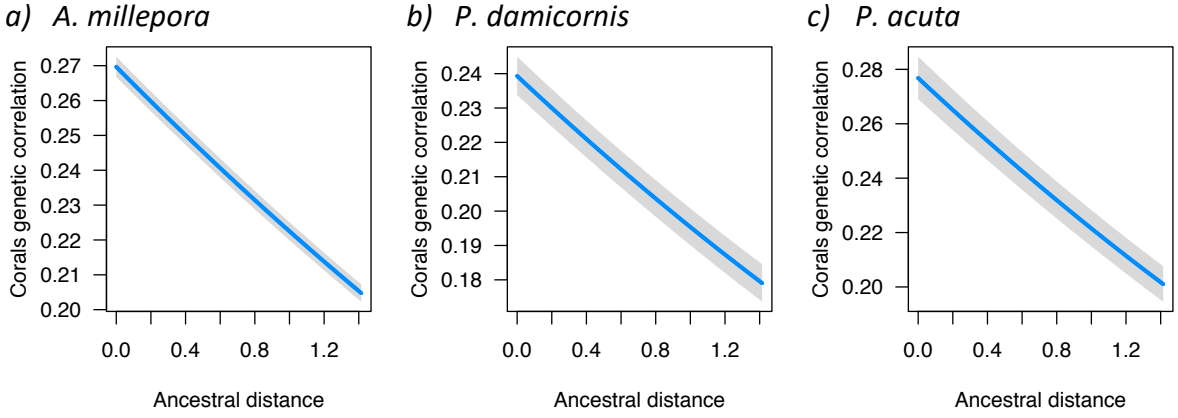
		A. millepora		P. damicornis		P. acuta	
SST	Average warmest month (5 km)	1	63	1	47	0	43
	Average coldest month (5 km)	1		1		1	
	Overall average (5 km)	1		0		0	
	Overall standard deviation (5 km)	1		2		1	
	Standard deviation hottest month (5 km)	0		0		0	
	Standard deviation coldest month (5 km)	11		1		7	
	$BAF_{0°C}$ (5 km)	2		1		6	
	$BAF_{4°C}$ (5 km)	3		9		1	
	$BAF_{8°C}$ (5 km)	1		1		0	
	BAF_{CRW} (5 km)	2		3		0	
	Average warmest month (1km)	5		2		4	
	Average coldest month (1km)	12		4		3	
	Overall average (1km)	0		1		1	
	Overall standard deviation (1 km)	4		3		2	
	Standard deviation hottest month (1 km)	2		1		0	
	Standard deviation coldest month (1 km)	2		4		1	
	$BAF_{0°C}$ (1 km)	10		8		8	
	$BAF_{4°C}$ (1 km)	3		3		3	
	$BAF_{8°C}$ (1 km)	2		1		1	
	BAF_{CRW} (1 km)	0		1		4	
Alkalinity	Average highest month	1	11	0	7	0	14
	Average lowest month	0		0		0	
	Overall average	0		1		0	
	Overall standard deviation	2		0		0	
	Standard deviation highest month	0		1		0	
	Standard deviation lowest month	8		5		14	
Chlorophyll concentration	Average highest month	5	15	4	22	0	27
	Average lowest month	1		4		8	
	Overall average	4		5		10	
	Overall standard deviation	1		3		4	
	Standard deviation highest month	3		4		5	
	Standard deviation lowest month	1		2		0	
Sea current velocity	Average highest month	0	4	0	4	1	2
	Average lowest month	2		0		1	
	Overall average	0		3		0	
	Overall standard deviation	2		0		0	
	Standard deviation highest month	0		0		0	
	Standard deviation lowest month	0		1		0	
Suspended particulate matter	Average highest month	5	8	0	1	1	2
	Average lowest month	2		0		0	
	Overall average	1		1		1	
Salinity	Average highest month	5	19	0	9	0	12
	Average lowest month	0		0		0	
	Overall average	0		0		0	
	Overall standard deviation	6		6		0	
	Standard deviation highest month	0		0		2	
	Standard deviation lowest month	8		3		10	

7.4. Supplementary material from article D

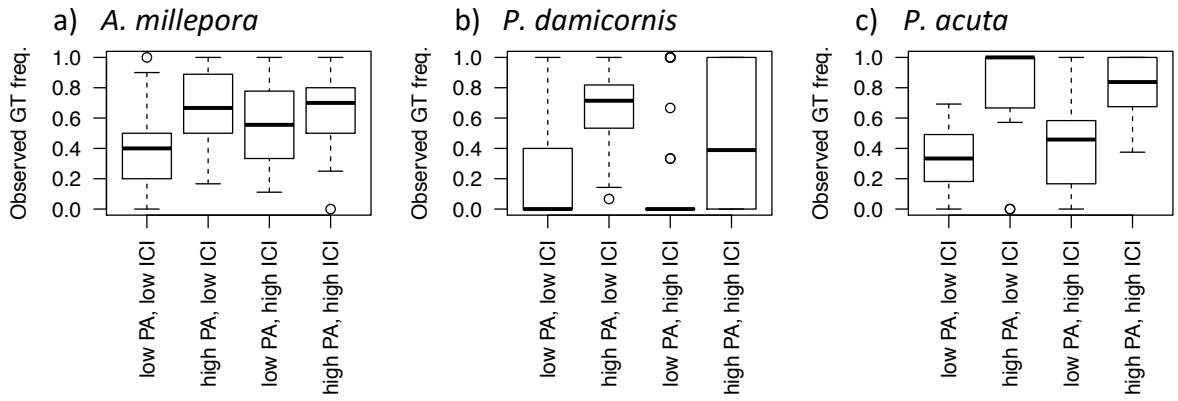
Supplementary Figure 4-1. Predictive accuracy evaluation of genotype-environment association models. For every potentially adaptive genotype against heat uncovered in a previous seascape genomics study (Selmoni et al., 2020), the resulting association model was cross-validated using a "leave-one-population-out" approach. The plots display the correlation between the expected and observed genotype frequencies. In the three boxplots on the left, the predictive accuracy is evaluated for potentially adaptive genotype-environment association models of the three studied species (*Acropora millepora, Pocillopora damicornis and Pocillopora acuta*). The three boxplots on the right display the results of the same cross-validation approach on 1000 random genotypes from the genome of the three species.



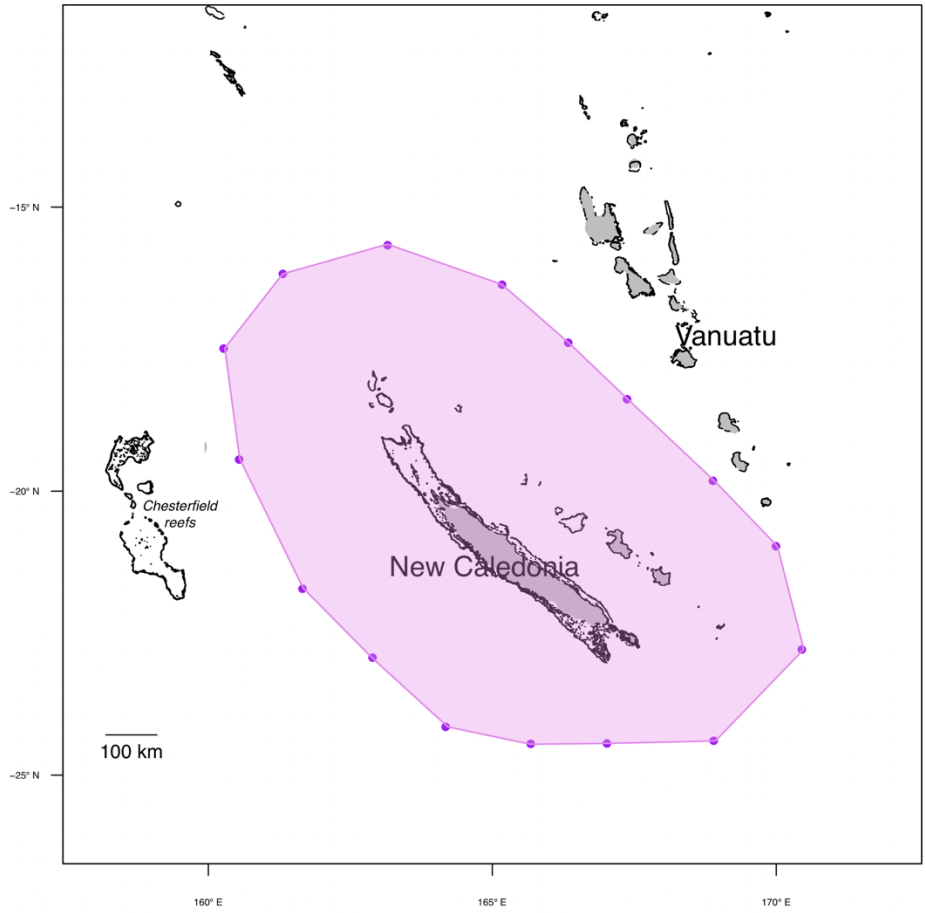
Supplementary Figure 4-2. Ancestral distance and genetic correlation between corals. The three plots display genetic correlations between pairs of corals sampled in New Caledonia as a function of their ancestral distance (blue line, with the grey band showing the 95% interval of confidence). Genetic correlations were computed as the correlation of single-nucleotide-polymorphisms, while ancestral distance is the difference in admixture from the ancestral populations. Each plot displays this association for a different species (a: *Acropora millepora*, b: *Pocillopora damicornis*, c: *Pocillopora acuta*). For each of the three species, the plots display a strong association between genetic correlation between corals and their ancestral distances.



Supplementary Figure 4-3. Observed frequency of adaptive genotypes to heat. The three boxplots display, for each of the three studied species (a: *Acropora millepora*, b: *Pocillopora damicornis*, c: *Pocillopora acuta*), the observed frequency of genotypes putatively adaptive to heat stress (y-axis) at reefs predicted with different values of probability of heat stress adaptation (low PA:PA<0.6, high PA: PA≥0.6) and inbound connectivity index (low ICI: ICI<1500 km^2, high ICI: ICI≥1500 km^2).



Supplementary Figure 4-4. Boundaries of the study area. The purple area displays the boundaries of the study area, located approximately 250 km around the most peripheral reefs of New Caledonia (excluding Chesterfield Reefs, located in the western part of the archipelago).



7.5. The ManaCo consortium

Using modern conservation tools for innovative management of coral reefs: the MANACO Consortium

Oliver Selmoni[1,2], Gaël Lecellier[2,3], Lara Ainley[4], Antoine Collin[5], Raimana Doucet[6], Vaimiti Dubousquet[7], Hudson Feremaito[8], Edouard Ito Waia[9], Stuart Kininmonth[10], Hélène Magalon[11], Siola'a Malimali[12], Ateliana Maugateau[13], Anders Meibom[14,15], Stephen Mosese[16], Malika René-Trouillefou,[17] Noriyuki Satoh[18], Madeleine JH van Oppen[19,20], André Xozamé[21], Maxime Yékawene[9], Stéphane Joost[1†] and Véronique Berteaux-Lecellier[2†]

[1]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[2]UMR250/9220 ENTROPIE IRD-CNRS-Ifremer-UNC-UR, Labex CORAIL, Noumea, New Caledonia
[3]UVSQ, Université de Paris-Saclay, Versailles, France
[4]Ministry of Marine Resources, Rarotonga, Cook Islands
[5]Centre de Géoécologie Littorale, EPHE-PSL Université Paris, CNRS LETG, Labex CORAIL, Dinard, France
[6]Direction de l'environnement de Polynésie française, Papeete, French Polynesia
[7]Délégation à la recherche, Government of French polynesia, Papeete, French Polynesia
[8]Vanuatu Fisheries Department, Port Vila, Vanuatu
[9]Collectif Bwé-Dï, Nengone, Maré, New Caledonia
[10]School of Marine Studies, The University of the South Pacific, Suva, Fiji
[11]UMR250/9220 ENTROPIE IRD-CNRS-Ifremer-UNC-UR, Labex CORAIL, Université de la Réunion, St Denis, France
[12]Ministry of Fisheries, Tonga
[13]Service territorial de l'Environnement, Wallis, Wallis et Futuna
[14]School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[15]Center for Advanced Surface Analysis, Institute of Earth Sciences, Université de Lausanne, CH-1015 Lausanne, Switzerland
[16]Ministry of Fisheries and Marine Resources, Honiara, Solomon Islands
[17]Laboratoire de Biologie des Organismes et Ecosystèmes Aquatiques (BOREA), Labex Corail, Université des Antilles, MNHN, SU, UA, CNRS, IRD, Pointe-à-Pitre Cedex, Guadeloupe
[18]Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate School, Okinawa, Japan
[19] School of BioSciences, The University of Melbourne, Victoria, Australia
[20]Australian Institute of Marine Science, Townsville MC, Queensland, Australia
[21]Collectif Meketrepune, Drehu, Lifou, New Caledonia
[†]These authors have contributed equally to this work

Postprint version of the article published in *Frontiers in Marine Science:*
Selmoni, O. et. al (2020). Using modern conservation tools for innovative management of coral reefs: the MANACO Consortium. Frontiers in Marine Science, doi: 10.3389/fmars.2020.00609.

