Exploiting whole-genome sequencing to understand the evolution of corals and their symbionts

Thesis submitted by:

Jia Zhang

MSc (Genetics)

0000-0003-1161-4836

For the degree of Doctor of Philosophy

College of Public Health, Medical and Veterinary Sciences

James Cook University

Townsville, Queensland

Australia

May 2023

Statement of Sources Declaration

I declare that this thesis entitled “Exploiting whole-genome sequencing to understand the evolution of corals and their symbionts” is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Every reasonable effort has been made to gain permission and acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Jia Zhang
25 May 2023

Acknowledgement

It’s been a gratifying journey on this long yet fulfilling path to my PhD, which could not have been accomplished without my exceptional supervisors. First and foremost, I would like to express my deepest gratitude and appreciation to my primary supervisor, Ira Cooke, for his guidance, support, encouragement, and mentorship over the years. I will forever be grateful for your patience and friendship. I would also like to thank Professor David Miller for offering his valuable insights during my challenging moments. Additionally, my heartfelt thanks go to Dr Cheong Xin Chan for introducing the alignment-free method into this thesis and kindly providing me with a workspace in your lab.

I would like to thank the former and present members of the Marine Omics lab group, David Miller’s Lab, and codeR-TSV for the knowledge you shared, and the encouragement and support you provided. I would also like to thank Dr Zoe Richards for introducing us to the incredible Kimberley corals and for all the support and feedback along the journey. Also, I am grateful to Dr Line Bay for providing the data for this thesis. Also, my gratitude goes to my friends and colleagues in Townsville - Felicity, Beatriz, Legana, Casey, Ramona, Bruna, Sally, Sandlie, Refilwe, Ji, Jun, Sophie, and Chris, for making my experience here memorable.

I want to give special thanks to my family for always being there for me and letting me follow my heart all these long years. Thank you to my partner, Kexin, for the unflinching support you have given me for the past ten years, particularly throughout this PhD and the unique challenges that have come with it. I couldn’t have done this without you.

Finally, I thank everyone for empowering me to walk the path of science, never losing sight of the larger challenges in research.

Statement of the Contribution of Others

I performed all the data analyses and figure generation and drafted all chapters. I was also primarily responsible for the interpretation of the results, experimental design, and conceptualisation. The following tables show the persons and organisations who made substantial contributions to this thesis.

Nature of Assistance	Contribution	Titles, Names, and Affiliations of Co-Contributors
Intellectual support	Conceptual idea, supervision and editorial assistance	**Dr Ira Cooke** College of Public Health, Medical and Veterinary Sciences, James Cook University **Professor David Mille**r College of Public Health, Medical and Veterinary Sciences, James Cook University **Dr Cheong Xin Chan** College of Public Health, Medical and Veterinary Sciences, James Cook University
	Advice on Data Analysis	**Dr Ira Cooke** College of Public Health, Medical and Veterinary Sciences, James Cook University **Dr Cheong Xin Chan** College of Public Health, Medical and Veterinary Sciences, James Cook University
	Support for the publication (Chapter 3): Details are listed with each co-author	**Dr Zoe T Richards** School of Molecular and Life Sciences, Curtin University: Field work, study design and conceptualisation, editorial support, funding. **Dr Arne A S Adam** School of Molecular and Life Sciences, Curtin University: Early access to unpublished ddRAD data for checking mislabelled samples **Dr Cheong Xin Chan** The University of Queensland, School of Chemistry and Molecular Biosciences: Advice on the d2s method used to analyse symbiont profiles, editorial support, and access to supercomputer facilities. **Associate Professor Chuya Shinzato** Atmosphere and Ocean Research Institute, The University of Tokyo: Early access to genome data, editorial support. **Dr James Gilmour** Australia Institute of Marine Science, Indian Oceans Marine Research Centre: Editorial support. **Dr Luke Thomas** Australia Institute of Marine Science, Indian Oceans Marine Research Centre:

		Editorial support, assistance with DNA extractions **Professor Jan M Strugnell** Department of Marine Biology and Aquaculture, James Cook University: Editorial support and advice on population genetic methods **Professor David J Miller** College of Public Health, Medical and Veterinary Sciences, James Cook University: Editorial support, study design and conceptualisation, funding. **Dr Ira Cooke** College of Public Health, Medical and Veterinary Sciences, James Cook University: Editorial support, advice on data analysis, study design and conceptualisation, funding.
Financial support	Fee offset/waiver Stipend	James Cook University Postgraduate Research Scholarship
	Research funding	Australian Research Council Linkage Projects Higher Degree by Research Enhancement Scheme
	Computational infrastructure	James Cook University High-Performance Computing National Computational Infrastructure National Facility systems through the NCI Merit Allocation Scheme (Project d85)
Data collection	Field work/Lab work	**Dr Zoe Ricards** School of Molecular and Life Sciences, Curtin University **Dr Line Bay** Australian Institute of Marine Science, Townsville, Australia

General abstract

Coral reefs are diverse ecosystems with exceptionally high economic, cultural and intrinsic value, however, they are increasingly under threat due to climate change. Studying the genetics of corals provides a powerful tool that can be used to design, risk-assess, and monitor coral reef management strategies. It also improves our basic understanding of coral biology by providing insights into the molecular mechanisms of coral adaptation. Current population genomics studies integrated with high-throughput sequencing methods and computational analyses are greatly expanding our understanding of coral speciation and evolution. Importantly, the findings of coral population genetics studies indicate the current genetic diversity of corals and suggest their evolutionary responses to past climate change which inform us about their further survival strategies. This thesis uses whole-genome sequencing to conduct in-depth studies of coral reef biodiversity, the impacts of climate change on coral populations, and the development of innovative techniques for investigating the genetics of coral-associated Symbiodiniaceae, focusing on *Acropora* species.

In Chapter 2, I analysed shallow-coverage whole-genome sequencing data of 228 *Acropora tenuis* colonies across 330km in the central Great Barrier Reef and explored their genetic diversity patterns. The results demonstrate a strong divergence between Magnetic Island and other reefs yet minor differentiation between inshore and offshore. The lack of genetic differentiation between north inshore and offshore reefs across the majority of the genome is contrasted by four genomic regions with genetic patterns shaped by strong population structure. The individual heterozygosities of each group within the four locus indicates chromosomal inversions. This discovery highlights chromosomal inversions as a significant structural variation in coral evolution since they strongly suppress local recombination. Additionally, this research underscores the role of population genomics in detecting inversions in non-model species without high-coverage sequencing data or long-read sequencing.

In Chapter 3, I leverage whole-genome sequencing technologies to understand the long-term effects of climate change on coral reefs, examining the response of *Acropora digitifera* to past climatic events. Genome-wide variations reveal three distinct clusters in northwestern Australia. Their rapid divergence since the last glacial maximum is associated with founder effects and restricted gene flow. Combined with two methods for detecting signatures of selection and simulated genomic data, I identified selective sweeps across the genome which are too strong to be explained by demographic history alone. Interestingly, these

signatures overlap with genes that show different patterns of functional enrichment between inshore and offshore habitats, suggesting the different molecular modifications based on the local environment. Importantly, a potential genetic marker was specifically identified in the inshore population as likely to have undergone strong selection associated with the adaptation of inshore reefs to climate change since the last glacial maximum.

In Chapter 4, I introduce an alignment-free approach for investigating the genetics of coral-associated Symbiodiniaceae. This method applies *k*-mer-based D_2^S statistics directly to short reads of symbiont origin in whole genome sequencing datasets of corals. This tool can resolve the genetic differentiation of symbiont reads in coral samples without the need for reference genomes or deep coverage. When all samples are dominated by a single genus, the result provides within-genus (species to population resolution) variations of symbiont communities in corals. This chapter underscores the roles of symbionts in coral differentiation and adaptation and proposes a new method to be integrated into data analysis of future coral genomic studies.

Together, this thesis expands our knowledge of coral biodiversity and evolution in different reef systems. Firstly, this thesis attempts to link chromosome inversions to population structure in coral and adapting genomic methods originally designed for humans to non-model organism studies. The results indicates that ocean geomorphology, demographic history, and local adaptation drive spatial genetic structure in population of spawning coral with pelagic dispersing larvae. Finally, the findings provide insight into the future adaptation of corals by assessing the current genetic diversity and detecting genetic markers that are contributing to coral resilience.

Table of Contents

List of Figures

CHAPTER 1

General Introduction

Coral reefs are among the most biodiverse ecosystems on earth and provide ecosystem services in the form of coastal protection, fisheries and tourism to millions of people around the world. Coral reefs face recurrent natural disturbances, such as tropical cyclones, heatwaves, and outbreaks of the crown of thorns starfish (Nyström et al., 2000). As a result of global warming, marine heatwaves are now recognised as one of the most severe threats to reefs globally (Spalding & Brown, 2015). This is because, under heat stress, the symbiosis of the coral host and its photosynthetic dinoflagellate will break down which results in coral bleaching, a state in which corals are more susceptible to disease and mortality (L. Hughes et al., 2018). With global warming leading to an increase in the frequency and severity of coral bleaching events, the future state of coral reefs will depend on the ability of species and populations to adapt (Colton et al., 2022; Matz et al., 2018). Understanding whether this will occur, how rapidly, and if human interventions can help is now a major focus of research, which includes expanding knowledge of the genetic diversity and connectivity of coral populations, as well as their demographic and adaptation histories.

Reef-building corals (order Scleractinia) are an ancient and species-rich taxonomic group (DeVantier et al., 2020) that includes major radiations at both deep and shallow time scales (Quattrini et al., 2020). Two major groupings exist within the Scleractinia, the complexa and robusta (Dustan, 1996; Stolarski et al., 2011), but fossil evidence suggests that significant morphological variation existed prior to this major divergence. Among extant corals, many cryptic species have recently been identified reflecting habitat differences, historical disturbances, reproduction, and interplay of species (Ladner & Palumbi, 2012; Matias et al., 2022; Rosser, 2015; Warner et al., 2015). In one recent example, three cryptic coral species were resolved using genomic data despite many of them being indistinguishable due to ancient morphological stasis (Bongaerts et al., 2021).

Molecular evidence combined with fossil records is revealing the evolutionary history of corals in increasing detail, providing a link between coral diversity and past climate change (Prada et al., 2016; Santodomingo et al., 2015). At deep evolutionary timescales (tens to hundreds of millions of years) past shifts in Earth's climate are thought to have caused massive coral extinctions as well as bursts of diversification, possibly reflecting increased niche availability (Prada et al., 2016; Quattrini et al., 2020). The Acroporidae is a particularly

diverse family of corals that may have benefited from the onset of glacial cycling as they are fast-growing and able to rapidly colonise new habitats. One study suggests that gene flow between acroporid species occurred during ancient diversification events, a phenomenon that could potentially promote the exchange of adaptive traits (Mao et al., 2018). A more recent study comparing eighteen acroporid genomes suggests that they shared a conserved gene repository despite a deep divergence back to the Eocene and Oligocene when the ocean was much warmer than today (Shinzato et al., 2020).