Abstract

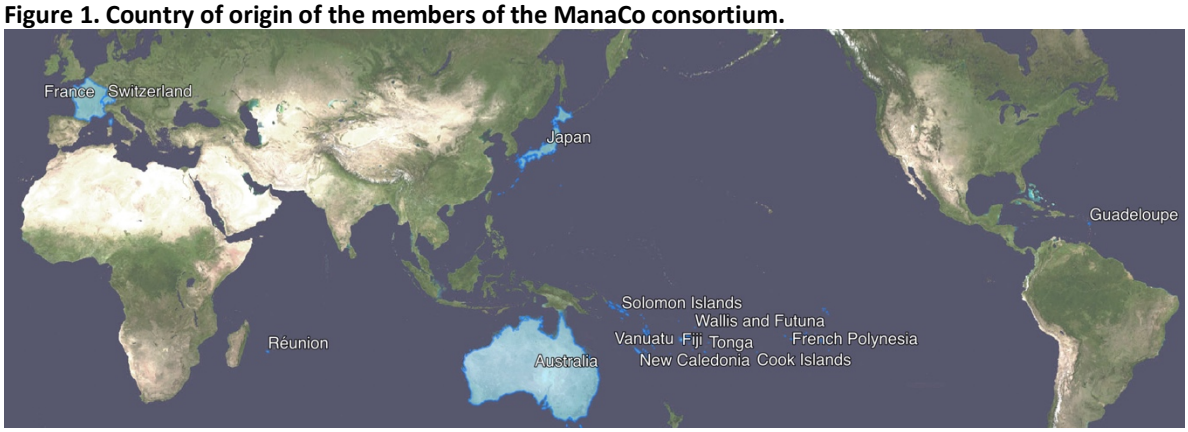
Coral reefs are under threat and innovative management strategies are urgently required. However, discoveries from innovative fields of coral reef research are rarely transposed in

practical conservation actions. This is mainly due to the difficulties in knowledge exchange between scientists and conservation stakeholders.
The ManaCo consortium (http://manaco.ird.nc/) is an international network federating conservation stakeholders and researchers in a common effort to preserve the coral reefs. The focus is on using modern tools to build a bridge between indigenous knowledge and scientific innovation.
ManaCo aims to orientate research towards relevant conservation needs and to facilitate the transposition of research into concrete management strategies. This will allow to coordinate a collaborative response against coral reef decline. We invite anyone sharing the same interests in joining us.

Introduction

Over the last decades, coral reefs have suffered a major decline due to the degradation of water quality, the increase of disease and predation, and the rise of sea surface temperatures (Ateweberhan et al., 2013; De'ath et al., 2012; Wilkinson, 2008). In the most severe cases, anomalous heat waves have already caused local coral losses of up to 50% (Hughes et al., 2017). Climatic projections predict stressful environmental conditions to become more frequent in the years to come, jeopardizing the future of coral reefs (van Hooidonk et al., 2016).
Coral reef conservation requires innovation in management strategies to cope with these threats (Mumby & Steneck, 2008). Recent research innovation at the intersection of genetics, oceanography, remote sensing and computer science can provide valuable insights to reinforce conservation strategies (Beger et al., 2014; Magris et al., 2014; Maina et al., 2011; van Oppen et al., 2017). However, the transposition of these scientific developments into a conservation perspective is hindered for several reasons. First, there is insufficient training of decision makers in the aforementioned disciplines which in some cases (*e.g.* genetics) has nourished skepticism toward the scientific process (Frankham, 2010; Joost et al., 2011). Second, the main format for dissemination of research findings is scientific publications, which are often technical and difficult to interpret for a non-specialist reader (Bainbridge, 2014). Literature search is also a time-consuming task, and often articles may not be open access (Gossa, Fisher, & Milner-Gulland, 2015). The use of software support (*e.g.* web applications) synthesizing information relevant for conservation has been advocated to fill these gaps (Hoban et al., 2013; Westgate et al., 2018). Another obstacle is often the difficulty of research projects in adequately representing the conservation situations for which they should provide solutions. This is due to the unbalance in scientific output between countries, that results in works covering spatial and temporal scales that are not necessarily relevant for other conservation contexts (Bainbridge, 2014; Rose et al., 2018). Furthermore, new guidelines for conservation should fit in with already established frameworks and acknowledge the importance of pre-existing preservation criteria (*e.g.* social importance, traditional knowledge; Roux et al., 2006). To do so, it is essential that the channel of communication between scientists, conservation managers and policy makers facilitates a multidirectional knowledge exchange (Bainbridge, 2014; Roux et al., 2006). The more frequent and sustained these exchanges, the greater mutual trust that develops (Roux et al., 2006). Last, lack of funding can be an obstacle to the transition from research to practical

Figure 1. Country of origin of the members of the ManaCo consortium.



conservation actions (Knight et al., 2008). International collaboration can help mitigate this issue by boosting the application for common funds.

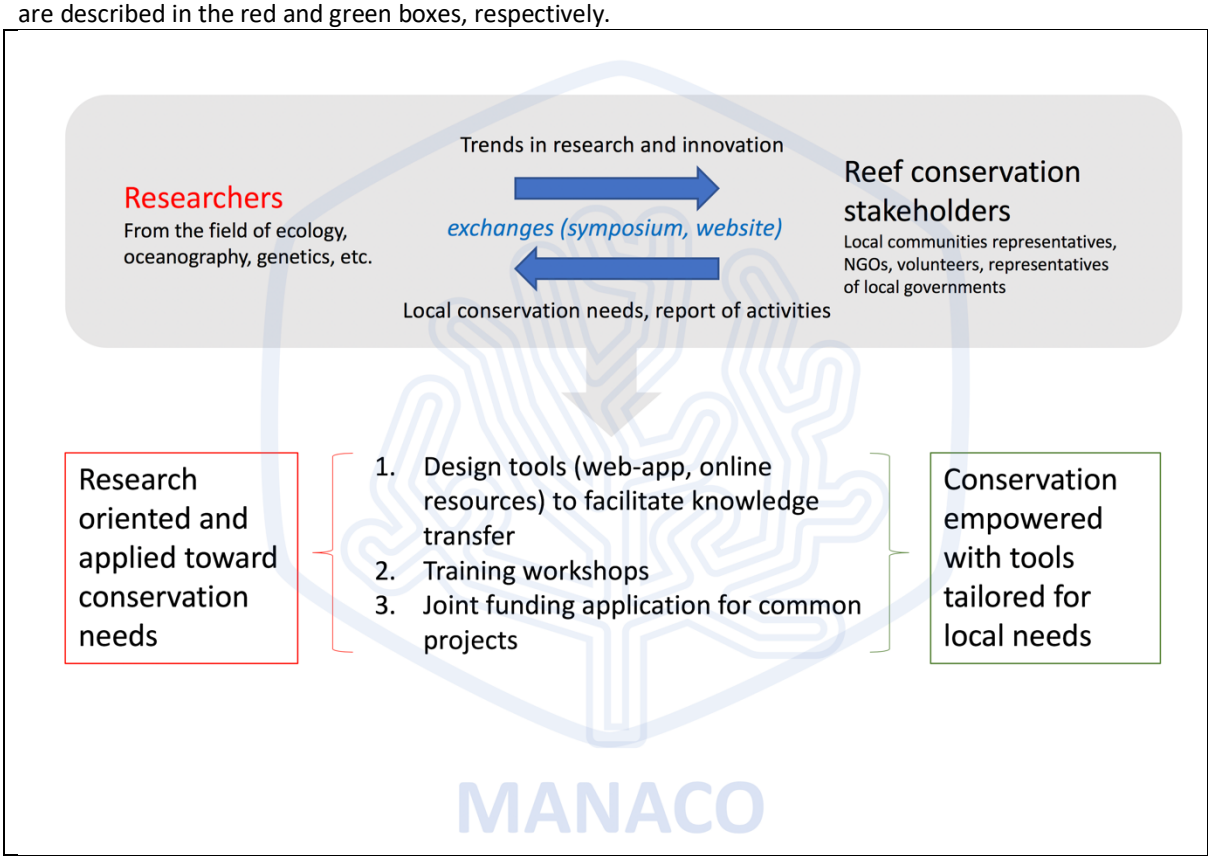
The ManaCo consortium

In December 2019, a symposium organized in Noumea (New Caledonia) gathered about 60 participants to discuss innovative approaches to reinforce coral reef conservation strategies. This meeting brought together scientists and reef conservation stakeholders from 13 countries distributed across the South and North Pacific, Caribbean, Indian Ocean and Europe (Fig. 1). Reef preservation stakeholders reported the state of their local conservation activities, while coral researchers presented the highlights of their research work. By means of a questionnaire, attendees provided their opinions concerning the main obstacles encountered in the exchange between research and applied conservation. The most frequent issues mentioned were about the lack of clarity in communication, the unbalanced representativeness of scientific studies and social/economic pressures. The answers to the questionnaires prepared the ground for a round table discussion, where the need for joining forces was unquestioned by all attendees. This led to the signature of a letter of intent for the creation of the ManaCo consortium.
ManaCo stands for “Modern tools for innovative coral MANAgement and COnservation” (http://manaco.ird.nc/) and is an international group that federates local communities, volunteers and stakeholders with researchers in a common effort to preserve the coral reefs. Its focus will be on using modern and easy-to-use tools to build a bridge between indigenous/local knowledge and scientific innovation, in particular to promote exchange and sharing of technical expertise, knowledge and resources between reef conservation stakeholders and researchers.

The plan of action

The plan of the ManaCo consortium activities consists of two recurring phases (Fig. 2). The first phase is the cross-talk between members to highlight trends and innovations in coral reef research and to point out critical issues in reef conservation. These discussions will take place

Figure 2. Schematic view of the ManaCo plan of action. The plan of action is composed of two phases. During the first phase (grey box), the discussion between reef conservation stakeholders and researchers aims to identify a specific research topic addressing a precise conservation issue. In the second phase (under the grey box), a strategy is established to facilitate the inherent transfer of knowledge. The instruments used in this phase are the design of tools using recent technologies (web-app, online resources), training workshops or the application for joint funding. The expected benefits for researchers and reef conservation stakeholders are described in the red and green boxes, respectively.



during bi-annual face-to-face meetings and via virtual consultations coordinated through the consortium website. Based on these exchanges, members of the consortium will identify a specific research topic addressing a precise conservation issue, as well as the associated obstacles in the transfer of knowledge.

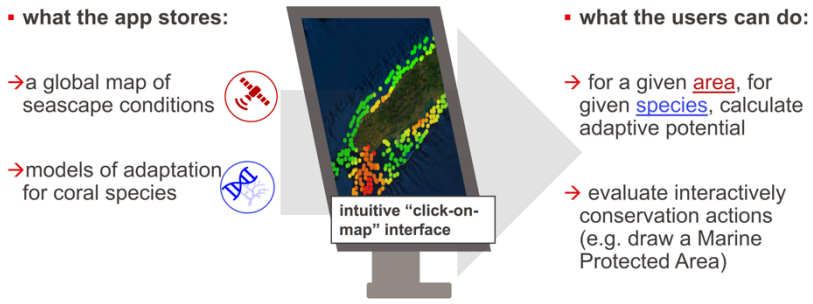
The second phase aims to overcome these obstacles. Three complementary instruments are proposed:

1) Development of tools using recent technologies: web-applications or online resources conceived to facilitate the access to scientific information relevant for specific conservation needs (*e.g.* see http://www.congressgenetics.eu/). The design of these tools should focus on developing intuitive platforms, self-empowering and open access (example in box. 1);
2) Training workshops: these events will be organized to promote knowledge transfer requiring dynamic exchanges. This can involve a wide range of topics, from the standardization of field procedures to training courses on specific software solutions;
3) Funding applications for common projects: members will team up to define common projects promoting innovation in reef conservation. Joint responses to calls for proposal will be performed through the consortium, leveraging the interdisciplinarity and internationality of the ManaCo team.

Expected benefits

The plan of action of the ManaCo consortium is expected to bring mutual benefits for all the actors involved in coral reef conservation and research. From a researcher point of view, the ManaCo activities will facilitate application of research work in concrete conservation actions. This collaboration will also provide valuable feedback from the field that will improve the efficiency of the conservation measures proposed as well as the scientific knowledge. In addition, the participation in round tables gathering different stakeholders of reef conservation will promote the uptake of research work among the general public.
On the other hand, reef conservation stakeholders will take advantage of a proactive context to discuss specific conservation needs. Insights for conservation will be more relevant for local requirements, and be available in an accessible format. Furthermore, the ManaCo tools will allow reef conservation stakeholders to personally define and customize preservation guidelines and therefore ensure the compatibility with traditional methods.

Box 1. The digital reef adaptive potential evaluator (DRAPEAU). DRAPEAU is an example of a modern tool that will be developed in the frame of ManaCo. The round tables at the ManaCo symposium highlighted the difficulty in interpreting genetic studies on coral adaptation in a conservation perspective. The goal of DRAPEAU is to make such information easily accessible to coral reef stakeholders working on any reef system. The foundations of this tool lay in the observation that reefs recurrently exposed to stressful conditions develop resistance. Seascape genomics predicts the potential for adaptation to environmental conditions measured by remote sensing (e.g. sea water temperature variations, pH, turbidity levels, proximity to populated areas; Selmoni, Rochat, Lecellier, et al., 2020). DRAPEAU will allow to explore these predictions trough an intuitive "click-on-map" interface. The user can customize the calculations for a given area of interest and for a given set of species. An interactive mode will also allow to evaluate the predicted impact of the location of marine protected areas on the other reef areas. An online tutorial will allow the users to familiarize with the functionalities of the app. All the predictions are exportable in various formats, therefore ensuring compatibility with pre-existing conservation frameworks. A video demonstration of DRAPEAU is available in the supplementary material (online version).



Open membership

The expertise and the roles of the members of the consortium transcend the fields of coral conservation and research: we can boast representatives of local communities, volunteers, decision makers of marine conservation, coral reef ecologists, physiologists, geneticists, oceanographers, etc. These actors join the ManaCo consortium from all around the world,

from small islands territories to large countries (Fig. 1). To date, ManaCo team include members from Australia, Cook Islands, Fiji, France, Guadeloupe, Japan, La Réunion, New Caledonia, French Polynesia, Solomon Islands, Switzerland, Tonga, Vanuatu and Wallis and Futuna.
ManaCo membership is open to any coral reef actor, and can be requested directly on the ManaCo website (http://manaco.ird.nc/).

Conclusion

The future of coral reefs is under threat and the need for innovative solutions is echoing worldwide. This global challenge can only be tackled by collaborative responses with solutions adapted to the needs and peculiarities of local contexts, such as the Transnational Red Sea Center (TRSC; Kleinhaus et al., 2020), the Southeast Florida Coral Reef Initiative (SEFCRI; https://floridadep.gov/CoralReefs), the Coral Triangle Initiative (CTI; https://www.conservation.org/projects/coral-triangle-initiative) or the Coral Reef Alliance (https://coral.org/). Like the ManaCo consortium, these networks coordinate collaborative efforts, and they promote the exchanges between actors involved in coral reef conservation and research. A desirable step in the future is to establish a link between these different networks, and it is with this in mind that a representative of the TRSC was present in Noumea to participate in the Manaco Workshop.

Acknowledgements

We are grateful to the following partners (alphabetical order) who contributed to ManaCo meeting and thus to the birth of the consortium: Ambassade de France en Australie, Centre National de la Recherche Scientifique, Consortium pour la Recherche l'Enseignement Supérieur et l'Innovation de Nouvelle-CAlédonie, Ecole Polytechnique Fédérale de Lausanne, Institut de Recherche pour le Développement, International Coral Reef Initiative, LabEx Corail, Pacific Community, United Nations Environment Progam, Université des Antilles, Université de la Réunion.

8. Bibliography

Abebe, T. D., Naz, A. A., & Léon, J. (2015). Landscape genomics reveal signatures of local adaptation in barley (Hordeum vulgare L.). *Frontiers in Plant Science*, *6*(October), 813. https://doi.org/10.3389/fpls.2015.00813

Ahmad, M., Srinivasula, S. M., Wang, L., Talaniaii, R. V., Litwack, G., Fernandes-Alnemri, T., & Alnemri, E. S. (1997). CRADD, a novel human apoptotic adaptor molecule for caspase-2, and FasL/tumor necrosis factor receptor-interacting protein RIP. *Cancer Research*, *57*(4), 615–619.

Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, *22*(13), 1600–1607. https://doi.org/10.1093/bioinformatics/btl140

Andréfouët, S., Cabioch, G., Flamand, B., & Pelletier, B. (2009). A reappraisal of the diversity of geomorphological and genetic processes of New Caledonian coral reefs: A synthesis from optical remote sensing, coring and acoustic multibeam observations. *Coral Reefs*, *28*(3), 691–707. https://doi.org/10.1007/s00338-009-0503-y

Anthony, K. R. N., Kline, D. I., Diaz-Pulido, G., Dove, S., & Hoegh-Guldberg, O. (2008). Ocean acidification causes bleaching and productivity loss in coral reef builders. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(45), 17442–17446. https://doi.org/10.1073/pnas.0804478105

Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., … Voolstra, C. R. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports*, *6*(1), 1–15. https://doi.org/10.1038/srep39734

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Consortium, G. O. (2000). Gene Ontology: Tool for The Unification of Biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Ateweberhan, M., Feary, D. A., Keshavmurthy, S., Chen, A., Schleyer, M. H., & Sheppard, C. R. C. (2013). Climate change impacts on coral reefs: Synergies with local effects, possibilities for acclimation, and management implications. *Marine Pollution Bulletin*, *74*(2), 526–539. https://doi.org/10.1016/j.marpolbul.2013.06.011

Ayre, D. J., & Hughes, T. P. (2000). Genotypic diversity and gene flow in brooding and spawning corals along the great barrier reef, Australia. *Evolution*, *54*(5), 1590–1605. https://doi.org/10.1111/j.0014-3820.2000.tb00704.x

Bainbridge, I. (2014). PRACTITIONER'S PERSPECTIVE: How can ecologists make conservation policy more evidence based? Ideas and examples from a devolved perspective. *Journal of Applied Ecology*, *51*(5), 1153–1158. https://doi.org/10.1111/1365-2664.12294

Balkenhol, N., Dudaniec, R. Y., Krutovsky, K. V., Johnson, J. S., Cairns, D. M., Segelbacher, G., … Joost, S. (2017). *Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations*. Springer, Cham. https://doi.org/10.1007/13836_2017_2

Ball, I. R., Possingham, H. P., & Watts, M. E. (2009). Marxan and relatives: software for spatial conservation prioritization. *Spatial Conservation Prioritization: Quantitative Methods and Computational Tools*.

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., … Zhang, J. (2015). UniProt: A hub for protein information. *Nucleic Acids Research*, *43*(D1), D204–D212. https://doi.org/10.1093/nar/gku989

Baums, I. B. (2008). A restoration genetics guide for coral reef conservation. *Molecular Ecology*, *17*(12), 2796–2811. https://doi.org/10.1111/j.1365-294X.2008.03787.x

Baums, I. B., Baker, A. C., Davies, S. W., Grottoli, A. G., Kenkel, C. D., Kitchen, S. A., … Shantz, A. A. (2019). Considerations for maximizing the adaptive potential of restored coral populations in the western Atlantic. *Ecological Applications*, *29*(8). https://doi.org/10.1002/eap.1978

Bay, R. A., & Palumbi, S. R. (2014). Multilocus adaptation associated with heat resistance in reef-building corals. *Current Biology*, *24*(24), 2952–2956. https://doi.org/10.1016/j.cub.2014.10.044

Beger, M., Selkoe, K. A., Treml, E., Barber, P. H., von der Heyden, S., Crandall, E. D., … Riginos, C. (2014). Evolving coral reef conservation with genetic information. *Bulletin of Marine Science*, *90*(1), 159–185. https://doi.org/10.5343/bms.2012.1106

Bellwood, D. R., Hughes, T. P., Folke, C., & Nyström, M. (2004). Confronting the coral reef crisis. *Nature*. https://doi.org/10.1038/nature02691

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., … Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). *Molecular Ecology*, *25*(20), 5073–5092. https://doi.org/10.1111/mec.13811

Berkelmans, R., Weeks, S. J., & Steinberga, C. R. (2010). Upwelling linked to warm summers and bleaching on the Great Barrier Reef. *Limnology and Oceanography*, *55*(6), 2634–2644. https://doi.org/10.4319/lo.2010.55.6.2634

Beveridge, M., & Simmons, L. W. (2006). Panmixia: An example from Dawson's burrowing bee (Amegilla dawsoni) (Hymenoptera: Anthophorini). *Molecular Ecology*, *15*(4), 951–957. https://doi.org/10.1111/j.1365-294X.2006.02846.x

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., … Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, *31*(1), 365–370. https://doi.org/10.1093/nar/gkg095

Bonesso, J. L., Leggat, W., & Ainsworth, T. D. (2017). Exposure to elevated sea-surface temperatures below the bleaching threshold impairs coral recovery and regeneration following injury. *PeerJ*, *2017*(8). https://doi.org/10.7717/peerj.3719

Boulanger, E., Dalongeville, A., Andrello, M., Mouillot, D., & Manel, S. (2020). Spatial graphs highlight how multi-generational dispersal shapes landscape genetic patterns. *Ecography*, ecog.05024. https://doi.org/10.1111/ecog.05024

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370. https://doi.org/10.1007/BF02294361

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Breheny, P., & Burchett, W. (2017). Visualization of regression models using visreg. *R Journal*, *9*(2), 56–71. https://doi.org/10.32614/rj-2017-046

Brodie, A., Azaria, J. R., & Ofran, Y. (2016). How far from the SNP may the causative genes be? *Nucleic Acids Research*, *44*(13), 6046–6054. https://doi.org/10.1093/nar/gkw500

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal*, *9*(2), 378–400. https://doi.org/10.32614/rj-2017-066

Bruno, J. F., & Selig, E. R. (2007). Regional Decline of Coral Cover in the Indo-Pacific: Timing, Extent, and Subregional Comparisons. *PLoS ONE*, *2*(8), e711. https://doi.org/10.1371/journal.pone.0000711

Cabreros, I., & Storey, J. D. (2019). A likelihood-free estimator of population structure bridging admixture models and principal components analysis. *Genetics*, *212*(4), 1009–1029. https://doi.org/10.1534/genetics.119.302159

Cairns, S. D. (1999). Species richness of recent Scleractinia. *Atoll Research Bulletin*, *459*(459–465), 1–46. https://doi.org/10.5479/si.00775630.459.1

Canty, A., & Ripley, B. (2017). boot: Bootstrap R (S-Plus) Functions.

Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications*, *11*(7), 1094–1119. https://doi.org/10.1111/eva.12600

Carvajal-Rodríguez, A. (2008). Simulation of genomes: a review. *Current Genomics*, *9*(3), 155–159. https://doi.org/10.2174/138920208784340759

Cesar, H., Burke, L., & Pet-soede, L. (2003). The Economics of Worldwide Coral Reef Degradation. *Cesar Environmental Economics Consulting, Arnhem, and WWF-Netherlands*, *14*, 23.

Chaichoompu, K., Abegaz, F., Tongsima, S., James Shaw, P., Sakuntabhai, A., Pereira, L., & Van Steen, K. (2018). KRIS: Keen and Reliable Interface Subroutines for Bioinformatic Analysis version 1.1.1 from CRAN. Retrieved from https://cran.r-project.org/package=KRIS

Chao, Y., Li, Z., Farrara, J. D., & Hung, P. (2009). Blending Sea Surface Temperatures from Multiple Satellites and In Situ Observations for Coastal Oceans. *Journal of Atmospheric and Oceanic Technology*, *26*(7), 1415–1426. https://doi.org/10.1175/2009JTECHO592.1

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2015). NbClust : An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*(6), 1–36. https://doi.org/10.18637/jss.v061.i06

Chen, C. A., & Shashank, K. (2009). Taiwan as a connective stepping-stone in the Kuroshio Triangle and the conservation of coral ecosystems under the impacts of climate change. *Kuroshio Science*, *3*(1), 15–22. Retrieved from https://ci.nii.ac.jp/naid/120001597275

Chin, T. M., Vazquez-Cuervo, J., & Armstrong, E. M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sensing of Environment*, *200*, 154–169. https://doi.org/10.1016/j.rse.2017.07.029

CIESIN Columbia University. (2010). Gridded Population of the World (GPW), v3. Retrieved January 18, 2019, from http://sedac.ciesin.columbia.edu/data/set/gpw-v3-population-count

Cinner, J. E., Huchery, C., MacNeil, M. A., Graham, N. A. J., McClanahan, T. R., Maina, J., … Mouillot, D. (2016). Bright spots among the world's coral reefs. *Nature*, *535*(7612), 416–419. https://doi.org/10.1038/nature18607

Colberg, F., Brassington, G. B., Sandery, P., Sakov, P., & Aijaz, S. (2020). High and medium resolution ocean models for the Great Barrier Reef. *Ocean Modelling*, *145*, 101507. https://doi.org/10.1016/j.ocemod.2019.101507

Colli, L., Joost, S., Negrini, R., Nicoloso, L., Crepaldi, P., Ajmone-Marsan, P., … Zundel, S. (2014). Assessing the spatial dependence of adaptive loci in 43 European and Western Asian goat breeds using AFLP markers. *PLoS ONE*, *9*(1), e86668. https://doi.org/10.1371/journal.pone.0086668

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S. J., Kubiszewski, I., … Turner, R. K. (2014). Changes in the global value of ecosystem services. *Global Environmental Change*, *26*(1), 152–158. https://doi.org/10.1016/j.gloenvcha.2014.04.002

Courcelle, J., & Hanawalt, P. C. (1999). RecQ and RecJ process blocked replication forks prior to the resumption of replication in UV-irradiated Escherichia coli. *Molecular & General Genetics : MGG*, *262*(3), 543–551. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10589843

Courtial, L., Picco, V., Grover, R., Cormerais, Y., Rottier, C., Labbe, A., … Ferrier-Pagès, C. (2017). The c-Jun N-terminal kinase prevents oxidative stress induced by UV and thermal stresses in corals and human cells. *Scientific Reports*, *7*(1), 45713. https://doi.org/10.1038/srep45713

Cravatte, S., Kestenare, E., Eldin, G., Ganachaud, A., Lefèvre, J., Marin, F., … Aucan, J. (2015). Regional circulation around New Caledonia from two decades of observations. *Journal of Marine Systems*, *148*, 249–271. https://doi.org/10.1016/j.jmarsys.2015.03.004

Crossley, M. S., Chen, Y. H., Groves, R. L., & Schoville, S. D. (2017). Landscape genomics of Colorado potato beetle provides evidence of polygenic adaptation to insecticides. *Molecular Ecology*, *26*(22), 6284–6300. https://doi.org/10.1111/mec.14339

Cunning, R., Bay, R. A., Gillette, P., Baker, A. C., & Traylor-Knowles, N. (2018). Comparative analysis of the Pocillopora damicornis genome highlights role of immune system in coral evolution. *Scientific Reports*, *8*(1), 16134. https://doi.org/10.1038/s41598-018-34459-8

Dalleau, M., Andréfouët, S., Wabnitz, C. C. C., Payri, C., Wantiez, L., Pichon, M., … Benzoni, F. (2010). Use of habitats as surrogates of biodiversity for efficient coral reef conservation planning in Pacific Ocean islands. *Conservation Biology*, *24*(2), 541–552. https://doi.org/10.1111/j.1523-1739.2009.01394.x