Studying the complete sets of genes encoded by genomes of different species can provide insights into molecular innovations or constraints that have affected the distinct evolutionary histories of different lineages. For example, a complete histidine biosynthesis pathway was found in one of the major clades of scleractinian corals, Robusta, but not in Complexa and also outgroup species sea anemones (Ying et al., 2018). This unique presence of a functional pathway suggests Robusta may retain an additional pathway for nutrition.

Comparative genomics provides insights into the evolutionary history of different coral lineages, resolves phylogenetic relationships, and helps clarify gene functions. However, the timescale of evolutionary events separating distinct coral lineages (species, genera, and higher taxa) is typically greater than tens of millions of years (Mao et al., 2018; Shinzato et al., 2020). Understanding how corals responded to more recent extreme climatic events, such as the last glacial maximum, and how they have adapted to live in different local environments requires tools that examine differences between groups of closely related individuals (typically within a species). One such tool is population genomics, which focuses primarily on the study of allele frequencies within different operating units of a species. Studying allele frequency shifts can provide insights into the relative roles of mutation, drift, migration, and natural selection. Coral population genomics also provides estimates of the genetic composition and migration of local populations and can find potential genetic factors linked to phenotypic differences which in turn provides valuable guidance to conservation and restoration (Pinsky et al., 2023). Since allele frequencies are also sensitive to demographic changes, they can be used to reconstruct demographic histories, thereby furthering our understanding of past climate change effects on population size, divergence, and migration (Gutenkunst et al., 2009). This in turn helps provide predictive insights into future coral adaptation to global warming.

The development and advancement of genomic sequencing is revolutionising life sciences disciplines, especially molecular biology and genetics. Since their emergence in the late 1970’s, sequencing technologies have sustained exponential growth in throughput. Falling prices and increasing throughput largely empower the population genomics study with

whole-genome data. While traditional population genetic approaches using single or very few genetic markers have revealed hidden population structures for many marine organisms (Teske et al., 2015), inconsistent results were obtained across markers, possibly due to factors such as incomplete lineage sorting. An equally important limitation of single-marker techniques is that they lack the ability to accurately infer demographic history and identify non-neutral differentiation from the background (Xu et al., 2017; L. Zhao et al., 2021). The term “population genomics”, which emerged with the advancement of genomic sequencing technology provides a new framework to understand the evolution of populations (Combosch & Vollmer, 2015; Hancock-Hanser et al., 2013; Shinzato et al., 2015; S. Wang et al., 2012). With the potential to study the change of numerous genomic loci or the whole genome, population genomics is used to investigate unbiased genome-wide effects of evolutionary processes (such as bottlenecks and hybridisation) and to identify genomic regions associated with adaptation or selection.

One of the advantages of sequencing-based molecular techniques is that they are often not taxon-specific. Thus, although many of the advances in the field have been driven by research in Humans and model species, there have also been large gains in other taxonomic groups. This has been the case for research on the molecular and evolutionary biology of corals where other (non-sequencing-based) techniques are often extremely impractical due to the great difficulty and long timescales involved in keeping corals in the lab. Through sequencing corals, we are gaining much more knowledge about how they evolve, differentiate, and respond to past and present climate changes.

In this chapter, I will first introduce the current genomic approaches and short-read sequencing strategies in population genomic studies, and their respective advantages and limitations. Then I will illustrate their usages in population genomic studies with a focus on coral reefs. I will also introduce two important reef systems that we studied in this thesis: the Great Barrier Reef and the Kimberley Reefs from Northwestern Australia. Finally, I will discuss the potential value of unmapped reads in population genomic studies which are usually discarded, especially for coral holobionts.

1.1 Genomic approaches in population genomics

Until recently, population genetic studies of corals have most frequently been conducted with few markers from a very limited proportion of the genome, such as mitochondrial loci (Palumbi et al., 2023), microsatellites (Nakajima et al., 2017; Sturm et al., 2020), and a single nuclear gene (Flot et al., 2013; Souza et al., 2017). Genetic inference based on single or few genetic markers has limited power across many types of analyses simply due to the

reduced number of independent loci available. It is unable to distinguish between processes that occur genome-wide (often neutral processes) and those that are restricted to certain loci such as selection. Individual markers may also lead to erroneous inferences of population or species divergence due to incomplete lineage sorting whereas this effect can be easily measured and accounted for in genome-wide datasets. High-throughput sequencing technology has dramatically increased our ability to obtain polymorphism data simultaneously across many genome-wide loci (Ekblom & Galindo, 2011). This has led to the development of population genomic approaches that are based on tens of thousands of genomic single nucleotide polymorphisms (SNPs) allowing more unbiased statistical inferences on the population structure and connectivity of marine species (Combosch & Vollmer, 2015; Reitzel et al., 2013). In this section, I will focus on techniques available for generating genome-wide SNP data. These techniques are all currently in widespread use and exist along a spectrum of tradeoffs in cost, complexity and potential for bias.

1.1.1 Reduced-representation genome sequencing

The development of restriction site-associated sequencing (RAD-seq) approaches has revolutionised population genetic studies with the ability to genotype numerous markers across the genome using a single cost-effective protocol. All RAD-seq methods rely on the use of restriction enzymes to cut genomic DNA semi-randomly across the entire genome while doing so consistently and reproducibly so that the same set of loci can be sequenced across many samples (Figure 1.1). RAD-seq methods include variations using a different or different number of restriction enzymes, combined with size selection techniques during library preparation. Although these methods differ in cost, the number of loci and degree of bias, they all provide an inexpensive way to randomly target genomic regions across an entire genome (Andrews et al., 2016; Peterson et al., 2012). RAD-seq methods are generally inexpensive because they generate information from only a subset of the genome, and broadly applicable because they operate without the necessity to have prior information on the genome of the target species. Thus, RAD-seq methods are widely used in ecological and evolutionary studies of non-model organisms (Andrews et al., 2016).

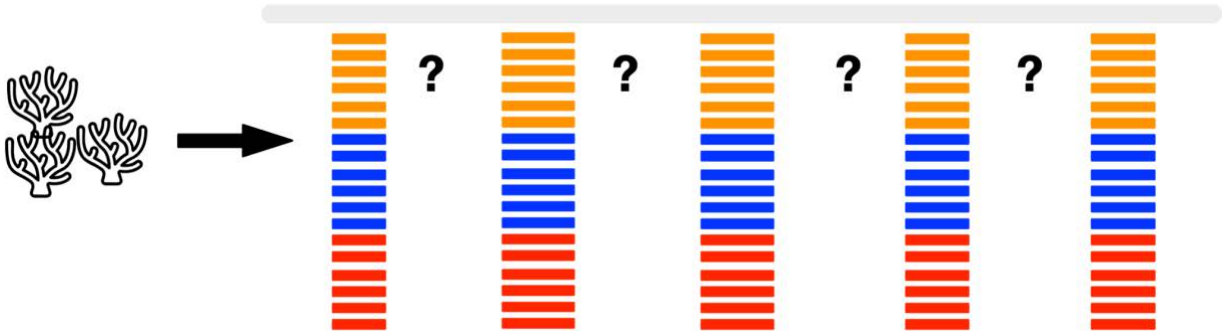


Figure 1.1 The illustration of RAD sequencing, short segments with colors represent the distribution of sequenced reads from each sample of populations. Without an available reference assembly, the coordinates of each segment and the distance between them can be unknown.

The power of RAD-seq for population genetic studies has been demonstrated for answering questions about genomic adaptation, population structure, genetic diversity, and phylogenomics, among others (Burgon et al., 2020). It is more likely to identify the true population structure than a single gene since more genetic markers better capture the neutral genetic background among populations and because overall signatures of divergence can be measured even when some loci may exhibit incomplete lineage sorting. Furthermore, RAD-seq can also be applied as genome-wide sequencing in capturing data from coral holobiont including symbiotic anthozoans. For example, completely overlapped genetic statistics were observed when RAD-seq data of symbiotic sea anemones were compared against the simulated aposymbiotic datasets (Titus & Daly, 2022).

While RAD-seq largely sequences a random subset of the genome, target-capture sequencing methods overcome some of the biases from RAD-seq by sequencing the same subset of DNA across genomes (Jones & Good, 2016; McCartney-Melstad et al., 2016). Depending on the specificity of the probe design, targeted sequencing enables the inference of deep evolutionary history (Grinblat et al., 2021; Zayasu et al., 2021). Meanwhile, the feature to capture homologous sequences across genomes contributed to the broad usage of target-capture sequencing in phylogenomic studies, because the conserved regions contain insufficient informative variants to infer population history at recent time scales (Manthey et al., 2016; B. T. Smith et al., 2014).

Transcriptome sequencing (RNA-seq) is also a key reduced-representation sequencing strategy used in population genetic studies of non-model organisms (Popovic et al., 2020). Like RAD-seq, RNA-seq does not necessarily require a reference genome. It only sequences the fractions of transcribed regions (i.e. expressed transcripts) in the genome, thus the evolutionary inferences are mainly based on functional genes. Since the targeted regions contain protein-coding genes which can be searched against the existing genomic

database for functional annotation, it can provide a stronger biological context for variants (Ekblom & Galindo, 2011). However, the complexity of transcript splicing in eukaryotic organisms and the large variation in gene expression limit the power of RNA-seq in population genomics (Andrews & Luikart, 2014; Todd et al., 2016). Costs of population RNA-seq are also high, since deep sequencing is required to obtain sufficient coverage across many loci, and the cost of preparing libraries for sequencing is typically much higher than for DNA-based approaches. It is also difficult to get accurate functional information for non-model species such as cnidarians, simply based on sequence similarity.

Compared with whole-genome sequencing, reduced-representation genomic sequencing allows low-cost sequencing at high coverage even without access to the reference genome and largely reduces the computational complexity. However, the limitations of reduced-representation genomic sequencing like RADseq are apparent. In addition to the sequencing and genotyping errors that come with all second-generation sequencing methods, RADseq methods have some unique limitations that could bring bias and error to the analysis (Andrews et al., 2016; Puritz et al., 2014). Allele dropout happens when a restriction enzyme recognition site fails to detect the right cutting location due to polymorphism and causes genotyping errors (Gautier et al., 2013). This could bias population genetic statistics like genetic diversity and F_{ST} since it may result in a heterozygous genotype being called as a homozygous site. Also, sequencing library preparation of RAD-seq methods tends to produce more PCR duplicates resulting in higher genotype uncertainty, an issue that is particularly problematic for inferences relying on high certainty in genotypes at single loci (Andrews et al., 2014). To mitigate these caveats, specific tools and improvements in both library preparation protocols (Peterson et al., 2012; Toonen et al., 2013; S. Wang et al., 2012) and the bioinformatic analysis pipeline have been developed (Melo & Hale, 2018; Rochette & Catchen, 2017). Despite the fast improvements in sequencing technology, costs, and more accessible genome reference sequences, RAD-seq methods currently are still essential tools for studying the genomics of natural populations with their good balance in flexibility and cost-effectiveness.