Dance, A. (2019, November 28). These corals could survive climate change — and help save the world's reefs. *Nature*. Nature Research. https://doi.org/10.1038/d41586-019-03629-7

Daniel, D. C., Dagdanova, A. V., & Johnson, E. M. (2013). The MCM and RecQ Helicase Families: Ancient Roles in DNA Replication and Genomic Stability Lead to Distinct Roles in Human Disease. In *The Mechanisms of DNA Replication*. InTech. https://doi.org/10.5772/52961

Darling, E. S., Alvarez-Filip, L., Oliver, T. A., McClanahan, T. R., & Côté, I. M. (2012). Evaluating life-history strategies of reef corals from species traits. *Ecology Letters*, *15*(12), 1378–1386. https://doi.org/10.1111/j.1461-0248.2012.01861.x

Dash, P., Ignatov, A., Martin, M., Donlon, C., Brasnett, B., Reynolds, R. W., … Poulter, D. (2012). Group for High Resolution Sea Surface Temperature (GHRSST) analysis fields inter-comparisons-Part 2: Near real time web-based level 4 SST Quality Monitor (L4-SQUAM). *Deep-Sea Research Part II: Topical Studies in Oceanography*, *77–80*, 31–43. https://doi.org/10.1016/j.dsr2.2012.04.002

Davey, J., Hohenlohe, P., Etter, P., Boone, J., Catchen, J., & Blaxter, M. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499–510. https://doi.org/10.1038/nrg3012

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511802843

Davy, S. K., Allemand, D., & Weis, V. M. (2012). Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and Molecular Biology Reviews : MMBR*, *76*(2), 229–261. https://doi.org/10.1128/MMBR.05014-11

De'ath, G., Fabricius, K. E., Sweatman, H., & Puotinen, M. (2012). The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17995–17999.

https://doi.org/10.1073/pnas.1208909109
De Kort, H., Vandepitte, K., Bruun, H. H., Closset-Kopp, D., Honnay, O., & Mergeay, J. (2014). Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species Alnus glutinosa. *Molecular Ecology*, *23*(19), 4709–4721. https://doi.org/10.1111/mec.12813
De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*(5), 1383–1399. https://doi.org/10.1111/mec.12182
Desalvo, M. K., Sunagawa, S., Voolstra, C. R., & Medina, M. (2010). Transcriptomic responses to heat stress and bleaching in the elkhorn coral Acropora palmata. *Marine Ecology Progress Series*, *402*, 97–113. https://doi.org/10.3354/meps08372
Desalvo, M. K., Voolstra, C. R., Sunagawa, S., Schwarz, J. A., Stillman, J. H., Coffroth, M. A., … Medina, M. (2008). Differential gene expression during thermal stress and bleaching in the Caribbean coral Montastraea faveolata. *Molecular Ecology*, *17*(17), 3952–3971. https://doi.org/10.1111/j.1365-294X.2008.03879.x
Diaz-Pulido, G., & McCook, L. J. (2002). The fate of bleached corals: Patterns and dynamics of algal recruitment. *Marine Ecology Progress Series*, *232*, 115–128. https://doi.org/10.3354/meps232115
DiBattista, J. D., Travers, M. J., Moore, G. I., Evans, R. D., Newman, S. J., Feng, M., … Berry, O. (2017). Seascape genomics reveals fine-scale patterns of dispersal for a reef fish along the ecologically divergent coast of Northwestern Australia. *Molecular Ecology*, *26*(22), 6206–6223. https://doi.org/10.1111/mec.14352
Dittmar, E. L., Oakley, C. G., Conner, J. K., Gould, B. A., & Schemske, D. W. (2016). Factors influencing the effect size distribution of adaptive substitutions. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1828), 20153065. https://doi.org/10.1098/rspb.2015.3065
Dixon, P. M. (2001). Statistics Preprints Statistics The Bootstrap Recommended Citation. *Encyclopedia of Environmetrics*. https://doi.org/10.1002/9780470057339.vab028
Donner, S. D., Rickbeil, G. J. M., & Heron, S. F. (2017). A new, high-resolution global mass coral bleaching database. *PloS One*, *12*(4), e0175490. https://doi.org/10.1371/journal.pone.0175490
Dray, S., & Dufour, A.-B. (2007). The **ade4** Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.18637/jss.v022.i04
Dudaniec, R. Y., Yong, C. J., Lancaster, L. T., Svensson, E. I., & Hansson, B. (2018). Signatures of local adaptation along environmental gradients in a range-expanding damselfly (*Ischnura elegans*). *Molecular Ecology*, *27*(11), 2576–2593. https://doi.org/10.1111/mec.14709
Dunn, S. R., Pernice, M., Green, K., Hoegh-Guldberg, O., & Dove, S. G. (2012). Thermal Stress Promotes Host Mitochondrial Degradation in Symbiotic Cnidarians: Are the Batteries of the Reef Going to Run Out? *PLoS ONE*, *7*(7), e39024. https://doi.org/10.1371/journal.pone.0039024
Dunn, S. R., Schnitzler, C. E., & Weis, V. M. (2007). Apoptosis and autophagy as mechanisms of dinoflagellate symbiont release during cnidarian bleaching: Every which way you lose. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1629), 3079–3085. https://doi.org/10.1098/rspb.2007.0711
Duruz, S., Sevane, N., Selmoni, O., Vajana, E., Leempoel, K., Stucki, S., … Joost, S. (2019). Rapid identification and interpretation of gene-environment associations using the new R.SamBada landscape genomics pipeline. *Molecular Ecology Resources*, 1755–0998.13044. https://doi.org/10.1111/1755-0998.13044
Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1–15. https://doi.org/10.1038/hdy.2010.152
EU Copernicus Marine Service. (2017). Global Ocean - In-Situ-Near-Real-Time Observations. Retrieved February 2, 2017, from http://marine.copernicus.eu
Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815. https://doi.org/10.1080/0266476042000214501
Fine, M., Gildor, H., & Genin, A. (2013). A coral reef refuge in the Red Sea. *Global Change Biology*, *19*(12), 3640–3647. https://doi.org/10.1111/gcb.12356
Fisher, R., Radford, B. T., Knowlton, N., Brainard, R. E., Michaelis, F. B., & Caley, M. J. (2011). Global mismatch between research effort and conservation needs of tropical coral reefs. *Conservation Letters*, *4*(1), 64–72. https://doi.org/10.1111/j.1755-263X.2010.00146.x
Forester, B. R., Landguth, E. L., Hand, B. K., & Balkenhol, N. (2018). Landscape Genomics for Wildlife Research (pp. 1–40). Springer, Cham. https://doi.org/10.1007/13836_2018_56
Frankham, R. (2010, September 1). Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*. Elsevier. https://doi.org/10.1016/j.biocon.2010.05.011
Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382
Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699. https://doi.org/10.1093/molbev/mst063
Fuller, Z. L., Mocellin, V. J. L., Morris, L., Cantin, N., Shepherd, J., Sarre, L., … Przeworski, M. (2019). Population genetics of the coral Acropora millepora: Towards a genomic predictor of bleaching. *BioRxiv*, 867754. https://doi.org/10.1101/867754
Garcia, T. P., & Marder, K. (2017, February 1). Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington's Disease as a Model. *Current Neurology and Neuroscience Reports*. Current Medicine Group LLC 1. https://doi.org/10.1007/s11910-017-0723-4
Gawroński, P., Pawełkowicz, M., Tofil, K., Uszyński, G., Sharifova, S., Ahluwalia, S., … Bolibok-Brągoszewska, H. (2016). DArT Markers Effectively Target Gene Space in the Rye Genome. *Frontiers in Plant Science*, *7*, 1600. https://doi.org/10.3389/fpls.2016.01600
Gélin, P., Fauvelot, C., Bigot, L., Baly, J., & Magalon, H. (2018). From population connectivity to the art of striping Russian dolls: the lessons from *Pocillopora* corals. *Ecology and Evolution*, *8*(2), 1411–1426. https://doi.org/10.1002/ece3.3747
Gélin, P., Pirog, A., Fauvelot, C., & Magalon, H. (2018). High genetic differentiation and low connectivity in the coral Pocillopora damicornis type β at different spatial scales in the Southwestern Indian Ocean and the Tropical Southwestern Pacific. *Marine Biology*, *165*(10). https://doi.org/10.1007/s00227-018-3428-6
Gélin, P., Postaire, B., Fauvelot, C., & Magalon, H. (2017). Reevaluating species number, distribution and endemism of the coral genus Pocillopora Lamarck, 1816 using species delimitation methods and microsatellites. *Molecular Phylogenetics and Evolution*, *109*, 430–446. https://doi.org/10.1016/j.ympev.2017.01.018

Gossa, C., Fisher, M., & Milner-Gulland, E. J. (2015). The research-implementation gap: How practitioners and researchers from developing countries perceive the role of peer-reviewed literature in conservation science. *ORYX*, *49*(1), 80–87. https://doi.org/10.1017/S0030605313001634

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10). https://doi.org/10.1371/journal.pgen.1000695

Halpern, B. S., & Kappel, C. V. (2013). Extinction risk in a changing ocean. In *Saving a Million Species: Extinction Risk from Climate Change* (pp. 285–307). Island Press-Center for Resource Economics . https://doi.org/10.5822/978-1-61091-182-5_16

Harris, S. E., & Munshi-South, J. (2017). Signatures of positive selection and local adaptation to urbanization in white-footed mice (*Peromyscus leucopus*). *Molecular Ecology*, *26*(22), 6336–6350. https://doi.org/10.1111/mec.14369

Harvell, C. D., & Grosberg, R. K. (1988). The timing of sexual maturity in clonal animals. *Ecology*, *69*(6), 1855–1864. https://doi.org/10.2307/1941162

Hecht, B. C., Matala, A. P., Hess, J. E., & Narum, S. R. (2015). Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Molecular Ecology*, *24*(22), 5573–5595. https://doi.org/10.1111/mec.13409

Hénin, C., Guillerm, J., & Chabert, L. (1984). Circulation superficielle autour de la Nouvelle-Cal{é}donie. *Oc{é}anographie Tropicale*, *19*(2), 113–126. Retrieved from https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=8786422

Henry, M. D., Yancey, S. D., & Kushner, S. R. (1992). Role of the heat shock response in stability of mRNA in Escherichia coli K-12. *Journal of Bacteriology*, *174*(3), 743–748. https://doi.org/10.1128/jb.174.3.743-748.1992

Hijmans, R. J. (2016). raster: Geographic Data Analysis and Modeling. Retrieved from https://cran.r-project.org/package=raster

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hoban, S., Arntzen, J. W., Bertorelle, G., Bryja, J., Fernandes, M., Frith, K., … Bruford, M. W. (2013). Conservation Genetic Resources for Effective Species Survival (ConGRESS): Bridging the divide between conservation research and practice. *Journal for Nature Conservation*, *21*(6), 433–437. https://doi.org/10.1016/j.jnc.2013.07.005

Hock, K., Wolff, N. H., Ortiz, J. C., Condie, S. A., Anthony, K. R. N., Blackwell, P. G., & Mumby, P. J. (2017). Connectivity and systemic resilience of the Great Barrier Reef. *PLoS Biology*, *15*(11). https://doi.org/10.1371/journal.pbio.2003355

Holderegger, R., Herrmann, D., Poncet, B., Gugerli, F., Thuiller, W., Taberlet, P., … Manel, S. (2008). Land ahead: using genome scans to identify molecular markers of adaptive relevance. *Plant Ecology & Diversity*, *1*(2), 273–283. https://doi.org/10.1080/17550870802338420

Houlbrèque, F., & Ferrier-Pagès, C. (2009, February). Heterotrophy in tropical scleractinian corals. *Biological Reviews*. https://doi.org/10.1111/j.1469-185X.2008.00058.x

Howells, E. J., Bauman, A. G., Vaughan, G. O., Hume, B. C. C., Voolstra, C. R., & Burt, J. A. (2020). Corals in the hottest reefs in the world exhibit symbiont fidelity not flexibility. *Molecular Ecology*, *29*(5), 899–911. https://doi.org/10.1111/mec.15372

Howells, E. J., Berkelmans, R., van Oppen, M. J. H., Willis, B. L., & Bay, L. K. (2013). Historical thermal regimes define limits to coral acclimatization. *Ecology*, *94*(5), 1078–1088. https://doi.org/10.1890/12-1257.1

Hughes, T. P., Anderson, K. D., Connolly, S. R., Heron, S. F., Kerry, J. T., Lough, J. M., … Wilson, S. K. (2018). Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science*, *359*(6371), 80–83. https://doi.org/10.1126/science.aan8048

Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., … Wilson, S. K. (2017). Global warming and recurrent mass bleaching of corals. *Nature*, *543*(7645), 373–377. https://doi.org/10.1038/nature21707

Hughes, T. P., Kerry, J. T., Baird, A. H., Connolly, S. R., Dietzel, A., Eakin, C. M., … Torda, G. (2018). Global warming transforms coral reef assemblages. *Nature*, *556*(7702), 492–496. https://doi.org/10.1038/s41586-018-0041-2

Hughes, T. P., Kerry, J. T., Connolly, S. R., Baird, A. H., Eakin, C. M., Heron, S. F., … Torda, G. (2019, January 10). Ecological memory modifies the cumulative impact of recurrent climate extremes. *Nature Climate Change*. Nature Publishing Group. https://doi.org/10.1038/s41558-018-0351-2

Hume, B., D'Angelo, C., Burt, J., Baker, A. C., Riegl, B., & Wiedenmann, J. (2013). 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*, *72*(2), 313–322. https://doi.org/10.1016/j.marpolbul.2012.11.032

Imhoff, J. F., Labes, A., & Wiese, J. (2011). Bio-mining the microbial treasures of the ocean: New natural products. *Biotechnology Advances*, *29*(5), 468–482. https://doi.org/10.1016/j.biotechadv.2011.03.001

Ishikawa, Y., Wirz, J., Vranka, J. A., Nagata, K., & Bächinger, H. P. (2009). Biochemical characterization of the prolyl 3-hydroxylase 1·Cartilage-associated protein·cyclophilin B complex. *Journal of Biological Chemistry*, *284*(26), 17641–17647. https://doi.org/10.1074/jbc.M109.007070

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., & Bork, P. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research*, *36*(Database issue), D250-4. https://doi.org/10.1093/nar/gkm796

Job, S. (2018). *Réseau d'observation des récifs coralliens (RORC) de Nouvelle-Calédonie. Campagne 2017-2018. Rapport Pays. Rapport CORTEX. Pour le compte de : Conservatoire d'espaces naturels de Nouvelle-Calédonie - Province des îles Loyauté - Observatoire de l'environne*.

Johnston, E. C., Forsman, Z. H., Flot, J. F., Schmidt-Roach, S., Pinzón, J. H., Knapp, I. S. S., & Toonen, R. J. (2017). A genomic glance through the fog of plasticity and diversification in Pocillopora. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-06085-3

Johnstone, I. M. (2001). On the Distribution of the Largest Eigenvalue in Principal Components Analysis. *The Annals of Statistics*, *29*(2), 295–327.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Joost, S. (2006). *The geographical dimension of genetic diversity: a GIScience contribution for the conservation of animal genetic resources*. EPFL. https://doi.org/10.5075/EPFL-THESIS-3454

Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., & Taberlet, P. (2007). A spatial analysis method (SAM) to detect

candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*, *16*(18), 3955–3969. https://doi.org/10.1111/j.1365-294X.2007.03442.x

Joost, S., Colli, L., Bonin, A., Biebach, I., Allendorf, F. W., Hoffmann, I., … Bruford, M. W. (2011). Promoting collaboration between livestock and wildlife conservation genetics communities. *Conservation Genetics Resources*, *3*(4), 785–788. https://doi.org/10.1007/s12686-011-9457-9

Kawecki, T. J. (2008). Adaptation to Marginal Habitats. *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 321–342. https://doi.org/10.1146/annurev.ecolsys.38.091206.095622

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Keller, L., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*(5), 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kenkel, C. D., Almanza, A. T., & Matz, M. V. (2015). Fine-scale environmental specialization of reef-building corals might be limiting reef recovery in the Florida Keys. *Ecology*, *96*(12), 3197–3212. https://doi.org/10.1890/14-2297.1

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., … Uszynski, G. (2012). Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods in Molecular Biology*, *888*, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Kimura, T., Tun, K., & Chou, L. M. (2018). *Status of Coral Reefs in East Asian Seas Region: 2018*. Tokyo, Japan.

Kleinhaus, K., Al-Sawalmih, A., Barshis, D. J., Genin, A., Grace, L. N., Hoegh-Guldberg, O., … Fine, M. (2020). Science, Diplomacy, and the Red Sea's Unique Coral Reef: It's Time for Action. *Frontiers in Marine Science*, *7*, 90. https://doi.org/10.3389/fmars.2020.00090

Krueck, N. C., Ahmadia, G. N., Green, A., Jones, G. P., Possingham, H. P., Riginos, C., … Mumby, P. J. (2017). Incorporating larval dispersal into MPA design for both conservation and fisheries. *Ecological Applications*, *27*(3), 925–941. https://doi.org/10.1002/eap.1495

Krueger, T., Horwitz, N., Bodin, J., Giovani, M. E., Escrig, S., Meibom, A., & Fine, M. (2017). Common reef-building coral in the northern red sea resistant to elevated temperature and acidification. *Royal Society Open Science*, *4*(5), 170038. https://doi.org/10.1098/rsos.170038

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694

Kültz, D. (2005). Molecular and Evolutionary Basis of the Cellular Stress Response. *Annual Review of Physiology*, *67*(1), 225–257. https://doi.org/10.1146/annurev.physiol.67.040403.103635

LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (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

Landguth, E. L., & Cushman, S. A. (2010). Cdpop: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, *10*(1), 156–161. https://doi.org/10.1111/j.1755-0998.2009.02719.x

Laporte, M., Pavey, S. A., Rougeux, C., Pierron, F., Lauzent, M., Budzinski, H., … Bernatchez, L. (2016). RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, *25*(1), 219–237. https://doi.org/10.1111/mec.13466

Lee, K., Tong, L. T., Millero, F. J., Sabine, C. L., Dickson, A. G., Goyet, C., … Key, R. M. (2006). Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. *Geophysical Research Letters*, *33*(19), L19605. https://doi.org/10.1029/2006GL027207

Leempoel, K., Duruz, S., Rochat, E., Widmer, I., Orozco-terWengel, P., & Joost, S. (2017). Simple Rules for an Efficient Use of Geographic Information Systems in Molecular Ecology. *Frontiers in Ecology and Evolution*, *5*, 33. https://doi.org/10.3389/fevo.2017.00033

Lefèvre, J., Marchesiello, P., Jourdain, N. C., Menkes, C., & Leroy, A. (2010). Weather regimes and orographic circulation around New Caledonia. *Marine Pollution Bulletin*, *61*(7–12), 413–431. https://doi.org/10.1016/j.marpolbul.2010.06.012

Lester, S., Halpern, B., Grorud-Colvert, K., Lubchenco, J., Ruttenberg, B., Gaines, S., … Warner, R. (2009). Biological effects within no-take marine reserves: a global synthesis. *Marine Ecology Progress Series*, *384*, 33–46. https://doi.org/10.3354/meps08029

Lett, C., Verley, P., Mullon, C., Parada, C., Brochier, T., Penven, P., & Blanke, B. (2008). A Lagrangian tool for modelling ichthyoplankton dynamics. *Environmental Modelling & Software*, *23*(9), 1210–1214. https://doi.org/10.1016/J.ENVSOFT.2008.02.005

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., & Ralph, P. (2019). Local PCA shows how the effect of population structure differs along the genome. *Genetics*, *211*(1), 289–304. https://doi.org/10.1534/genetics.118.301747

Li, J., Wang, Q. E., Zhu, Q., El-Mahdy, M. A., Wani, G., Prætorius-Ibba, M., & Wani, A. A. (2006). DNA damage binding protein component DDB1 participates in nucleotide excision repair through DDB2 DNA-binding and cullin 4a ubiquitin ligase activity. *Cancer Research*, *66*(17), 8590–8597. https://doi.org/10.1158/0008-5472.CAN-06-1115

Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., … Morse, D. (2015). The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science*, *350*(6261), 691–694. https://doi.org/10.1126/science.aad0408

Lind, M. I., & Spagopoulou, F. (2018, September 1). Evolutionary consequences of epigenetic inheritance. *Heredity*. Nature Publishing Group. https://doi.org/10.1038/s41437-018-0113-y

Liu, G., Strong, A. E., & Skirving, W. (2003). Remote sensing of sea surface temperatures during 2002 Barrier Reef coral bleaching. *Eos, Transactions American Geophysical Union*, *84*(15), 137–141. https://doi.org/10.1029/2003EO150001

Logan, C. A., Dunne, J. P., Eakin, C. M., & Donner, S. D. (2014). Incorporating adaptive responses into future projections of coral bleaching. *Global Change Biology*, *20*(1), 125–139. https://doi.org/10.1111/gcb.12390

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*(5), 1031–1046. https://doi.org/10.1111/mec.13100

Louis, Y. D., Bhagooli, R., Kenkel, C. D., Baker, A. C., & Dyall, S. D. (2017, January 1). Gene expression biomarkers of heat stress in scleractinian corals: Promises and limitations. *Comparative Biochemistry and Physiology Part - C: Toxicology and Pharmacology*. Elsevier Inc. https://doi.org/10.1016/j.cbpc.2016.08.007

Lowry, D. B. (2012). Local adaptation in The model plant. *New Phytologist*, *194*(4), 888–890. https://doi.org/10.1111/j.1469-8137.2012.04146.x

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Loya, Y., Sakai, K., Yamazato, K., Nakano, Y., Sambali, H., & van Woesik, R. (2001). Coral bleaching: the winners and the losers. *Ecology Letters*, *4*(2), 122–131. https://doi.org/10.1046/j.1461-0248.2001.00203.x

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews. Genetics*, *4*(12), 981–994. https://doi.org/10.1038/nrg1226

Lukoschek, V., Riginos, C., & van Oppen, M. J. H. (2016). Congruent patterns of connectivity can inform management for broadcast spawning corals on the Great Barrier Reef. *Molecular Ecology*, *25*(13), 3065–3080. https://doi.org/10.1111/mec.13649

Lundgren, P., Vera, J. C., Peplow, L., Manel, S., & van Oppen, M. J. H. (2013). Genotype - environment correlations in corals from the Great Barrier Reef. *BMC Genetics*, *14*(1), 9. https://doi.org/10.1186/1471-2156-14-9

Lutz, A., Raina, J. B., Motti, C. A., Miller, D. J., & Van Oppen, M. J. H. (2015). Host coenzyme Q redox state is an early biomarker of thermal stress in the coral Acropora millepora. *PLoS ONE*, *10*(10), e0139290. https://doi.org/10.1371/journal.pone.0139290

Lv, F.-H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., … Ajmone Marsan, P. (2014). Adaptations to Climate-Mediated Selective Pressures in Sheep. *Molecular Biology and Evolution*, *31*(12), 3324–3343. https://doi.org/10.1093/molbev/msu264

Madden, T., & Coulouris, G. (2008). BLAST Command Line Applications User Manual BLAST Command Line Applications User Manual - BLAST ® ... National Center for Biotechnology Information (US). Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK279690/

Magris, R. A., Pressey, R. L., Weeks, R., & Ban, N. C. (2014, February 1). Integrating connectivity and climate change into marine conservation planning. *Biological Conservation*. Elsevier. https://doi.org/10.1016/j.biocon.2013.12.032

Maina, J., McClanahan, T. R., Venus, V., Ateweberhan, M., & Madin, J. (2011). Global gradients of coral exposure to environmental stresses and implications for local management. *PLoS ONE*, *6*(8), e23064. https://doi.org/10.1371/journal.pone.0023064

Maina, J., Venus, V., McClanahan, T. R., & Ateweberhan, M. (2008). Modelling susceptibility of coral reefs to environmental stress using remote sensing data and GIS models. *Ecological Modelling*, *212*(3), 180–199. https://doi.org/10.1016/j.ecolmodel.2007.10.033

Manel, S., Albert, C. H., & Yoccoz, N. G. (2012). Sampling in landscape genomics. *Methods in Molecular Biology*, *888*, 3–12. https://doi.org/10.1007/978-1-61779-870-2_1

Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S., … Fortin, M. J. (2010). Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, *19*(17), 3760–3772. https://doi.org/10.1111/j.1365-294X.2010.04717.x

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6018555

Manthey, J. D., & Moyle, R. G. (2015). Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. *Molecular Ecology*, *24*(14), 3628–3638. https://doi.org/10.1111/mec.13258

Maor-Landaw, K., Karako-Lampert, S., Ben-Asher, H. W., Goffredo, S., Falini, G., Dubinsky, Z., & Levy, O. (2014). Gene expression profiles during short-term heat stress in the red sea coral Stylophora pistillata. *Global Change Biology*, *20*(10), 3026–3035. https://doi.org/10.1111/gcb.12592

Maor-Landaw, K., & Levy, O. (2016). Gene expression profiles during short-term heat stress; branching vs. massive Scleractinian corals of the Red Sea. *PeerJ*, *2016*(3). https://doi.org/10.7717/peerj.1814

Marchesiello, P., Lefèvre, J., Vega, A., Couvelard, X., & Menkes, C. (2010). Coastal upwelling, circulation and heat balance around New Caledonia's barrier reef. *Marine Pollution Bulletin*, *61*(7–12), 432–448. https://doi.org/10.1016/j.marpolbul.2010.06.043

Marhaver, K. L., Edwards, R. A., & Rohwer, F. (2008). Viral communities associated with healthy and bleaching corals. *Environmental Microbiology*, *10*(9), 2277–2286. https://doi.org/10.1111/j.1462-2920.2008.01652.x

Marshall, R. J. (1989). The Predictive Value of Simple Rules for Combining Two Diagnostic Tests. *Biometrics*, *45*(4), 1213. https://doi.org/10.2307/2531772

Mason, N. J., Fiore, J., Kobayashi, T., Masek, K. S., Choi, Y., & Hunter, C. A. (2004). TRAF6-dependent mitogen-activated protein kinase activation differentially regulates the production of interleukin-12 by macrophages in response to Toxoplasma gondii. *Infection and Immunity*, *72*(10), 5662–5667. https://doi.org/10.1128/IAI.72.10.5662-5667.2004

Matsuda, K. (2017). PCR-Based Detection Methods for Single-Nucleotide Polymorphism or Mutation: Real-Time PCR and Its Substantial Contribution Toward Technological Refinement. In *Advances in Clinical Chemistry* (Vol. 80, pp. 45–72). Academic Press Inc. https://doi.org/10.1016/bs.acc.2016.11.002

Matz, M. V., Treml, E. A., Aglyamova, G. V., & Bay, L. K. (2018). Potential and limits for rapid genetic adaptation to warming in a Great Barrier Reef coral. *PLoS Genetics*, *14*(4), e1007220. https://doi.org/10.1371/journal.pgen.1007220