1.1.2 Whole-genome sequencing

With the generation of genome assemblies for non-model organisms no longer being impeded by sequencing cost or computational approaches, whole-genome sequencing (WGS) of a population of individuals is becoming a viable option for the generation of genome-wide variant information. While sequencing the whole genome of individuals at high coverage is the “perfect” situation for population genomic data analysis, sequencing costs make it prohibitively expensive, especially for organisms with large genomes (Cheng et al.,

2014). To work around this, pooled and low-coverage approaches have been developed in an attempt to obtain some of the benefits of whole genome sequencing (low bias, dense marker coverage) at reduced costs. Sequencing the whole genome of pools of individuals (Pool-seq) is perhaps the most cost-effective way to obtain whole genome allele frequency information for a large number of individuals (Ferretti et al., 2013; Schlötterer et al., 2014). DNA is pooled prior to library preparation, thereby eliminating one of the major costs of sequencing. However, it is recommended that many individuals are sequenced in each pool to reduce biases from unequal representation across samples (Ferretti et al., 2013; Schlötterer et al., 2014)). Although the information on individual samples is discarded in Pool-seq, it remains an attractive approach because many population genetic inferences can be made purely based on the variant number, position and allele frequency in populations. Results from a Pool-seq data study revealed high consistency of population diversity estimates and minor allele frequency with sequencing data of individuals at a much lower cost per individual (Kurland et al., 2019; Schlötterer et al., 2014; Zhu et al., 2012). However, pool-seq has several limitations. Since it needs to be performed on a large sample size (>40 individuals) it has limited utility for species that are endangered or difficult to collect (Futschik & Schlötterer, 2010; Zhu et al., 2012). Pool-seq also has difficulties in differentiating low-frequency alleles from sequencing errors, equal sequencing representation of individuals in the pool, and identifying alignment errors based on coverage variation (Anand et al., 2016). Finally, the design of pool-seq experiments must be informed by prior knowledge of population structure since pools should not include individuals from multiple populations which makes it difficult to identify substructures in taxa with prevalent cryptic diversity, such as corals. In conclusion, the success of a pool-seq experiment is especially dependent on a high-quality sampling design and sequencing, as well as a stringent data quality control (Anderson et al., 2014; Fuentes-Pardo & Ruzzante, 2017; Kofler & Schlötterer, 2014).

Given the downsides of Pool-Seq, low-coverage whole-genome sequencing (lcWGS) overcomes some of the limitations and is also a popular strategy to retrieve complete genomic information with limited budgets (Therkildsen & Palumbi, 2017). Studies based on both simulated and empirical data suggest that sequencing a large number of individuals at low coverage has greater power in population genetic parameter estimates than sequencing a few samples at high depth (Buerkle & Gompert, 2013; Fumagalli, 2013; Y. Li et al., 2011). The lcWGS compromises sequencing depth while maintaining the ability to obtain genome-wide data at the individual level. Although costs are typically higher than RAD-seq or Pool-seq, the gap may be small, especially for organisms with small-sized genomes (Lou et al., 2021).

The sequencing depth of lcWGS is usually <5x per individual, meaning that most loci genome-wide are unlikely to be accurately genotyped. Thus the genetic analyses of lcWGS data are based on a statistical framework (probabilistic genotype calling) that accounts for the uncertainty in genotypes (Buerkle & Gompert, 2013; Korneliussen et al., 2014; Nielsen et al., 2011). Different models are available to estimate the genotype likelihood (GL) of each locus based on the reads covering the site, read mapping qualities and read base qualities (Korneliussen et al., 2014). In addition, these GLs are the basis of downstream population genetic analysis that does not rely on genotype calling or single nucleotide polymorphism (Fumagalli et al., 2013, 2014). Therefore, the use of lcWGS data is limited to population genetic tools designed to work with genotype likelihoods rather than called genotypes. Furthermore, genotype likelihoods are sensitive to read qualities, depth and statistical models, thus we need to be extra cautious to sequencing errors and recalibrate read base quality (Lou & Therkildsen, 2022).

In studies where sufficient funds are available, it may be possible to sequence at a higher depth per individual, which provides additional benefits compared with the lcWGS approach. The starting point for such studies is around 10x sequencing depth as this is able to cover more than 99% of the whole genome and will provide accurate genotypes across the majority of loci (Ellegren, 2014; Jiang et al., 2019). With the provision of genome-wide genotype data at the individual level, this approach is the most comprehensive for population genomics. Although the optimal sequencing depth per individual varies depending on the species lineage and genome complexity, most whole-genome re-sequencing projects of natural populations generated 10-20x data to ensure the quality and completeness of variants (Fuller, Mocellin, et al., 2020). Studies with whole-genome re-sequencing data overcome the limitations of RAD-seq methods, pool-seq and shallow whole-genome sequencing methods with both fine-scale genetic markers and accurate individual genotypes. For example, the demographic history can be reconstructed with the whole genome polymorphism data of one or few individuals based on the theory of the Sequentially Markovian Coalescent (SMC), such as PSMC (H. Li & Durbin, 2011), MSMC (Schiffels & Durbin, 2014) and SMC++ (Terhorst et al., 2016).

Genome-wide genotype data of populations also provide access to haplotype data through computational phasing (S. R. Browning & Browning, 2011). Haplotype data are the linkage information of alleles along the individual chromosomes and encode information on both mutation as well as recombination. For natural populations with no pedigree information, tools like Beagle (B. L. Browning et al., 2021) and SHAPEIT (Delaneau et al., 2013) can apply algorithms to obtain phasing data using unrelated individuals in the population. The

“horizontal” signals from haplotypes have been shown to be powerful in characterising the genomic local ancestry (S. R. Browning et al., 2018) and in detecting recent selective sweeps (Simonson et al., 2010; J. Zhang et al., 2022; Zheng et al., 2020). Furthermore, recent adaptation leaves genomic regions with extended homologous haplotypes that can be directly identified using phased variant data via extended haplotype homozygosity (EHH) and related statistics. Although most methods to calculate EHH statistics require phased data, this is not strictly necessary if sample sizes are large (>200) (Klassmann & Gautier, 2022).

1.1.3 The importance of high-quality reference genomes

Apart from the benefits of sequencing strategies discussed above, the power of population genomics for natural populations will also largely expand with a high-quality reference genome. There are two main reasons for this: (1) genome re-sequencing data is largely constrained to identifying variants within the reference assembly, and (2) the spatial arrangement of variants (genomic context) is often an important factor affecting their interpretation in relation to both neutral and adaptive processes. However, many eukaryotic genomes are characterised by complex repetitive structures which are difficult to sequence and assemble (Treangen & Salzberg, 2012), especially large plant genomes (Pellicer et al., 2010) or polyploid genomes (Peng et al., 2022). Even for humans, a complete characterisation of the whole genome sequence was only generated recently (Nurk et al., 2022; Schneider et al., 2017; T. Zhao et al., 2020). At least in part as a result of these difficulties, a significant proportion of mutations in the genome are ignored in current non-model organism genomic studies, potentially leading to biased statistical conclusions (Domanska et al., 2018). Therefore, in the genome-wide scan of selective sweeps, a missing selection signal could actually mean an unmapped selection signal, especially for those highly polygenic adaptations.

Chromosome-level assemblies can be generated using a combination of short- and long-read sequencing technologies combined with a Hi-C scaffolding approach to improve the contiguity of the assemblies (Totikov et al., 2021). A chromosome-level assembly or a genome with a genetic linkage map or physical map can be extremely useful for evaluating the genetic statistics across the genome along each chromosome with a Manhattan plot. With locations and distances among loci being visualised, clustered signals like selective sweeps can be easily picked up as the act of evolutionary forces on a gene or a locus will also skew the allele frequencies of the neighbouring loci due to a process referred to as genetic hitchhiking (Luikart et al., 2018; McKinney et al., 2016). For example, genomic islands or regions that are highly differentiated between two closely related species or

populations due to gene flow barriers or selection may be present in the genome as blocks (Malinsky et al., 2015), and therefore only detectable if the underlying reference assembly is sufficiently contiguous. Information on the arrangement of genetic loci is therefore particularly valuable in distinguishing clustered signals from evolutionary processes of interest against the background of noise.

The recombination rate is a critical parameter for making inferences in evolutionary and genetic studies. Although genetic linkage maps contain information on the recombination frequency of genome fractions, a high-density recombination map of a species provides a novel tool to map traits to genotypes and assist breeding in the future (Luo et al., 2020). The variation in recombination rate across the genome effectively shapes the local genomic diversity. Models that incorporate recombination rate and LD information are able to simulate and reveal the demographic history in unprecedented detail (Excoffier et al., 2013; Terhorst et al., 2016). Recombination rate also provides good guidance for an appropriate choice of window size in genome-wide window-based scans to ensure the best statistical power. Furthermore, the precision and power of methods for detecting selective sweeps based on linkage disequilibrium, like iHS/EHH, depend on background recombination and mutation rates (Lotterhos, 2019).

Lastly, the known gene functions are the final stepping stones for us to understand the biology of genetic changes. Genomic studies are able to identify thousands of traits associated with genetic variants or highly expressed genes, however, half of which often do not have an annotated function. While building model organism systems and making use of CRISPR/Cas 9 gene editing tools can be effective to test gene function hypotheses, the analysis of gene co-expression networks is proven useful in linking ‘dark genes’ to gene networks with known functions (Cleves et al., 2019).

1.2 Population genomic studies of coral reef taxa

Genomic tools bring huge improvements in the accuracy, convenience and scale of traditional population genetic studies while focusing on the most long-standing subjects that are tightly linked to ecology, evolution, and conservation biology. By studying the basics of genetic composition in groups of individuals, we gain fundamental knowledge of the genetic status of the population and insights into the mechanism of evolution. While the conceptual frameworks of population genetics have been widely used in studying terrestrial organisms (Mattle-Greminger et al., 2018; Qiu et al., 2012; Salojärvi et al., 2017; S. Zhang et al., 2020), big knowledge gaps exist in marine systems simply because species are more difficult to observe, obtain, or trace in the complex marine environment. Genomic tools are starting to

reveal the genetic diversity of marine organisms and define the almost invisible boundaries and connectivities in the ocean through genetic differentiation and gene flow. Here I review some studies using a genomic approach to answer key questions in the population genetics of shallow coral reefs which is also the focus of this thesis.

1.2.1 Genetic variation and diversity

Genetic diversity captures both the number of polymorphisms and their frequency in a population. Metrics of genetic diversity such as individual heterozygosity, expected heterozygosity and nucleotide diversity (π) are important parameters for coral reef conservation management and restoration as they can be indicators of adaptive potential, or provide warning signs for populations that are isolated, in decline, or that have experienced a severe bottleneck. The quantity of overall genetic variation in the population can also be used to infer effective population size (N_e) which is a metric routinely used in the assessment of marine populations for conservation risk management (Hare et al., 2011). Recent study found that coral can maintain high level of genetic diversity despite of severe bleaching and cyclones (Underwood et al., 2018). However, since many coral species have large ranges and harbour very large amounts of genetic diversity these measures may be of limited value. Despite dramatic (>90%) recent declines in *Acropora* throughout the Caribbean, genetic diversity estimated with data from Genotyping by Sequencing (GBS) remained high (Drury et al., 2016).