Matz, M. V., Treml, E., & Haller, B. C. (2019). Predicting coral adaptation to global warming in the Indo-West-Pacific. *BioRxiv*, 722314. https://doi.org/10.1101/722314

Mayer, A. M. S., & Gustafson, K. R. (2006). Marine pharmacology in 2003–2004: Anti-tumour and cytotoxic compounds. *European Journal of Cancer*, *42*(14), 2241–2270. https://doi.org/10.1016/j.ejca.2006.05.019

Mazor, T., Possingham, H. P., & Kark, S. (2013). Collaboration among countries in marine conservation can achieve substantial efficiencies. *Diversity and Distributions*, *19*(11), 1380–1393. https://doi.org/10.1111/ddi.12095

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mckerns, M. M., Strand, L., Sullivan, T., Fang, A., & Aivazis, M. A. G. (2011). *Building a Framework for Predictive Science*. *PROC. OF THE 10th PYTHON IN SCIENCE CONF*. Retrieved from https://arxiv.org/abs/1202.1056

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Retrieved from http://conference.scipy.org/proceedings/scipy2010/mckinney.html

McLeod, E., Salm, R., Green, A., & Almany, J. (2009). Designing marine protected area networks to address the impacts of climate change. *Frontiers in Ecology and the Environment*, *7*(7), 362–370. https://doi.org/10.1890/070211

Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., … Donlon, C. (2019). Satellite-based time-series of sea-surface

temperature since 1981 for climate applications. *Scientific Data*, *6*(1), 223. https://doi.org/10.1038/s41597-019-0236-x

Meyer, E., & Weis, V. M. (2012). Study of cnidarian-algal symbiosis in the "omics" age. *Biological Bulletin*, *223*(1), 44–65. https://doi.org/223/1/44 [pii]

Moberg, F., & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, *29*(2), 215–233. https://doi.org/10.1016/S0921-8009(99)00009-9

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G. Y., & Myles, S. (2015). LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3: Genes, Genomes, Genetics*, *5*(11), 2383–2390. https://doi.org/10.1534/g3.115.021667

Mora, C., Andréfouët, S., Costello, M. J., Kranenburg, C., Rollo, A., Veron, J., … Myers, R. A. (2006, June 23). Coral reefs and the global network of marine protected areas. *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.1125295

Mumby, P. J., & Steneck, R. S. (2008, October). Coral reef management and conservation in light of rapidly evolving ecological paradigms. *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2008.06.011

Muñoz-Gómez, S. A., Slamovits, C. H., Dacks, J. B., & Wideman, J. G. (2015). The evolution of MICOS: Ancestral and derived functions and interactions. *Communicative and Integrative Biology*, *8*(6), 1–5. https://doi.org/10.1080/19420889.2015.1094593

Mydlarz, L. D., McGinty, E. S., & Harvell, C. D. (2010). What are the physiological and immunological responses of coral to climate warming and disease? *The Journal of Experimental Biology*, *213*(6), 934–945. https://doi.org/10.1242/jeb.037580

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, *82*(4), 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nakajima, Y., Nishikawa, A., Iguchi, A., & Sakai, K. (2010). Gene Flow and Genetic Diversity of a Broadcast-Spawning Coral in Northern Peripheral Populations. *PLoS ONE*, *5*(6), e11149. https://doi.org/10.1371/journal.pone.0011149

NASA. (2016). NASA Earth Observations (NEO). Retrieved May 10, 2017, from http://neo.sci.gsfc.nasa.gov/

Nathan, R. (2006). Long-distance dispersal of plants. *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.1124975

National Oceanic and Atmospheric Administration. (2017). NOAA DATA CATALOG. Retrieved February 2, 2017, from https://data.noaa.gov/dataset/

Neupane, A., Nepal, M. P., Benson, B. V, MacArthur, K. J., & Piya, S. (2013). Evolutionary history of mitogen-activated protein kinase (MAPK) genes in Lotus, Medicago, and Phaseolus. *Plant Signaling and Behavior*, *8*(11), e27189. https://doi.org/10.4161/psb.27189

Nielsen, D. A., Petrou, K., & Gates, R. D. (2018). Coral bleaching from a single cell perspective. *ISME Journal*, *12*(6), 1558–1567. https://doi.org/10.1038/s41396-018-0080-6

Nishikawa, A. (2008). Degree and Pattern of Gene Flow in Several Scleractinian Corals in the Ryukyu Archipelago, Southern Japan. *Https://Doi.Org/10.2984/1534-6188(2008)62[413:DAPOGF]2.0.CO;2*, *62*(3), 413–422. https://doi.org/10.2984/1534-6188(2008)62[413:DAPOGF]2.0.CO;2

Nishikawa, A., Katoh, M., & Sakai, K. (2003). Larval settlement rates and gene flow of broadcast-spawning (Acropora tenuis) and planula-brooding (Stylophora pistillata) corals. *Marine Ecology Progress Series*, *256*, 87–97. https://doi.org/10.3354/meps256087

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009, February 1). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*. John Wiley & Sons, Ltd (10.1111). https://doi.org/10.1111/j.1365-294X.2008.03946.x

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., … Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, *456*(7218), 98–101. https://doi.org/10.1038/nature07331

Oakley, C. A., & Davy, S. K. (2018). Cell Biology of Coral Bleaching (pp. 189–211). Springer, Cham. https://doi.org/10.1007/978-3-319-75393-5_8

Oakley, C. A., Durand, E., Wilkinson, S. P., Peng, L., Weis, V. M., Grossman, A. R., & Davy, S. K. (2017). Thermal Shock Induces Host Proteostasis Disruption and Endoplasmic Reticulum Stress in the Model Symbiotic Cnidarian Aiptasia. *Journal of Proteome Research*, *16*(6), 2121–2134. https://doi.org/10.1021/acs.jproteome.6b00797

OECD. (2017). *Marine Protected Areas. Economics, Management and Effective Policy Mixes*. *OECD Publishing*. Paris: OECD Publishing. https://doi.org/10.2307/2261484

Osman, E. O., Smith, D. J., Ziegler, M., Kürten, B., Conrad, C., El-Haddad, K. M., … Suggett, D. J. (2018). Thermal refugia against coral bleaching throughout the northern Red Sea. *Global Change Biology*, *24*(2), e474–e484. https://doi.org/10.1111/gcb.13895

Palardy, J. E., Rodrigues, L. J., & Grottoli, A. G. (2008). The importance of zooplankton to the daily metabolic carbon requirements of healthy and bleached corals at two depths. *Journal of Experimental Marine Biology and Ecology*, *367*(2), 180–188. https://doi.org/10.1016/j.jembe.2008.09.015

Palazzo, A. F., & Gregory, T. R. (2014). The Case for Junk DNA. *PLoS Genetics*, *10*(5). https://doi.org/10.1371/journal.pgen.1004351

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, *13*(1 SUPPL.), 146–158. https://doi.org/10.1890/1051-0761(2003)013[0146:PGDCAT]2.0.CO;2

Palumbi, S. R., Barshis, D. J., Traylor-Knowles, N., & Bay, R. A. (2014). Mechanisms of reef coral resistance to future climate change. *Science*, *344*(6186). Retrieved from http://science.sciencemag.org/content/344/6186/895

Pardo-Diaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, *6*(4), 445–464. https://doi.org/10.1111/2041-210X.12324

Paris-Limouzy, C. B. (2011). Reef Interconnectivity/Larval Dispersal (pp. 881–889). Springer, Dordrecht. https://doi.org/10.1007/978-90-481-2639-2_138

Pariset, L., Joost, S., Marsan, P., & Valentini, A. (2009). Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics*, *10*(1), 7. https://doi.org/10.1186/1471-2156-10-7

Patel, R., Rinker, L., Peng, J., & Chilian, W. M. (2018). Reactive Oxygen Species: The Good and the Bad. In *Reactive Oxygen Species (ROS) in Living Cells*. InTech. https://doi.org/10.5772/intechopen.71547

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), 2074–2093. https://doi.org/10.1371/journal.pgen.0020190

Penin, L., Vidal-Dupiol, J., & Adjeroud, M. (2013). Response of coral assemblages to thermal stress: are bleaching intensity and spatial patterns consistent between events? *Environmental Monitoring and Assessment*, *185*(6), 5031–5042. https://doi.org/10.1007/s10661-012-2923-3

Pluess, A. R., Frank, A., Heiri, C., Lalagüe, H., Vendramin, G. G., & Oddou-Muratorio, S. (2016). Genome-environment association study

suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytologist*, *210*(2), 589–601. https://doi.org/10.1111/nph.13809

Porté, S., Valencia, E., Yakovtseva, E. A., Borràs, E., Shafqat, N., Debreczeny, J. E., … Parés, X. (2009). Three-dimensional structure and enzymatic function of proapoptotic human p53-inducible quinone oxidoreductase PIG3. *Journal of Biological Chemistry*, *284*(25), 17194–17205. https://doi.org/10.1074/jbc.M109.001800

Pratchett, M. S., McCowan, D., Maynard, J. A., & Heron, S. F. (2013). Changes in bleaching susceptibility among corals subject to ocean warming and recurrent bleaching in Moorea, French Polynesia. *PloS One*, *8*(7), e70443. https://doi.org/10.1371/journal.pone.0070443

Pratchett, M. S., Thompson, C. A., Hoey, A. S., Cowman, P. F., & Wilson, S. K. (2018). Effects of Coral Bleaching and Coral Loss on the Structure and Function of Reef Fish Assemblages (pp. 265–293). Springer, Cham. https://doi.org/10.1007/978-3-319-75393-5_11

Prindle, M. J., & Loeb, L. A. (2012, December). DNA polymerase delta in dna replication and genome maintenance. *Environmental and Molecular Mutagenesis*. NIH Public Access. https://doi.org/10.1002/em.21745

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Python Software Foundation. (2018). Python Language Refrence, version 3.5. Retrieved from www.python.org

QGIS development team. (2009). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from http://www.qgis.org/

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Retrieved from https://www.r-project.org/

Rabinowicz, P. D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L. D., Stein, L., … Martienssen, R. A. (1999). Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nature Genetics*, *23*(3), 305–308. https://doi.org/10.1038/15479

Ray, N., Currat, M., Foll, M., & Excoffier, L. (2010). SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, *26*(23), 2993–2994. https://doi.org/10.1093/bioinformatics/btq579

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. https://doi.org/10.1111/mec.13322

Ren, B., Liu, M., Ni, J., & Tian, J. (2018, November 2). Role of selenoprotein f in protein folding and secretion: Potential involvement in human disease. *Nutrients*. MDPI AG. https://doi.org/10.3390/nu10111619

Ricaurte, M., Schizas, N. V., Ciborowski, P., & Boukli, N. M. (2016). Proteomic analysis of bleached and unbleached Acropora palmata, a threatened coral species of the Caribbean. *Marine Pollution Bulletin*, *107*(1), 224–232. https://doi.org/10.1016/j.marpolbul.2016.03.068

Riginos, C., Crandall, E. D., Liggins, L., Bongaerts, P., & Treml, E. A. (2016). Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. *Current Zoology*, *62*, doi: 10.1093/cz/zow067. https://doi.org/10.1093/cz/zow067

Robbins, S. J., Singleton, C. M., Chan, C. X., Messer, L. F., Geers, A. U., Ying, H., … Tyson, G. W. (2019, December 1). A genomic view of the reef-building coral Porites lutea and its microbial symbionts. *Nature Microbiology*. Nature Research. https://doi.org/10.1038/s41564-019-0532-4

Robinson, J. P. W., Wilson, S. K., & Graham, N. A. J. (2019). Abiotic and biotic controls on coral recovery 16 years after mass bleaching. *Coral Reefs*, *38*(6), 1255–1265. https://doi.org/10.1007/s00338-019-01831-7

Rochat, E., & Joost, S. (2019). Spatial Areas of Genotype Probability (SPAG): predicting the spatial distribution of adaptive genetic variants under future climatic conditions. *BioRxiv*, 2019.12.20.884114. https://doi.org/10.1101/2019.12.20.884114

Roche, R. C., Williams, G. J., & Turner, J. R. (2018, March 1). Towards Developing a Mechanistic Understanding of Coral Reef Resilience to Thermal Stress Across Multiple Scales. *Current Climate Change Reports*. Springer. https://doi.org/10.1007/s40641-018-0087-0

Rodgers, K. S., Lorance, K., Richards Donà, A., Stender, Y., Lager, C., & Jokiel, P. L. (2017). Effectiveness of coral relocation as a mitigation strategy in Kāne'ohe Bay, Hawai'i. *PeerJ*, *5*, e3346. https://doi.org/10.7717/peerj.3346

Roff, G., Doropoulos, C., Rogers, A., Bozec, Y. M., Krueck, N. C., Aurellado, E., … Mumby, P. J. (2016). The Ecological Role of Sharks on Coral Reefs. *Trends in Ecology and Evolution*, *31*(5), 395–407. https://doi.org/10.1016/j.tree.2016.02.014

Rose, D. C., Sutherland, W. J., Amano, T., González-Varo, J. P., Robertson, R. J., Simmons, B. I., … Mukherjee, N. (2018). The major barriers to evidence-informed conservation policy and possible solutions. *Conservation Letters*, *11*(5), e12564. https://doi.org/10.1111/conl.12564

Rosic, N. N., Pernice, M., Dove, S., Dunn, S., & Hoegh-Guldberg, O. (2011). Gene expression profiles of cytosolic heat shock proteins Hsp70 and Hsp90 from symbiotic dinoflagellates in response to thermal stress: Possible implications for coral bleaching. *Cell Stress and Chaperones*, *16*(1), 69–80. https://doi.org/10.1007/s12192-010-0222-x

Roux, D. J., Rogers, K. H., Biggs, H. C., Ashton, P. J., & Sergeant, A. (2006). Bridging the science-management divide: Moving from unidirectional knowledge transfer to knowledge interfacing and sharing. *Ecology and Society*, *11*(1). https://doi.org/10.5751/ES-01643-110104

Ruiz-Jones, L. J., & Palumbi, S. R. (2017). Tidal heat pulses on a reef trigger a fine-tuned transcriptional response in corals to maintain homeostasis. *Science Advances*, *3*(3), e1601298. https://doi.org/10.1126/sciadv.1601298

Ruppert, E., Fox, R., & Barnes, R. (2004). *Invertebrate Zoology*. Cengage Learning.

Ruppert, J. L. W., Travers, M. J., Smith, L. L., Fortin, M. J., & Meekan, M. G. (2013). Caught in the Middle: Combined Impacts of Shark Removal and Coral Loss on the Fish Communities of Coral Reefs. *PLoS ONE*, *8*(9), e74648. https://doi.org/10.1371/journal.pone.0074648

Ryan, W. B. F., Carbotte, S. M., Coplan, J. O., O'Hara, S., Melkonian, A., Arko, R., … Zemsky, R. (2009). Global multi-resolution topography synthesis. *Geochemistry, Geophysics, Geosystems*, *10*(3). https://doi.org/10.1029/2008GC002332

Safaie, A., Silbiger, N. J., McClanahan, T. R., Pawlak, G., Barshis, D. J., Hench, J. L., … Davis, K. A. (2018). High frequency temperature variability reduces the risk of coral bleaching. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04074-2

Sampayo, E. M., Ridgway, T., Franceschinis, L., Roff, G., Hoegh-Guldberg, O., & Dove, S. (2016). Coral symbioses under prolonged environmental change: Living near tolerance range limits. *Scientific Reports*, *6*(1), 36271. https://doi.org/10.1038/srep36271

Schmidt-Roach, S., Lundgren, P., Miller, K. J., Gerlach, G., Noreen, A. M. E., & Andreakis, N. (2013). Assessing hidden species diversity in the coral Pocillopora damicornis from Eastern Australia. *Coral Reefs*, *32*(1), 161–172. https://doi.org/10.1007/s00338-012-0959-z

Schmidt-Roach, S., Miller, K. J., Lundgren, P., & Andreakis, N. (2014). With eyes wide open: a revision of species within and closely related to the *Pocillopora damicornis* species complex (Scleractinia; Pocilloporidae) using morphology and genetics. *Zoological Journal of the*

Linnean Society, *170*(1), 1–33. https://doi.org/10.1111/zoj.12092

Seabold, S., & Perktold, J. (2010). Statsmodels: econometric and statistical modeling with Python. In *9th Python in Science Conference* (pp. 57–61). Retrieved from http://statsmodels.sourceforge.net/

Selig, E. R., & Bruno, J. F. (2010). A global analysis of the effectiveness of marine protected areas in preventing coral loss. *PloS One*, *5*(2), e9278. https://doi.org/10.1371/journal.pone.0009278

Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., … Toonen, R. J. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, *554*(554), 1–19. https://doi.org/10.3354/meps11792

Selkoe, K. A., Scribner, K. T., & Galindo, H. M. (2015). Waterscape Genetics - Applications of Landscape Genetics to Rivers, Lakes, and Seas. In *Landscape Genetics* (pp. 220–246). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118525258.ch13

Selmoni, O., Lecellier, G., Ainley, L., Collin, A., Doucet, R., Dubousquet, V., … Berteaux-Lecellier, V. (2020). Using Modern Conservation Tools for Innovative Management of Coral Reefs: The MANACO Consortium. *Frontiers in Marine Science*, *7*, 609. https://doi.org/10.3389/fmars.2020.00609

Selmoni, O., Lecellier, G., Magalon, H., Vigliola, L., Benzoni, F., Peignon, C., … Berteaux-Lecellier, V. (2020). Seascape genomics reveals candidate molecular targets of heat stress adaptation in three coral species. *BioRxiv*, 2020.05.12.090050. https://doi.org/10.1101/2020.05.12.090050

Selmoni, O., Rochat, E., Lecellier, G., Berteaux-Lecellier, V., & Joost, S. (2020). Seascape genomics as a new tool to empower coral reef conservation strategies: an example on north-western Pacific Acropora digitifera. *Evolutionary Applications*, 588228. https://doi.org/10.1101/588228

Selmoni, O., Vajana, E., Guillaume, A., Rochat, E., & Joost, S. (2020). Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach. *Molecular Ecology Resources*, *20*(1). https://doi.org/10.1111/1755-0998.13095

Shanks, A. L., Grantham, B. A., & Carr, M. H. (2003). Propagule Dispersal Distance and the Size and Spacing of Marine Reserves. *Ecological Applications*, *13*(1). https://doi.org/https://doi.org/10.1890/1051-0761(2003)013[0159:PDDATS]2.0.CO;2

Sharma, S., Doherty, K. M., Brosh, R. M., & Jr. (2006). Mechanisms of RecQ helicases in pathways of DNA metabolism and maintenance of genomic stability. *The Biochemical Journal*, *398*(3), 319–337. https://doi.org/10.1042/BJ20060450

Shinzato, C., Mungpakdee, S., Arakaki, N., & Satoh, N. (2015). Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago. *Scientific Reports*, *5*(1), 18211. https://doi.org/10.1038/srep18211

Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., … Satoh, N. (2011). Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature*, *476*(7360), 320–323. https://doi.org/10.1038/nature10249

Shoguchi, E., Beedessee, G., Tada, I., Hisata, K., Kawashima, T., Takeuchi, T., … Shinzato, C. (2018). Two divergent Symbiodinium genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. *BMC Genomics*, *19*(1), 458. https://doi.org/10.1186/s12864-018-4857-9

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., … Satoh, N. (2013). Draft assembly of the symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Current Biology*, *23*(15), 1399–1408. https://doi.org/10.1016/j.cub.2013.05.062

Smits, V. A. J., Cabrera, E., Freire, R., & Gillespie, D. A. (2019, February 1). Claspin – checkpoint adaptor and DNA replication factor. *FEBS Journal*. Blackwell Publishing Ltd. https://doi.org/10.1111/febs.14594

Son, Y., Kim, S., Chung, H. T., & Pae, H. O. (2013). Reactive oxygen species in the activation of MAP kinases. In *Methods in Enzymology* (Vol. 528, pp. 27–48). Academic Press Inc. https://doi.org/10.1016/B978-0-12-405881-1.00002-1

Sottile, M. L., & Nadin, S. B. (2018, May 1). Heat shock proteins and DNA repair mechanisms: an updated overview. *Cell Stress and Chaperones*. Cell Stress and Chaperones. https://doi.org/10.1007/s12192-017-0843-4

Statisticat, & LCC. (2018). LaplacesDemon: Complete Environment for Bayesian Inference. Bayesian-Inference.com.

Storey, J. D. (2003). The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value. *Annals of Statistics*, *31*(6), 2013–2035.

Storlazzi, C. D., van Ormondt, M., Chen, Y.-L., & Elias, E. P. L. (2017). Modeling Fine-Scale Coral Larval Dispersal and Interisland Connectivity to Help Designate Mutually-Supporting Coral Reef Marine Protected Areas: Insights from Maui Nui, Hawaii. *Frontiers in Marine Science*, *4*, 381. https://doi.org/10.3389/fmars.2017.00381

Stronen, A. V., Jędrzejewska, B., Pertoldi, C., Demontis, D., Randi, E., Niedziałkowska, M., … Czarnomska, S. D. (2015). Genome-wide analyses suggest parallel selection for universal traits may eclipse local environmental selection in a highly mobile carnivore. *Ecology and Evolution*, *5*(19), 4410–4425. https://doi.org/10.1002/ece3.1695

Stucki, S., Orozco-terWengel, P., Bruford, M. W., Colli, L., Masembe, C., Negrini, R., … Consortium, the N. (2017). High performance computation of landscape genomic models integrating local indices of spatial association. *Molecular Ecology Resources*, *17*(5), 1072–1089. https://doi.org/10.1111/1755-0998.12629

Sully, S., Burkepile, D. E., Donovan, M. K., Hodgson, G., & van Woesik, R. (2019). A global analysis of coral bleaching over the past two decades. *Nature Communications*, *10*(1), 1–5. https://doi.org/10.1038/s41467-019-09238-2

Tchernov, D., Kvitt, H., Haramaty, L., Bibbyd, T. S., Gorbunov, M. Y., Rosenfeld, H., & Falkowski, P. G. (2011). Apoptosis and the selective survival of host animals following thermal bleaching in zooxanthellate corals. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9905–9909. https://doi.org/10.1073/pnas.1106924108

The IMAGE Consortium. (2019). IMAGE - Innovative Management of Animal Genetic resources. Retrieved September 13, 2018, from http://www.imageh2020.eu/index.php

Theodorou, P., Radzevičiūtė, R., Kahnt, B., Soro, A., Grosse, I., & Paxton, R. J. (2018). Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.). *Proceedings of the Royal Society B: Biological Sciences*, *285*(1877), 20172806. https://doi.org/10.1098/rspb.2017.2806

Thomas, C. J., Lambrechts, J., Wolanski, E., Traag, V. A., Blondel, V. D., Deleersnijder, E., & Hanert, E. (2014). Numerical modelling and graph theory tools to study ecological connectivity in the Great Barrier Reef. *Ecological Modelling*, *272*, 160–174. https://doi.org/10.1016/j.ecolmodel.2013.10.002

Thomas, L., Kennington, W. J., Evans, R. D., Kendrick, G. A., & Stat, M. (2017). Restricted gene flow and local adaptation highlight the vulnerability of high-latitude reefs to rapid environmental change. *Global Change Biology*, *23*(6), 2197–2205. https://doi.org/10.1111/gcb.13639

Thomas, L., Rose, N. H., Bay, R. A., López, E. H., Morikawa, M. K., Ruiz-Jones, L., & Palumbi, S. R. (2018). Mechanisms of Thermal Tolerance

in Reef-Building Corals across a Fine-Grained Environmental Mosaic: Lessons from Ofu, American Samoa. *Frontiers in Marine Science*, *4*, 434. https://doi.org/10.3389/fmars.2017.00434

Thompson, D. M., & van Woesik, R. (2009). Corals escape bleaching in regions that recently and historically experienced frequent thermal stress. *Proceedings. Biological Sciences / The Royal Society*, *276*(1669), 2893–2901. https://doi.org/10.1098/rspb.2009.0591

Timar, L., & Phaneuf, D. J. (2009). Modeling the human-induced spread of an aquatic invasive: The case of the zebra mussel. *Ecological Economics*, *68*(12), 3060–3071. https://doi.org/10.1016/j.ecolecon.2009.07.011

Treml, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., & Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. In *Integrative and Comparative Biology* (Vol. 52, pp. 525–537). https://doi.org/10.1093/icb/ics101

U.S. Geological Survey. (2016). *Landsat—Earth observation satellites*. *USGS* (Vol. 2020). https://doi.org/10.3133/fs20153081

UNEP-WCMC, WorldFish-Center, WRI, & TNC. (2010). Global distribution of warm-water coral reefs, compiled from multiple sources including the Millennium Coral Reef Mapping Project. Version 1.3. Retrieved May 9, 2017, from http://data.unep-wcmc.org/datasets/1

Vajana, E., Barbato, M., Colli, L., Milanesi, M., Rochat, E., Fabrizi, E., … Ajmone Marsan, P. (2018). Combining landscape genomics and ecological modelling to investigate local adaptation of indigenous Ugandan cattle to East Coast fever. *Frontiers in Genetics*, *9*, 385. https://doi.org/10.3389/FGENE.2018.00385

Valmiki, M. G., & Ramos, J. W. (2009, March). Death effector domain-containing proteins. *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s00018-008-8489-0

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(SUPL.43), 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

van Etten, J. (2018). gdistance: Distances and Routes on Geographical Grids. Retrieved from https://cran.r-project.org/package=gdistance

Van Hooidonk, R., Maynard, J. A., & Planes, S. (2013). Temporary refugia for coral reefs in a warming world. *Nature Climate Change*, *3*(5), 508–511. https://doi.org/10.1038/nclimate1829

van Hooidonk, R., Maynard, J., Tamelander, J., Gove, J., Ahmadia, G., Raymundo, L., … Planes, S. (2016). Local-scale projections of coral reef futures and implications of the Paris Agreement. *Scientific Reports*, *6*, 39666. https://doi.org/10.1038/srep39666

van Oppen, M. J. H., Gates, R. D., Blackall, L. L., Cantin, N., Chakravarti, L. J., Chan, W. Y., … Putnam, H. M. (2017, September 1). Shifting paradigms in restoration of the world's coral reefs. *Global Change Biology*. https://doi.org/10.1111/gcb.13647

van Oppen, M. J. H., & Lough, J. M. (2009). Coral Bleaching — Patterns, Processes, Causes and Consequences (pp. 298–201). Springer. https://doi.org/10.1007/978-3-540-69775-6_11

van Oppen, M. J. H., Oliver, J. K., Putnam, H. M., & Gates, R. D. (2015, February 24). Building coral reef resilience through assisted evolution. *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences. https://doi.org/10.1073/pnas.1422301112

van Oppen, M. J. H., Willis, B. L., Van Rheede, T., & Miller, D. J. (2002). Spawning times, reproductive compatibilities and genetic structuring in the Acropora aspera group: Evidence for natural hybridization and semi-permeable species boundaries in corals. *Molecular Ecology*, *11*(8), 1363–1376. https://doi.org/10.1046/j.1365-294X.2002.01527.x