1.2.2 Coral population structure

Another important focus of population genetics is examining how population structure, which is the spatial distribution of genetic variation, aligns with geographic locations and dispersal patterns. For many years the dominant paradigm for marine species like corals with pelagic dispersing larvae was that they should be well-connected and therefore have little population structure (Palumbi. 2003; Cowen and Sponaugle 2009). Recent studies adopted genome sequencing and revealed complex factors drive strong population structure within small spatial scales (Matias et al., 2022; Teske et al., 2015; Thomas et al., 2020). Thus, a thorough investigation of population structure within samples is needed before undertaking any analyses or making management decisions.

Population structure is often quantified by the summary statistic, F_{ST} or fixation index (Holsinger & Weir, 2009) which measures population differentiation. Low F_{ST} values among populations generally suggest they share a well-mixed gene pool while high values indicate differentiation, either through neutral or adaptive processes. There are no strict criteria of F_{ST}

to define population structure as the value also depends on within-population diversity (Meirmans & Hedrick, 2011) making it difficult to compare across systems. In addition to F_{ST}, further information on population structure is often gained through tools developed to assign individuals to genetic clusters and to visualise the degree of admixture within individuals. These tools include principal component analysis (PCA) based methods and Bayesian model-based clustering methods (STRUCTURE).

Population genomics tools provide increased power to delineate populations, quantify genetic differentiation and detect sympatric cryptic diversity, which is common in coral species. For example, the population structure in corals previously undetectable with fewer markers such as microsatellites and allozymes was revealed by using a large number of genetic markers (Crawford & Oleksiak, 2016; Drury et al., 2016). It highlights the importance of natural selection in shaping population-specific genetic variation in different locations. Similarly, unique genetic discontinuity in corals was observed in the Arabian Peninsula where the sea area is bounded by the Gulf which is characterised by extreme environmental conditions like high water temperature and salinity, and the Sea of Oman which is connected to the ocean (Howells et al., 2016; Torquato et al., 2022). Furthermore, a combination of genomic and seascape data which is equivalent to landscape genomics in marine systems can provide a better understanding of the association between abiotic and biotic factors in coral population structure establishment (Riginos et al., 2016; Underwood et al., 2020).

Understanding the patterns of gene flow between populations is fundamental to coral reef conservation and management. Maintaining natural patterns of gene flow through targeted management efforts may be an important pre-condition for evolutionary adaptation (Colton et al., 2022). Methods that artificially assist gene flow such as transplants are sometimes used in an attempt to increase coral abundance on degraded reefs (Rinkevich, 2008; Young et al., 2012), and may also play a role in enhancing adaptive potential (Colton et al., 2022). Assisted gene flow approaches must be considered with caution however as they might cause problems including reduced genetic diversity, outbreeding depression, and genetic introgression (Baums 2008). Thus, surveys of gene flow across habitats are critical to safe and effective restoration. With the advent of next-generation sequencing tools, genome-wide surveys of single-nucleotide polymorphisms were carried out to study the threatened Caribbean elkhorn coral, *Acropora palmata*. Fine-scale population genetic structures were dissected, and the major barrier to gene flow was addressed (Devlin-Durante and Baums 2017). The first coral population genomic study with whole genome sequencing data was performed in *Acropora digitifera* populations in the southern Ryukyu Archipelago (sRA). It successfully resolved high-resolution population structure and revealed complex migration

patterns in the sRA (Shinzato et al. 2016b). Furthermore, in another study using sequences of microsatellites of *Acropora tenuis* populations across the south and north Ryukyu Archipelago, an unexpectedly complex migration pattern was observed despite the strong Kuroshio current (Zayasu et al. 2016). Eventually, the conservation plans for establishing marine protected areas (MPAs) and coral restoration can be performed effectively based on the dispersal patterns and local structures.

1.2.3 Coral demographic history

Population demographic history reflects the effective population size changes from past to present as well as processes of divergence and introgression affecting multiple populations. Demographic history shapes the genetic diversity of a population and can reflect the speciation process. When calculated genome-wide, summary statistics such as Tajima's D (Tajima, 1989) can be interpreted as indicators of demographic change. A much more detailed understanding of demographic history can be obtained through statistical tools based on coalescent theory, and that match theoretical expectations with observed data on genome-wide patterns of heterozygosity, allele frequency and recombination. One such class of statistical tools is based on the site-frequency spectrum (SFS), which is essentially a histogram of allele frequencies, potentially having multiple dimensions in situations where more than one population is being modelled. Tools like fastsimcoal2 (Excofffier et al., 2021) and ∂a∂i (Gutenkunst et al., 2009) can make use of the SFS and compare it with expected values under different demographic scenarios (obtained via explicit (fastsimcoal2) or heuristic (∂a∂i) coalescent processes) capturing effective population size changes, divergence, and migration.

Glacial cycles are likely to have been a major driver of demographic change in the marine environment as they include changes in the amount of shallow-water habitat (Ludt & Rocha, 2015) and ocean temperature (Bintanja & Wal, 2008). As seawater temperature and sea level changed periodically with these glacial-interglacial cycles, many taxa likely went extinct while opportunists took over the new niches. *Acropora* spp. are thought to be one of the major “winners” surviving the disruptive effects of these glacial cycles. This highly diverse genus includes some of the most common and prolific stony corals around the world, a fact that has been attributed to their high growth rate and a short time to reproductive maturity which is likely to have been an advantage during periods (Pliocene and especially the Quaternary) with massive sea-level fluctuations (Renema et al., 2016). The ability to link their demographic history to geological events allows a better understanding of coral evolutionary history. For example, the demographic modelling of multiple *Acropora* corals suggests a population expansion that began from 2 Mya, which coincides with the onset of

quaternary glacial cycles and the mass extinction of many other coral species (Mao et al., 2018). During the same period, population declines were observed in another group of reef-building corals in the genus Orbicella (Prada et al., 2016).

Similarly, demographic modelling seems to be able to describe the divergence and migration between coral reefs quantitatively and directionally. For example, demographic modelling with the diffusion approximation based on SFS allows us to treat connected coral reefs as separate demographic units and estimate the directionality of their historical migration. This provides support for the theory that corals along the GBR are adapting to heat stress by exchanging genetic variants (Matz et al., 2018). Furthermore, the ongoing divergence within coral species in different ecological environments was inferred using the improved diffusion approximation method Moments (Jouganous et al., 2017), which provides genomic insights into the coral speciation process (Prada & Hellberg, 2021; Rippe et al., 2021).

1.2.4 Adaptation to climate change

Simulations based on current population genomic data and future ocean temperature parameters suggest that the adaptive evolution of corals may keep pace with climate change under low carbon dioxide emission scenarios (Bay et al., 2017). This is contingent on findings in several coral species that variation in coral heat tolerance has a strong genetic basis (Dixon et al., 2015; Fuller, Mocellin, et al., 2020; Quigley et al., 2020) and the fact that high effective population sizes contribute to large stocks of standing genetic variation in many coral species, providing the raw material for this type of rapid adaptive response to changing environments (Torda et al., 2017). In a structured population, the geographic distribution of standing genetic variation may include higher frequencies of heat-tolerant alleles in hotter areas. Thus, adaptive variations combined with reef regional connectivity are key parameters used to predict coral heat resilience and to guide coral conservation actions, such as assisted gene flow (AGF) (Matz et al., 2020; Quigley et al., 2019).

Genetic simulations attempting to predict the effectiveness of conservation strategies such as AGF depend on the accurate identification of the genomic variants associated with heat tolerance. Genetic signatures in locally adapted heat-tolerant populations resulting from past selection provide a method for mapping these loci. Population genomics makes it possible to directly identify these based on patterns of allele frequencies and linkage disequilibrium resulting from selective sweeps. However, a major challenge with this approach is that similar signatures can arise through neutral demographic processes. Identification of genuine selective sweeps therefore also often involves building accurate neutral models as null hypotheses against which statistical measures indicative of selection can be compared.

Cooke *et al.* recently used this approach to look for signatures of selection in a coral population and used a composite likelihood ratio (CLR) ratio statistic implemented by SweepFinder2 (DeGiorgio et al., 2016) to scan for selective sweeps while controlling the false positive rates with simulated demographic models (Cooke et al., 2020). A strong signal was located in a region with tandem arrays of epidermal growth factor (EGF) domain-containing genes. This study illustrates the power of dense allele frequency information which allowed the CLR statistic to have much greater power compared with an F_{ST} outlier scan.

Despite the high potential for gene flow in broadcast spawning coral populations, many studies have identified candidate loci for adaptation to the local environment (Palumbi et al., 2014; Dixon et al., 2015; Fuller et al., 2019; Cooke et al., 2020; Howells et al., 2021; Smith et al., 2022; Zhang et al., 2022). Sometimes, this is even observed in habitats with fine-scale marginal environmental pressure, like lagoons and slopes in close physical proximity (Thomas et al., 2022). Alleles with highly divergent frequencies under high gene flow may therefore be inferred to confer a differential advantage to coral recruits in alternate habitats. If temperature is a major differential between habitats this provides evidence for the association of specific alleles with heat tolerance. Genes involved in DNA damage response generated by oxidative stress and UV-light damage were also detected as outliers of genome-wide F_{ST} values between *Acropora hyacinthus* sampled from low latitude and high latitude (J. Fifer et al., 2021) and some of these candidates were also identified as adaptive genes associated with heat waves in coral *Acropora digitifera* (Selmoni et al., 2020). In the Southern Persian/Arabian Gulf (PAG), a specific long homologous haplotype was detected in heat-tolerant brain coral *Platygyra daedalea* population compare with their ocean-side neighbours (E. G. Smith et al., 2022). These regions involve genes with methylation level changes in thermal and light stress experiments.

Other than the scenarios in which directional selection generates strong regional differentiation and reduced nucleotide diversity, the effect of local adaptation can be eroded by high gene flow. Thus, it is important to understand the different genetic responses in coral reefs when facing gene flow. A recent study identified a signature of balancing selection in co-chaperone of the Hsp 70 that persisted long in the population while strong gene flow homogenised genomes across different latitudes of the GBR, (Fuller, Mocellin, et al., 2020). Despite some great strides in understanding the genetic basis of adaption in corals, the wide distribution and diverse coral species usually have their own trajectories of how they evolve with past and current climate changes (Ayre & Hughes, 2004; Burgess et al., 2021; Pinsky et

al., 2023). Knowledge gaps exist in finding effective gene markers and verifying the function of long-list gene candidates for adaptation.