Van Woesik, R., Irikawa, A., & Loya, Y. (2004). Coral Bleaching: Signs of Change in Southern Japan. In *Coral Health and Disease* (pp. 119–141). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-06414-6_5

Vanhellemont, Q. (2020). Automated water surface temperature retrieval from Landsat 8/TIRS. *Remote Sensing of Environment*, *237*. https://doi.org/10.1016/j.rse.2019.111518

Verbeke, G., Molenberghs, G., & Rizopoulos, D. (2010). Random effects models for longitudinal data. In *Longitudinal Research with Latent Variables* (pp. 37–96). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11760-2_2

Vincent, B., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. (2013). Landscape genomics in atlantic salmon (salmo salar): Searching for gene-environment interactions driving local adaptation. *Evolution*, *67*(12), 3469–3487. https://doi.org/10.1111/evo.12139

Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature*, *461*(7261), 199–205. https://doi.org/10.1038/nature08451

Voolstra, C. R., Li, Y., Liew, Y. J., Baumgarten, S., Zoccola, D., Flot, J. F., … Aranda, M. (2017). Comparative analysis of the genomes of Stylophora pistillata and Acropora digitifera provides evidence for extensive differences between species of corals. *Scientific Reports*, *7*(1), 1–14. https://doi.org/10.1038/s41598-017-17484-x

Voolstra, C. R., Schnetzer, J., Peshkin, L., Randall, C. J., Szmant, A. M., & Medina, M. (2009). Effects of temperature on gene expression in embryos of the coral Montastraea faveolata. *BMC Genomics*, *10*, 627. https://doi.org/10.1186/1471-2164-10-627

Voolstra, C. R., Sunagawa, S., Matz, M. V., Bayer, T., Aranda, M., Buschiazzo, E., … Medina, M. (2011). Rapid evolution of coral proteins responsible for interaction with the environment. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0020392

Wallace, C. C. (1999). *Staghorn corals of the world : a revision of the coral genus Acropora (Scleractinia ; Astrocoeniina ; Acroporidae) worldwide, with emphasis on morphology, phylogeny and biogeography*. CSIRO.

Wang, B., Luo, X., Yang, Y. M., Sun, W., Cane, M. A., Cai, W., … Liu, J. (2019). Historical change of El Niño properties sheds light on future changes of extreme El Niño. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(45), 22512–22517. https://doi.org/10.1073/pnas.1911130116

Wang, L., Zhang, W., Li, Q., & Zhu, W. (2017). AssocTests: Genetic Association Studies.

Webb, T. J., & Mindel, B. L. (2015). Global patterns of extinction risk in marine and non-marine systems. *Current Biology*, *25*(4), 506–511. https://doi.org/10.1016/j.cub.2014.12.023

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358. https://doi.org/10.2307/2408641

Welle, P. D., Small, M. J., Doney, S. C., & Azevedo, I. L. (2017). Estimating the effect of multiple environmental stressors on coral bleaching and mortality. *PloS One*, *12*(5), e0175018. https://doi.org/10.1371/journal.pone.0175018

Wenzel, M. A., Douglas, A., James, M. C., Redpath, S. M., & Piertney, S. B. (2016). The role of parasite-driven selection in shaping landscape genomic structure in red grouse (*Lagopus lagopus scotica*). *Molecular Ecology*, *25*(1), 324–341. https://doi.org/10.1111/mec.13473

Westgate, M. J., Haddaway, N. R., Cheng, S. H., McIntosh, E. J., Marshall, C., & Lindenmayer, D. B. (2018, April 1). Software support for environmental evidence synthesis. *Nature Ecology and Evolution*. Nature Publishing Group. https://doi.org/10.1038/s41559-018-

0502-x

White, A. T., Vogt, H. P., & Arin, T. (2000). Philippine Coral Reefs Under Threat: The Economic Losses Caused by Reef Destruction. *Marine Pollution Bulletin*, *40*(7), 598–605. https://doi.org/10.1016/S0025-326X(00)00022-9

Wilkinson, C. (2008). *Status of Coral Reefs of the World: 2000 Edited by Clive Wilkinson*. Retrieved from www.aims.gov.au

Wilson, C., Terman, J. R., González-Billault, C., & Ahmed, G. (2016, October 1). Actin filaments—A target for redox regulation. *Cytoskeleton*. John Wiley and Sons Inc. https://doi.org/10.1002/cm.21315

Wilson, K. L., Tittensor, D. P., Worm, B., & Lotze, H. K. (2020). Incorporating climate change adaptation into marine protected area planning. *Global Change Biology*, *26*(6), 3251–3267. https://doi.org/10.1111/gcb.15094

Yamazato, K. (1999). Coral bleaching in Okinawa, 1980 vs 1998. *Journal of the Japanese Coral Reef Society*, *1999*(1), 83–87. https://doi.org/10.3755/jcrs.1999.83

Ying, H., Cooke, I., Sprungala, S., Wang, W., Hayward, D. C., Tang, Y., … Miller, D. J. (2018, November 2). Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. *Genome Biology*. BioMed Central Ltd. https://doi.org/10.1186/s13059-018-1552-8

Yoder, J. B., Stanton-Geddes, J., Zhou, P., Briskine, R., Young, N. D., & Tiffin, P. (2014). Genomic Signature of Adaptation to Climate in *Medicago truncatula*. *Genetics*, *196*(4), 1263–1275. https://doi.org/10.1534/genetics.113.159319

Young, C. N., Schopmeyer, S. A., & Lirman, D. (2012). A review of reef restoration and Coral propagation using the threatened genus Acropora in the Caribbean and western Atlantic. *Bulletin of Marine Science*, *88*(4), 1075–1098. https://doi.org/10.5343/bms.2011.1143

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, *11*(2), R14. https://doi.org/10.1186/gb-2010-11-2-r14

Yuasa, K., Ota, R., Matsuda, S., Isshiki, K., Inoue, M., & Tsuji, A. (2015). Suppression of death-associated protein kinase 2 by interaction with 14-3-3 proteins. *Biochemical and Biophysical Research Communications*, *464*(1), 70–75. https://doi.org/10.1016/j.bbrc.2015.05.105

Zangar, R. C., Davydov, D. R., & Verma, S. (2004, September 15). Mechanisms that regulate production of reactive oxygen species by cytochrome P450. *Toxicology and Applied Pharmacology*. Academic Press. https://doi.org/10.1016/j.taap.2004.01.018

Zhao, R. Z., Jiang, S., Zhang, L., & Yu, Z. Bin. (2019). Mitochondrial electron transport chain, ROS generation and uncoupling (Review). *International Journal of Molecular Medicine*. Spandidos Publications. https://doi.org/10.3892/ijmm.2019.4188

Zheng, M., Cierpicki, T., Burdette, A. J., Utepbergenov, D., Janczyk, P. L., Derewenda, U., … Derewenda, Z. S. (2011). Structural features and chaperone activity of the NudC protein family. *Journal of Molecular Biology*, *409*(5), 722–741. https://doi.org/10.1016/j.jmb.2011.04.018

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., & Voolstra, C. R. (2017). Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nature Communications*, *8*, 14213. https://doi.org/10.1038/ncomms14213

9. Curriculum vitae

Oliver Selmoni

PhD Student
Laboratory of Geographic Information Systems (LASIG), École Polytechnique Fédérale de Lausanne (EPFL)

@ oliver.selmoni@epfl.ch +41 76 477 3694

Avenue du Tribunal-Fédéral 38, 1005 Lausanne
https://www.researchgate.net/profile/Oliver_Selmoni

RESEARCH

ManaCo network
Laboratory of Geographic Information Systems (LASIG, EPFL, CH)

2019- Nouméa (NC) http://manaco.ird.nc

Coordinators: Véronique Berteaux-Lecellier (IRD), Stéphane Joost (EPFL)
Contribution to the creation of an international network for coral reef research and preservation through genomic analyses.

Red Sea Transnational Research Center
Laboratory of Geographic Information Systems (LASIG, EPFL, CH)

2019- Lausanne (CH)

Project leader: Anders Meibom (EPFL)
Contribution to the seascape genomics study on coral of the Red Sea. Supported by the Swiss Department of Foreign Affairs (DFAE).

SABLE project
Laboratory of Geographic Information Systems (LASIG, EPFL, CH)
Lab. of Marine Tropical Ecology of Indo-Pacific (ENTROPIE, IRD, New Caledonia)

2018-2019 Lausanne (CH), Nouméa (NC, FR)

Supervisors: Stéphane Joost (EPFL), Véronique Berteaux-Lecellier (IRD)
PhD research project on the investigation of local adaptation to climate change in corals of New Caledonia. Supported by the International Coral Reef Initiative (ICRI) and the United Nation Environment Program (UNEP).

IMAGE project
Laboratory of Geographic Information Systems (LASIG, EPFL, CH)

2016-2020 Lausanne (CH)

Supervisors: Stéphane Joost (EPFL)
Support in the creation of an innovative online repository for genetic resources of livestock animals. EU-H2020 project.

URBANGENE project
Laboratory of Geographic Information Systems (LASIG, EPFL, CH)

2016-2017 Lausanne (CH)

Supervisors: Stéphane Joost (EPFL)
Landscape genomics study on the impact of urbanization on genetic diversity of cosmopolitan species in Geneva (CH).

Lake Thun grayling project
Evolutionary Conservation Biology Lab (UNIL, CH)
Evolutionary Bioinformatics Lab (UNIL, CH)

2014-2016 Lausanne (CH)

Supervisors: Claus Wedekind (UNIL), Marc Robinson-Rechavi (UNIL)
Master thesis research studying the molecular effects of estrogen pollution on a salmonid species of Lake Thun (CH).

EDUCATION

PhD in landscape and seascape genomics
École Polytechnique Fédérale de Lausanne, EPFL (CH)
Institut de la Recherche pour le Développement (IRD, New Caledonia, France)

2016-2020 Lausanne (CH)

MSc in Molecular Life Science
- Major: Bioinformatics

University of Lausanne, UNIL

2014-2016 Lausanne (CH)

BSc in Biology
University of Lausanne, UNIL

MY INTERESTS

Graduated in bioinformatics, PhD thesis in landscape and seascape genomics. I am interested in creating channels to transpose research findings in wildlife conservation strategies.

LANGUAGES

Italian	Mother tongue
French	Bilingual proficiency
English	Fluent in written and spoken
German	Basic in written and spoken

2011-2014 Lausanne (CH)

TEACHING

Seascape genomics, instructor

2020- Quebec (Canada)

Co-instructor: Laura Benestan (CEFE)

One-week course for graduated researchers at Physalia courses (https://www.physalia-courses.org/) on the seascape genomics approach.

Landscape genomics, co-instructor

2017-2018 Berlin (Germany)

Main instructor: Stéphane Joost (EPFL)

One-week course for graduated researchers at Physalia courses (https://www.physalia-courses.org/) on the landscape genomics approach.

Design of webgis, assistant

2019- Lausanne (CH)

Teacher: Marc Soutter (EPFL)

Semester course for master students focusing on how to develop a web application containing geographic data.

Spatial statistics and analysis, assistant

2016-2018 Lausanne (CH)

Teacher: François Golay (EPFL)

Semester course for master students focusing on the understanding of statistical methods employed in spatial analyses.

Multivariate analysis, student assistant

2014 Lausanne (CH)

Teacher: Zoltan Kutalik (UNIL)

Semester course for bachelor students focusing on the understanding of multivariate statistics basis.

VOLUNTEERING and INTERSHIPS

Tropical marine conservation, intern

2016 Koh Phangan (Thailand)

Center for Oceanic Research and Education - south east Asia (CORESEA)

Two months internship on scientific diving and data collection for coral reef preservation and research.

Wildlife conservation, volunteer

2013 St. Eustatius (Dutch Antilles)

St. Eustatius National Park

One-month volunteering in support of the preservation of local terrestrial and marine wildlife.

SKILLS

Scientific programming

R | python | MATLAB | QGIS

Web design

HTML5 | Javascript | MongoDB | openlayers

Bioinformatics

DNA sequencing | RNA sequencing | RAD-sequencing

Experimental and molecular biology

DNA/RNA purification | Scientific diving

SCIENTIFIC PUBLICATION

Seascape genomics as a new tool to empower coral reef conservation strategies: an example on north-western Pacific *Acropora digitifera*.

2019 First author Article, Evolutionary Applications

https://doi.org/10.1101/588228

Sampling strategy optimization to increase statistical power in landscape genomics: a simulation-based approach.

2019 First author Article, Molecular Ecologies Resources

https://doi:10.1111/1755-0998.13095

Sex-specific changes in gene expression in response to estrogen pollution around the onset of sex differentiation in grayling (Salmonidae).

2019 First author Article, BMC Genomics

https://doi:10.1186/s12864-019-5955-z

Rapid identification and interpretation of gene-environment associations using the new R.SamBada landscape genomics pipeline.

2019 Co-author Article, Molecular Ecology Resources

https://doi:10.1111/1755-0998.13044

Indication of spatially random occurrence of Chlamydia-like organisms in Bufo bufo tadpoles from ponds located in the Geneva metropolitan area.

2019 Co-author Article, New Microbes New Infect.

https://doi:10.1016/j.nmni.2018.11.006

Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations

2017 Co-author Book chapter, Springer

https://doi:10.1007/13836_2017_2

Sex differentiation in grayling (Salmonidae) goes through an all-male stage and is delayed in genetic males who instead grow faster.

2017 First author Article, Scientific Reports

https://doi:10.1038/s41598-017-14905-9