1.3 Corals in natural extreme environments

Although in vitro bleaching experiments provide fundamental insights into the molecular biology of how corals respond to heat stress, these experiments are limited by their temporal and spatial scales and the ability to reproduce complex natural environments. Corals living in naturally marginal environments are therefore valuable study subjects for understanding how corals will respond to future climates. Importantly, however, no single site can capture the various stresses expected under future climate conditions. It is therefore essential that corals living in different scenarios are studied to explore the commonalities and limitations of coral adaptation (Camp et al., 2018). In this thesis, I surveyed two coral reef systems along the similar latitude from the eastern and western coast of Australia.

1.3.1 The Great Barrier Reef (GBR)

Located on the northeastern coast of Australia, the GBR is the largest coral reef system in the world and is famous for its biodiversity and broad latitude distribution (more than 2,300km) and temperature gradients extending (Hopley et al., 2007; Pandolfi & Kelley, 2011). The individual reefs of the GBR are highly connected with most no more than 50km from an adjacent reef. Although there is a predominant southward flowing offshore current (East Australian Current), dispersal is complex as the strength and direction of inshore currents is highly weather dependent. Maintaining this connectivity is key to the resilience of the coral reef ecosystem and to its recovery as well (Hock et al., 2017). Corals on the GBR continuously experience casualties in hard corals from tropical cyclones, the coral-eating crown-of-thorns starfish, water pollution, and bleaching (Cheal et al., 2017; De'ath et al., 2012; MacNeil et al., 2019; Ortiz et al., 2018). The proportion of coral death from bleaching is increasing with more frequent heat waves because of anthropogenic climate changes (Ainsworth et al., 2016; Guest, 2021). Meanwhile, variation in bleaching susceptibility in corals was observed in this broad-range latitude distribution and temperature gradients resulting from different thermal histories (Ainsworth et al., 2016; Lundgren et al., 2013).

Factors both historical and contemporary are shaping the population structure of corals on the GBR (Lukoschek et al., 2016; Smith-Keune & van Oppen, 2006; van Oppen et al., 2011). The formation of the GBR is thought to have commenced approximately 600 thousand years ago, and its reef structures have reestablished themself repeatedly during sea-level change associated with glacial cycles (Brodie & Cohn, 2021). Some major divisions observed in the

current coral populations are assumed to be a result of allopatric divergence between glacial refugia during Pleistocene glaciations (van Oppen et al., 2011). Although cryptic species and population structure of *Acropora* spp. were found along the GBR (Ladner & Palumbi, 2012; Matias et al., 2022), the ongoing southward flowing East Australian Current is thought to be the main gene flow driver for spreading adaptive alleles from low latitude corals and maintaining genetic diversity in southern reefs (Dixon et al., 2015; Fuller, Mocellin, et al., 2020; Lukoschek et al., 2016; Matz et al., 2018).

1.3.2 The Macrotidal coral reefs in the Kimberley region in northwest Australia

The Kimberley region of Western Australia (WA) is characterised by an extreme macrotidal environment with a tidal range of up to 12 meters while supporting highly diverse coral reefs (Richards et al., 2015). Corals on the reef flat in the Kimberley experience extended aerial exposure during low tides and strong tidal currents (Purcell, 2002). Intertidal corals living in the nearshore Kimberley region experience a daily temperature fluctuation of up to 7 degrees and maximum temperatures of up to 37 degrees Celcius along with subaerial exposure (Purcell, 2002; Richards et al., 2015). Combined with highly fluctuating temperature, water turbidity and pH, and dissolved oxygen, this extremely variable environment is thought to have been selected for highly resilient intertidal corals, possibly contributing to relatively low mortality during recent heat waves (Dandan et al., 2015; Lindsay et al., 2015; Rosser & Veron, 2011). However, more research is needed on this shallow tropical reef system to dissect to what extent and how these corals are capable of adapting to additional heat stress (Schoepf et al., 2015, 2019).

The modern coral reefs in the Kimberley bioregion were re-established after the Last Glacial Maximum (LGM) when sea-level rise inundated the continental shelf (Solihuddin et al., 2015). Despite the more recent reestablishments in Kimberley corals, the origins and connectivities within these coral communities remain unclear. Unlike the highly connected reef system on the GBR, many studies found population structure in brooding and broadcast spawning coral populations, reflecting relatively low connectivity among reefs (Adam et al., 2022; Thomas et al., 2017; Underwood et al., 2006, 2020). Although corals from the Kimberley regions especially the inshore reefs are found to be highly adapted to the local environment, the fact that these reefs are highly dependent on local recruitment and self-seeding may make them more vulnerable to further climate change (Adam et al., 2022; Gilmour et al., 2013; Thomas et al., 2017).

1.4 Diversity in Symbiodiniaceae: lessons from unmapped reads

Phenotypic variation in corals is determined by the interactions between host and symbiont genotypes and their environment, the coral-associated endosymbiotic algae (family Symbiodiniaceae) play an important role in environmental tolerance at the individual level (Berkelmans & van Oppen, 2006; Matsuda et al., 2022; Mote et al., 2021; Oliver & Palumbi, 2011; Torres et al., 2021). Coral-symbiont associations within species have also been shown to vary with a range of environmental factors including depth, temperature and host population structure (Bongaerts et al., 2013; Osman et al., 2020; Yorifuji et al., 2017). For example, high proportions of the known heat-tolerant symbiotic algae, *Durusdinium* genus, are frequently found in corals in high-temperature environments (Oliver & Palumbi, 2011). Furthermore, symbiont genetic diversity increases the functional diversity of coral-dinoflagellate assemblages and may help coral reefs survive climate change (Stat et al., 2008; Baskett et al., 2009). Thus, it is important to investigate the symbiont communities in corals for comprehensive evaluation of adaptation potential.

As intracellular residents, the DNA of Symbiodiniaceae is inevitably co-extracted with that of the coral host. As a result, the genome sequencing data of coral adult tissue often contains the sequences of their endosymbionts. Although the standard analysis of whole-genome resequencing data typically discards reads that do not map to the host reference genome these discarded reads may not only include reads from contamination, unknown structural variants, and highly diverged regions, but also the reads from symbiont genomes (Gouin et al., 2014). In reference genome construction, stringent filtering is needed to exclude data from symbionts in both experiment and data processing and to ensure the high quality of the genome assembly (Cooke et al., 2020; Fuller et al., 2019; Shinzato et al., 2011). However, the symbiont data from coral sequencing provides us with an extra portal to understand the association between symbiont diversity and host diversity, especially for population genomic studies in which hundreds of coral individuals are being collected.

To evaluate the within-family (genus to species level) genetic diversity of symbionts, many coral population genomic studies incorporate metabarcoding sequencing of the noncoding region of the circular plastid (psbA) or the Internal transcribed spacer 2 (ITS2) region for each coral sample (Davies et al., 2020; Matias et al., 2022; Thomas et al., 2022). This incurs extra costs, but may also be biased or lack of resolution due to the focus on variation within a single marker. Some tried to align reads to available Symbiodiniaceae genomes and quantify the relative read proportion in each genus in each coral sample (Fuller, Mocellin, et al., 2020). Cooke *et al.* used a *k*-mer-based tool Kraken (Wood & Salzberg, 2014) to determine the dominant symbiodiniacean genera and tried to explore the genetic diversity of

Cladocopium using a mitochondrial haplotype network (Cooke et al., 2020). However, these approaches are still limited by the unevaluated difference in evolutionary landscapes of mitochondrial and nuclear genomes in Symbiodiniaceae and difficulties in assembling symbiont mitochondrial genomes.

The reference assemblies have increased for Symbioniaceae including divergent genera and closely related species (Dougan, Bellantuono, et al., 2022; González-Pech et al., 2021; LaJeunesse et al., 2018). Recent studies based on these genome sequences have established that the phylogenetic relationships within and between genera can be evaluated based on an alignment-free approach using shared *k*-mer profiles (Dougan, González-Pech, et al., 2022; González-Pech et al., 2021). This highlights the potential of using *k*-mer-based methods to quantify the diversity of symbionts using coral whole-genome sequencing data. Even in the absence of a reference genome this method can make use of low-coverage genome-wide data and capture the hidden information from duplicated regions to evaluate sample-to-sample relationships based on differences in symbiont community composition.

1.5 Overall aim and thesis structure

The goal of this thesis is to apply advanced genomics techniques to improve our understanding of coral population genetics and evolution. To this purpose, I analysed whole-genome sequencing data of two *Acropora* corals located in the GBR and Kimberley region in northwestern Australia. In these studies, I i) identified variants and evaluated the population structure and admixture with genome-wide SNPs; ii) made inferences about the demographic history of coral populations and iii) applied a variety of statistics to explore genome-wide patterns of diversity and scan for selective sweeps. The gene candidates within these selective sweeps provide the potential molecular mechanism for heat tolerant. Finally, I applied a *k*-mer-based approach to make use of unmapped coral genome sequencing data to investigate the genetic diversity of coral-hosted symbionts and developed a pipeline to generalise this approach. I address these aims through three independent data chapters in this thesis, as outlined below:

Chapter 2 entitled "**Genomic variation in *Acropora tenuis* from inshore and offshore locations in the central Great Barrier Reef is structured by local selection, chromosomal inversions and differences in algal symbionts**":

I analysed low-coverage whole-genome resequencing data of 228 *Acropora tenuis* corals sampled along the central GBR to evaluate the genetic differentiation and cryptic divergence with respect to inshore-offshore and latitudinal gradients. All analyses were implemented

under the framework of genotype likelihoods. Under strong gene flow, the hypothesis of "genomic island of speciation" was examined by investigating the relationship between absolute divergence and relative divergence in populations. Using genome-wide F_{ST} scans I identified putative selective sweeps in corals from different habitats. Interestingly, I observed fractions of the genome with characteristic patterns of heterozygosity and local population structure indicative of chromosomal inversions. Low recombination within these inversions means they are strong drivers of structure in the heterogeneous genetic differentiation landscape for *A. tenuis* on the central GBR.

Chapter 3 entitled "**Evolutionary Responses of a Reef-building Coral to Climate Change at the End of the Last Glacial Maximum**":

This chapter used whole-genome resequencing of 75 *Acropora digitifera* corals sampled from three locations in the Kimberley region in northwestern Australia. In contrast to **Chapter 2**, the deeper sequencing used in this chapter allowed me to accurately call genotypes and phase haplotypes genome-wide. I inferred demographic history using methods based on the sequentially Markovian coalescent (SMC) and fitting coalescent models to the site frequency spectrum (SFS). Genomic regions with extended linkage disequilibrium (LD) that are likely linked to thermal tolerance in inshore corals were identified.

Chapter 4 entitled "**A workflow and *k*-mer-based approach to dissect the symbiont diversity in coral whole-genome sequencing data**":

The results of this chapter are based on simulated genome sequencing data and raw sequencing data from both **Chapter 2** and **Chapter 3** with the main focus on coral-associated algae genetic diversity. In this chapter, I developed and used a pipeline to harness extra unmapped reads from coral whole-genome sequencing data, build a *k*-mer count database for each sample, summarise the pairwise distance based on shared *k*-mers, and evaluate the relationship between samples based on the diversity of Symbiodiniaceae spp.

In **Chapter 5**, I summarise the major findings of the thesis and their implications for coral conservation. In this section, I also consider the limitations and novelty of my analysis and identify further research directions in the field of coral population genomics.

CHAPTER 2

Genomic variation in *Acropora tenuis* from inshore and offshore locations in the central Great Barrier Reef is structured by the local selection, chromosomal inversions and differences in algal symbionts

---

2.1 Abstract

Widespread use of genomic sequencing has revealed prevalent cryptic species and genetically distinct population subunits in reef-building corals, however, the evolutionary mechanisms that structure genetic diversity in these and other marine taxa with pelagic larvae remain poorly understood. Here, we used whole-genome sequencing of 228 individual colonies to study the genetic diversity of *Acropora tenuis* at inshore and offshore sites along a 300 km stretch of the central Great Barrier Reef (GBR). The genome-wide SNPs revealed a strong divergence between Magnetic Island and the other reefs, however, no genetic structure was observed between reefs from distinct inshore and offshore habitats. Further investigations of differentiation between inshore and offshore revealed loci with elevated F_{ST} and reduced nucleotide diversity indicative of strong local selection in inshore versus offshore populations. Although there was little genetic structure throughout the majority of the genome, we identified four genomic regions with clear local structure and patterns of individual heterozygosity indicative of chromosomal inversions. Despite limited genetic differentiation between inshore and offshore sites for the coral host, genetic distances based on *k*-mer of their symbiont sequences were clustered separately . By combining genetic data from both corals and their symbionts, this study provides insights into how coral populations differentiate across environmental gradient. It also demonstrates that chromosomal inversions can be detected from shallow whole genome sequencing data, paving the way for further investigation of this important class of structural variant in coral evolution.

Keywords: coral differentiation, chromosomal inversion, Central GBR, population genomics, symbiont composition diversity

2.2 Introduction

The key goal of population genetics is to understand the spatial and temporal distribution of genetic variants within and between species or populations. This underpins fundamental questions in evolutionary biology such as the formation of new species with gene flow and adaptation to changing environments. Recently, population genomic studies across a wide range of taxa have observed non-random genetic differences across the genome (Harr, 2006; Malinsky et al., 2015; Nadeau et al., 2012; Tine et al., 2014; Turner et al., 2005) leading to the hypothesis that gene flow may vary across genomic loci. More specifically, genomic regions with barriers to gene flow referred to as “islands of differentiation”, inferred by a high relative (F_{ST}) and absolute (D_{XY}) sequence divergence (Cruickshank & Hahn, 2014) have been suggested to play important roles in pre- and post-zygotic isolation and local adaptation (Harr, 2006; Malinsky et al., 2015; Pennisi, 2014). Thus the compelling concept of genomic islands of differentiation provides a potential explanation for speciation in circumstances with no geographic barrier to stop hybridisation (Barton & Bengtsson, 1986; Turner et al., 2005).

The cause of heterogeneous patterns of genetic differentiation across the genome remains incompletely understood. In the context of speciation, genomic islands were initially thought to arise due to divergent selection which reinforces reproductive isolation in regions where specialised genes are located. However, there is an increased realisation that this signature could emerge with alternative processes (Cruickshank & Hahn, 2014) unrelated to population divergences, such as linked selection (Burri et al., 2015; Corbett-Detig et al., 2015), recombination rate variation (Burri et al., 2015; Singhal et al., 2015), chromosome rearrangements (Burri et al., 2015; Harringmeyer & Hoekstra, 2022) and background selection in regions with low recombination rates (Charlesworth, 2012; Cutter & Payseur, 2013; Hudson & Kaplan, 1995). Even without the premise of divergence with gene flow, selection can act on the genome to reduce local genetic diversity by hitchhiking, accelerate lineage sorting, and ultimately give rise to heterogeneous patterns of differentiation. Given the wide range of speciation systems in wild populations, individual investigations are needed for each system to understand the mechanism of heterogeneous genetic differentiation.

Coral reefs are one of the most diverse ecosystems and have been estimated to contain more than 25% of all marine species (Knowlton, 2001; Forsman, 2005). In addition, it is now becoming clear that many coral populations harbour cryptic diversity and are much more structured than would be expected given the high potential for gene flow in mass-spawning marine taxa (Bongaerts et al., 2021; Matias et al., 2022). One factor that might explain some

of this cryptic diversity is the reshaping of coastal marine environments by cycles of glaciation during the Pleistocene, leading to periods of isolation in distinct refugia and subsequent range expansions (Cooke et al., 2020; J. E. Fifer et al., 2022; J. Zhang et al., 2022). Another factor contributing to the differentiation between coral populations is the high rate of local recruitment of coral larvae in seascapes where underwater geographic barriers such as currents and distances between isolated reefs may cause an overall reduction in gene flow (Thomas et al., 2015; Underwood et al., 2020). Finally, adaptation to local environments also contributes to cryptic diversity because specialised habitats facilitate the settlement of larvae with different genetic variants (van Oppen et al., 2018; Rippe et al., 2021). While these studies illustrate a range of ecological and genetic mechanisms driving diversity in corals, the relative paucity of genome-wide population genetic studies makes it difficult to generalise the relative importance of these mechanisms or predict their role in adaptation to future climatic conditions.

Genomic tools allow us to not only obtain accurate estimates of genetic statistics but also to investigate genome-wide patterns of variation in these statistics. With such data it becomes possible to identify divergences within individual regions due to the high density of loci. For this reason, whole-genome sequencing (WGS), is more effective than reduced representation techniques when characterising the genomic landscape of divergence and identifying local peaks in genetic differentiation (Szarmach et al., 2021).

Here, we applied a shallow (~3-5X per individual) whole genome sequencing approach to characterise genome-wide patterns of genetic variation in the common reef-building coral *Acropora tenuis* from the central Great Barrier Reef (GBR). Corals were sampled at inshore and offshore locations along a small latitudinal gradient spanning approximately 300km. Gene flow of *Acropora* corals along the GBR is influenced by the southward flowing East Australian Current in offshore sites as well as inshore currents that predominantly flow northward but are highly variable. Under this regime, genetic data and biophysical modelling suggest an asymmetric but consistent gene flow from north to south leading to rich genetic diversity in the south (Matz et al., 2018; Riginos et al., 2019). Overlaid on this broad-scale pattern, substantial differentiation was observed in the inshore central GBR between *A. tenuis* from Magnetic Island and other sampling sites elsewhere on the GBR (Cooke et al., 2020; Matias et al., 2022). A combination of these two studies shows that the *A. tenuis* population on Magnetic Island is a geographically restricted group that is almost absent everywhere else on the GBR. Demographic modelling in prior work suggests that Magnetic Island populations of *A. tenuis* are deeply diverged from those at surrounding reefs with an estimated split at around ~600 ka ago with secondary contact (Cooke et al., 2020). In a

previous study (Cooke et al., 2020), selective sweeps were identified in the inshore reefs along an approximately 300km transect northward from Magnetic Island indicating a history of strong selection in this region, however, without a comparison to offshore reefs it was not possible to determine whether these signatures are specifically related to adaptation to an inshore environment. A large-scale survey of the same species along the full extent of the GBR revealed three main clusters also with estimated divergence times between 0.5 and 1.5Mya. These population clusters were not strictly geographically separated but did differ in their association with inshore vs offshore reefs (shore location) and latitudinal distribution (Matias et al., 2022). While these studies provide a clear map of broad-scale genetic differentiation in *A. tenuis* on the GBR the mechanisms that allowed these clusters to diverge, and that maintain divergence in the face of gene flow remain unclear. One way to tackle the first of these questions is to examine recent and ongoing processes of divergence along environmental gradients. Along the central GBR there are clear environmental differences in turbidity, salinity and temperature between inshore and offshore reefs that could drive local selection and genetic divergence (Abrego et al., 2009; Rocker et al., 2017; Smith-Keune & van Oppen, 2006; Warner et al., 2015). Studies across a range of coral species have also shown the genetic variation patterns of associated photosynthetic algae are crucial for their adaptation to local environmental stresses (Camp et al., 2020; Hoadley et al., 2021; Howells et al., 2009; Manzello et al., 2018). The adaptive variation has been seen in between Symbiodiniaceae species and populations and was determined by the disturbance regimes and oceanographic transport (Howells et al., 2013; van Oppen et al., 2001; Howells et al; 2012). Furthermore, the response of coral in the local environment could be a pattern of symbiont genetic diversity that reflects the different patterns of the genetic composition of symbionts observed from their coral host (Howells et al., 2009; Howells et al.,2013). A combined investigation of both host and symbiont genetic composition could provide a more comprehensive view of the adaptation of corals (Howells et al., 2016; Manzello et al., 2018).

In this study, we extended an existing (Cooke et al., 2020) low-coverage whole-genome sequencing (lcWGS) dataset of *A. tenuis* from inshore reefs with additional lcWGS data from four offshore locations along similar latitudes. Genome-wide single nucleotide polymorphisms (SNPs) that were called jointly across these 212 samples were used to investigate the genetic differentiation among reefs across latitudes and inshore-offshore gradient locations (Supplementary Table 2.1). Besides the substantial differentiation between Magnetic Island and other reefs, we investigated the genomic landscape of differentiation between inshore (except MI) and offshore reefs, searching for regions with extreme F_{ST}. Since chromosomal inversions have previously been shown to contribute to differentiation between populations (Harringmeyer & Hoekstra, 2022; Mérot et al., 2021) we

also searched for these within the (non-MI) *A. tenuis* population using a method based on a local principle component analysis (PCA) scan throughout the genome. Furthermore, we made use of the photosynthetic algae (Symbiodiniaceae) DNA sequence inclusion in the sequencing data of coral samples to infer the associated genetic diversity in coral symbionts and to exam whether different patterns exist between reefs.

2.3 Results

We combined data from five inshore sites in previous study (Cooke et al., 2020) and four extra offshore sites to investigate the genetic patterns in the central Great Barrier Reef (Figure 2.1).

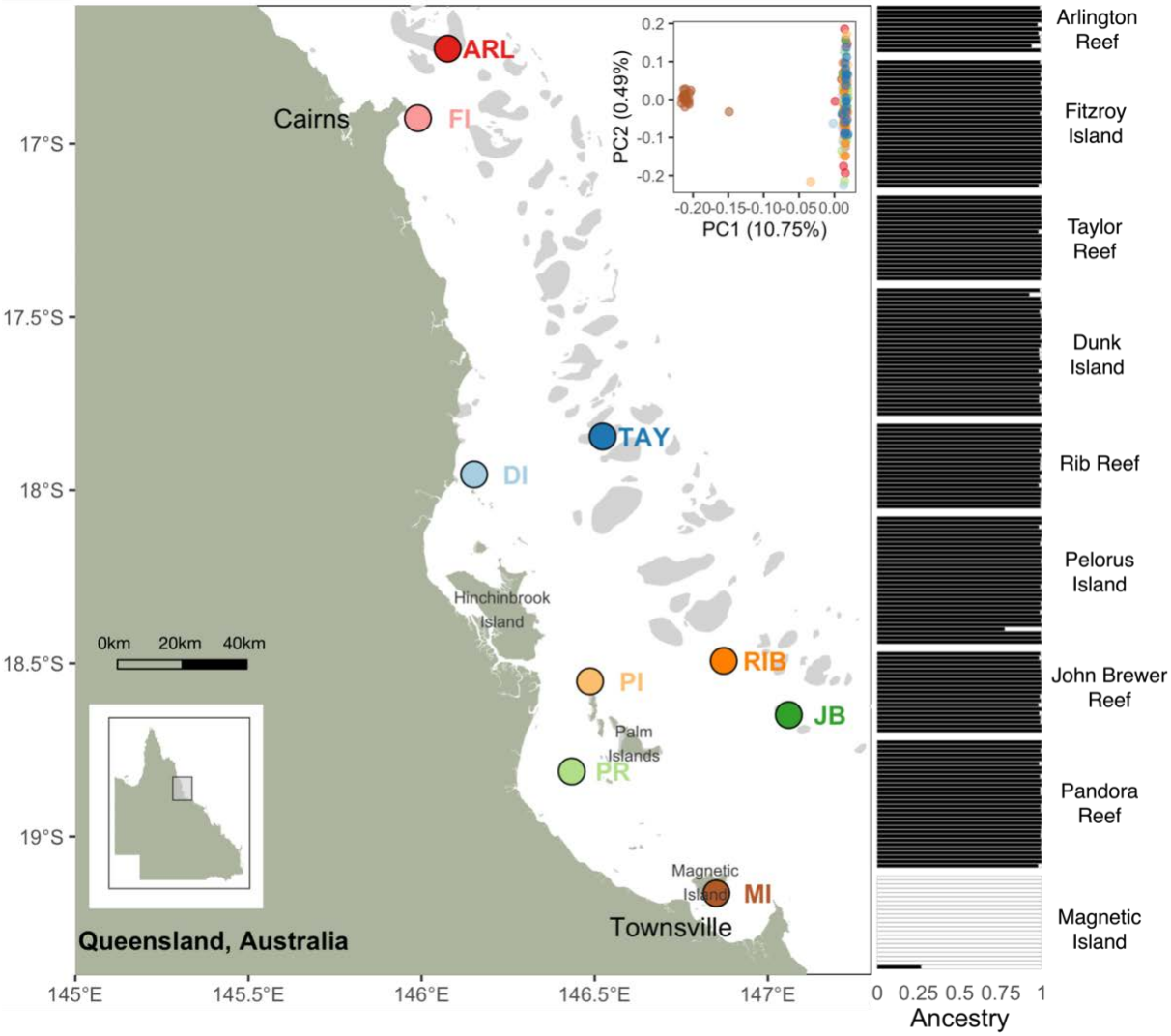


Figure 2.1 Sampling sites, population structure, and admixture of *Acropora tenuis* on the central Great Barrier Reef. Nine coloured dots on the map indicate the sampling location of this study with sample size for offshore reefs (Arlington Reef (ARL), Taylor Reef (TAY), Rib Reef (RIB), John Brewer Reef (JB)) are N=20, for north inshore reefs (Fitzroy Island (FI), Dunk Island (DI), Palm Islands (PI) are N=30, Pandora reef (PR)) and for Magnetic Island (MI) is N=28. The insect

PCA plot on top right corner shows the genetic relationship among all samples. The right panel displays admixture results with two ancestral groups (black and white).

2.3.1 Variant call set

After mapping and stringent quality controls (see Methods) over the samples and the reference genome assembly, 212 samples passed filtering and approximately 256 million accessible genomic bases were used in our analyses (Supplementary Table 2.1; Supplementary Table 2.2). 10 samples (9 from Arlington Reef, and 1 from John Brewer Reef) were discarded due to species misidentification (Supplementary Table 2.3; Supplementary Fig 2.2) while six samples from Magnetic Island were removed as clones or close relatives (Supplementary Fig 2.2). Variant calling based on GATK-like genotype likelihoods and the likelihood ratio test ($p<1e^{-6}$) implemented in ANGSD (Korneliussen et al., 2014) identified approximately 3.8 million (3,786,724) SNPs that were polymorphic within the genomes of all valid samples.

We observed similar sequence nucleotide diversity (π) ranging from 0.0060 to 0.0063 among reef locations (Supplementary Table 2.4). Negative Tajima's D values were observed in all reefs, however, we found that Tajima's D in offshore reefs was higher than inshore reefs and Magnetic Island also shows significant higher Tajima's D than north inshore reefs (p=2.2e-16; Wilcoxon rank sum test) (Supplementary Fig 2.9). Since the data was sequenced in two batches but with the same sequencing platform, we investigated the potential effect of batch sequencing using individual heterozygosity of each sample, and we noticed a large variance in individual heterozygosity estimates of inshore samples than offshore samples (Supplementary Fig 2.11). Thus, the difference in Tajima's D values between inshore and offshore sites could be either biological or explained by a difference between batches affecting the number of segregating sites. Since there was a slight difference in sequencing coverage between batches (higher in offshore) we also checked to see if this could explain differences in the variance on individual heterozygosity, however, we found that there was no significant correlation between sequencing coverage and individual heterozygosity in offshore reefs, and only a moderately correlation in inshore samples (R=0.48) (Supplementary Fig 2.20). Thus, despite stringent quality control measures is impossible to completely rule out the possibility of a batch effect that might influence genome-wide analyses such as genome-wide average F_{ST}. In our analyses that follow we largely focus on localised measures of differentiation that are robust to differences in sequencing data. In cases where our analyses are based on genome-wide averages, we explicitly remind readers that a batch effect cannot be ruled out as explaining the data.

2.3.2 Population structure in the GBR *A. tenuis*

We first explored the population structure of all *A. tenuis* samples by performing a principle component analysis with genome-wide SNPs using PCAngsd (Meisner & Albrechtsen, 2018). This revealed a clear separation between Magnetic Island and all other reefs with the most substantial divergence captured by the first eigenvector explaining 10.75% of the total variance (Figure 2.1; Supplementary Fig 2.5). In contrast, PC2 largely captured within-population diversity among all reefs other than Magnetic Island. Samples from Magnetic Island clustered tightly along PC2, whereas samples from other reefs were spread widely along PC2 without any pattern based on reef location. This lack of structure with respect to reef location was further supported by a plot of PC2 against PC3 (Supplementary Fig 2.5) and agrees with previous work that identified samples from Fitzroy Island and Orpheus Island as dominated by a single genetic cluster (cluster 1a in Matias et al., 2022). Admixture analysis with ngsAdmix (Skotte et al., 2013) supported a model with two ancestral source populations (K=2) with the split clearly partitioning samples into Magnetic Island and other reefs respectively (Figure 2.1; Supplementary Fig 2.7). Almost all samples had the complete assignment to one homogeneous ancestry lineage, with the exception of one admixed sample from Magnetic Island and three from other reefs (Figure 2.1). These four admixed samples are likely hybrids and we excluded them (Supplementary Table 2.1) in the following population-level analyses. Since Magnetic Island is the southern-most sample in our study we refer to other locations collectively as “north”, while further dividing this group into north inshore (Pandora Reef, Pelorus Island, Dunk Island, Fitzroy Island) and north offshore (John Brewer Reef, Ribbon Reef, Taylor Reef, Arlington Reef).

Estimates of genetic diversity and pairwise population differentiation were carried out based on reef locations. Similar to PCA analysis, we observed a consistently markedly higher fixation index (F_{ST}) between Magnetic Island and all other reefs (0.215-0.226) than between reefs within other locations (0.0091-0.021). This strong divergence between Magnetic Island and the other reefs is in agreement with the previous study of inshore GBR *A. tenuis* (Cooke et al., 2020). Here we show that reefs from MI are also distinct from other offshore reefs. There is no detectable relationship between geographic distance and F_{ST} within the north GBR samples (except MI), however, the F_{ST} values were slightly higher between Arlington Reefs and other reefs (Supplementary Fig 2.10). Clustering based on the matrix of pairwise F_{ST} further suggests a monophyly grouping of the north inshore reefs (Supplementary Fig 2.10). While PCA-based evidence supports our inference that corals from each reef location (except MI) are recruited from a mixed larvae pool, a recent study sampled the full expanse of the GBR and discovered cryptic divergence based on the latitude or inshore-offshore

location (Matias et al., 2022). Furthermore, our effective migration modelling analysis (EEMS) (Petkova et al., 2016) based on paired-wise identify-by-state (IBD) reveals a lower relative migration rate between inshore and offshore than within inshore and offshore samples (Supplementary Fig 2.12). Although we carried out stringent controls to avoid batch effects, we cannot completely rule out the possibility that this result and the monophyly of inshore samples in F_{ST} analyses are artifacts caused by the fact that our inshore and offshore samples were sequenced in separate batches. Overall, our analyses of population structure revealed that samples from Magnetic Island constitute a distinct genetic cluster, meanwhile, all other reefs from northward of Magnetic Island compose one largely homogeneous genetic group irrespective of their inshore-offshore and latitude locations.

2.3.3 Genomic islands of differentiation caused by selection

Lack of population structure in the north GBR samples suggests strong gene flow, however, strong differences in selection regimes between environments could still lead to differentiation at specific loci (Thomas et al., 2022). To identify regions of the genomic landscape with strong sequence divergence between the north inshore reefs and offshore reefs despite strong ongoing gene flow, we examined the genome-wide profiles of F_{ST} between north inshore samples and offshore samples. With F_{ST} values estimated in sliding windows with a size of 20kb and a step of 4kb (which we found to have the best resolution; Supplementary Fig 2.17 and Supplementary Fig 2.18), the results suggest an overall low level of sequence differentiation with only a small fraction of genomic windows showing high F_{ST} deviating from the genome average (Figure 2.2a). Such heterogeneous genome-wide sequence differentiation could arise due to barriers to gene flow, for example, due to chromosomal inversions which restrict recombination, leading to so-called “islands of differentation” between closely related populations. Alternatively, they could arise due to selective sweeps that bring up the frequency of the favoured allele, while reducing within-population nucleotide diversity at neighbouring loci (Malinsky et al., 2015; Poelstra et al., 2014). With the ecological set-up of high gene flow across reefs, we tested one hypothesis that the highly diverged regions are associated with restricted gene flow between populations by assessing the absolute sequence divergence (D_{XY}) in the highly differentiated regions (within windows with top 1% F_{ST} values). If these islands are indeed associated with restricted gene flow they should be associated with elevated absolute divergence (D_{XY}),

however, we found that D_{XY} was slightly lower in the high F_{ST} regions (Figure 2.2c).

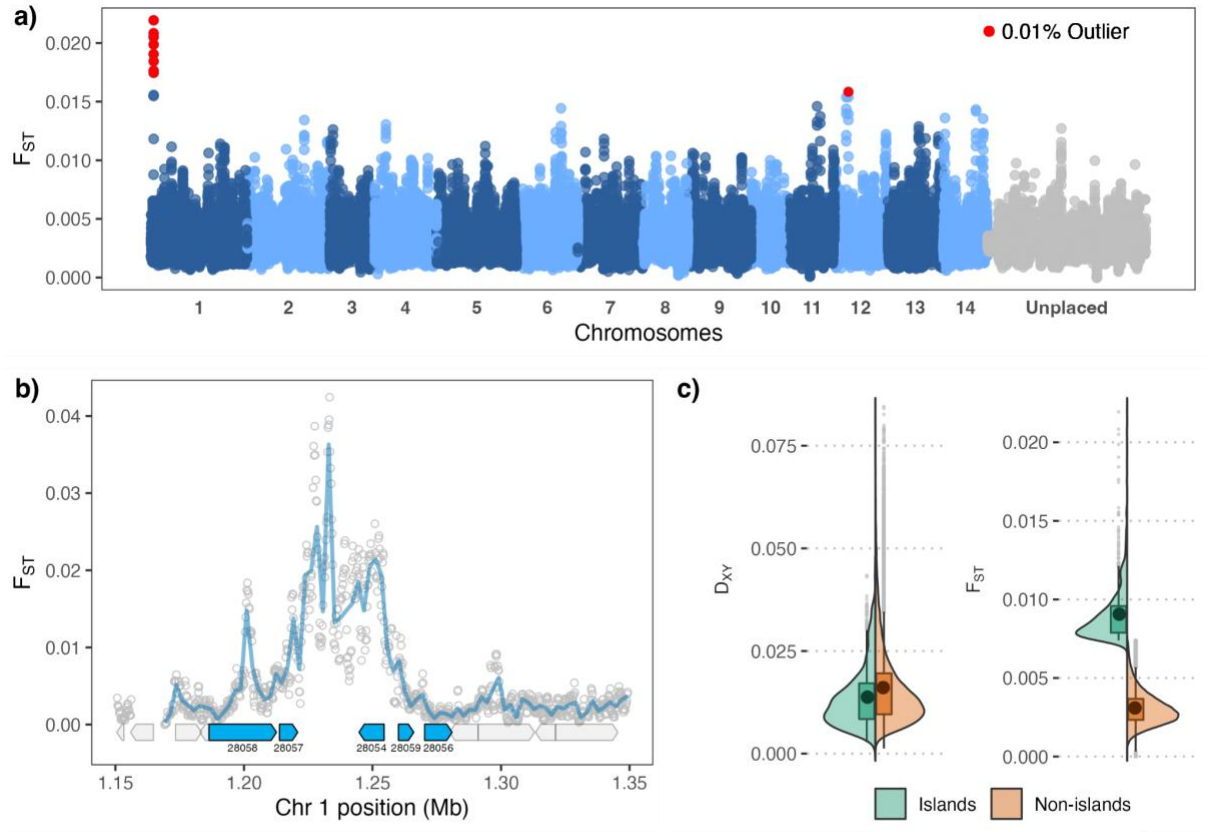


Figure 2.2 Genome-wide genetic differentiation pattern between the north inshore reefs and offshore reefs. a) Manhattan plot of genome-wide F_{ST} values for overlapped 20kb windows with 4kb step size, red dots represent the top 0.01% outliers. b) Close-up view of the estimated F_{ST} calculated in sliding windows of 1kb across the region of chromosome 1 surrounding the outliers. The coding regions of predicted genes were indicated below and genes within the high F_{ST} region are coloured in blue with the arrows pointing transcriptional direction. The unique numeric part of the gene IDs from the general feature format (GFF) file was used to label each gene for short (see Supplementary Table 2.5 for full details of all genes). c) Boxplot and split violin plots for measures of relative divergence F_{ST} and absolute divergence D_{XY} in the high F_{ST} regions (top 1%) and low F_{ST} regions.

The lower D_{XY} and reduced within-population genetic diversity (π) in islands led us to the alternative hypothesis that these islands are from linked selection (Figure 2.2 c; Supplementary Fig 2.14). Intriguingly, of the genome-wide distribution of F_{ST}, significant clusters of strong population differentiation were observed in the first chromosome of our pseudo-chromosome-level assembly across the whole genome with a windowed F_{ST} value of 15.1 SD higher than the average F_{ST} estimate and with the top 0.01% of observed F_{ST} values (Figure 2.2a). Furthermore, SNP loci within this region revealed different structuring patterns in the north inshore reefs and offshore reefs and the minor allele frequency (MAF) distributions suggest the inshore reefs abound with loci of high allele frequencies (Supplementary Fig 2.13).

This one localised region, which is less than 100kb, is overlapped with five predicted genes. Two genes in tandem sequence (28058, 28057) are homologs of HEAT Repeat Containing 1 (HEATR1) gene 28054 is homologous to ATP-dependent Clp protease proteolytic subunit (CLPP). Gene 28059 is not a homolog of known protein, however, it contains a polysaccharide deacetylase domain (PF01522.20), and gene 28056 is predicted as the G-protein-signaling modulator 2 (GPSM2). The CLPP gene was found to be the hub gene of differential expression which positively correlated to the combined treatment of antibiotics and heat (Connelly et al., 2022). GPSM2 modulates the activation of G-protein and was found to be highly expressed after short-term heat stress with sediment in *Pocillopora* (Poquita-Du et al., 2019). Within this region, we did not observe an aligned decrease in within-population genetic diversity (π) which suggests this differentiation is still in progress (Supplementary Fig 2.15). Meanwhile, the potential function of genes within this region suggests their roles in heat response.

2.3.4 PCA-based scan identified regions with a pattern of chromosome inversion

Since chromosomal inversions suppress recombination between heterokarytypes they provide a mechanism by which islands of differentiation can arise between locally adapted populations. Intuitively this occurs when multiple locally adapted alleles are captured on an inverted haplotype resulting in a “super-allele” that has a selective advantage over non-inverted haplotypes whose fitness can be reduced via recombination with migrant alleles (Kirkpatrick & Barton, 2006).

To identify genetic variation patterns consistent with inversion polymorphisms in the north (all reefs except Magnetic Island) population, we first performed a scan based on local principal component analysis (PCA) for each locus (Galinsky et al., 2016; Meisner et al., 2021). This scan (performed using PCAngsd on GL data) revealed four genomic regions ranging in size from 200Kb to 2Mb with exceptionally strong population structure compared with the genomic background in which no structure was present (Figure 2.3; Supplementary Fig 2.16; Supplementary Table 2.6). We did PCA analysis with genotype data of each locus in Figure 2.3 (top panel), visual inspection of the population structure within each region revealed three major clusters along PC1 as expected based on the three possible genotypes of an inversion polymorphism (Harringmeyer & Hoekstra, 2022; Huang et al., 2020). While all regions displayed this tripartite clustering, the boundaries between clusters were much

more distinct in some regions (L2-L4) than others (L1) (Supplementary Fig 2.16).

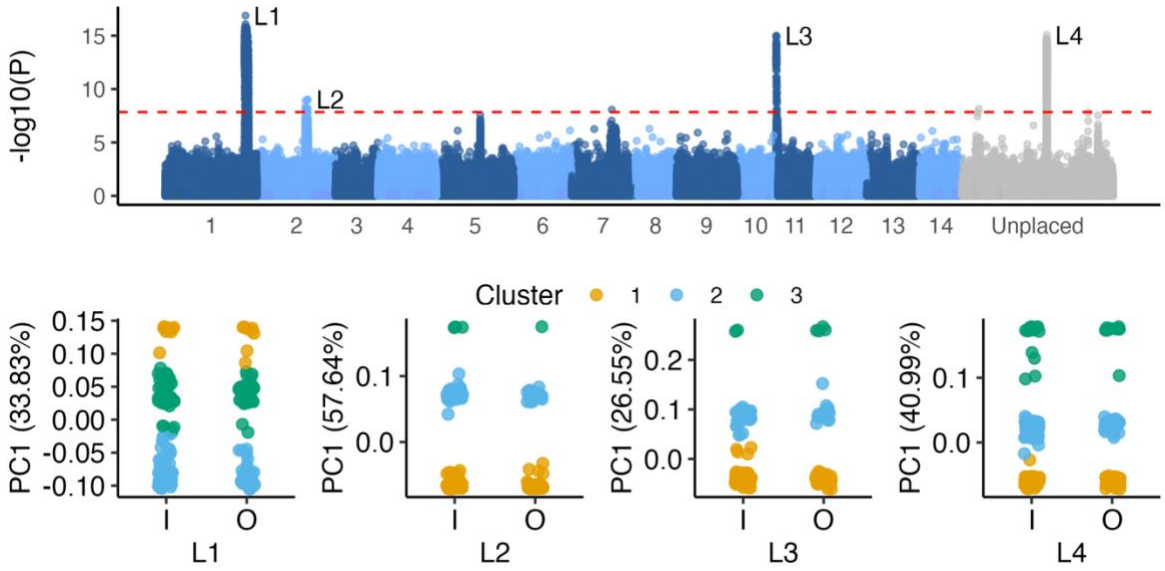


Figure 2.3 Genome-wide scores of selections along the first principle component in PCAngsd selection scan. The top four signals are denoted as L1, L2, L3, and L4 with sample loadings on PC1 across these loci accordingly. Reefs from inshore and offshore environments are distinctly represented on the x-axis (I: inshore; O: offshore). Samples were partitioned into three clusters based on their PC1 loadings using the K-means algorithm.

To further confirm that the clusters in PCA correspond to genotypes of an inversion we checked the individual heterozygosity in the three loci with a clear three-cluster pattern on their PCA plots (L2: Sc0000185, L3:Sc0000135, and L4:Sc0000214) by binning samples along the PC1. The individual heterozygosities of samples in the middle cluster are significantly higher than in the first and the last clusters (Figure 2.4). Therefore, these three clusters represent the three types of haplotypes (the homokaryotypes of reference arrangement, the heterokaryotypes and the homokaryotypes of the alternative arrangement) from within the inversion (Harringmeyer & Hoekstra, 2022; Mérot, 2020). A further check for Hardy–Weinberg equilibrium of the lead SNPs at loci revealed they are not significantly deviated from HWE (p<1e-6) (Supplementary Table 2.7) and the same result was obtained by genotyping samples based on cluster membership. Since inversions effectively block recombination in hetero karyotypes they allow selection to act more effectively if multiple beneficial alleles are captured within the inverted region (Betancourt et al., 2009). We investigated genes within inversions and found that some inversions, especially the largest ones can capture many genes whereas others have few (Supplementary Table 2.6). Even the largest inversions found here are relatively small compared with those found in other species (Harringmeyer & Hoekstra, 2022; Mérot et al., 2021). Since smaller inversions have a lower chance of capturing multiple selected alleles, it is possible that the inversions

identified here are largely neutral as no geographical pattern was found to structure allele frequencies for these inversions.

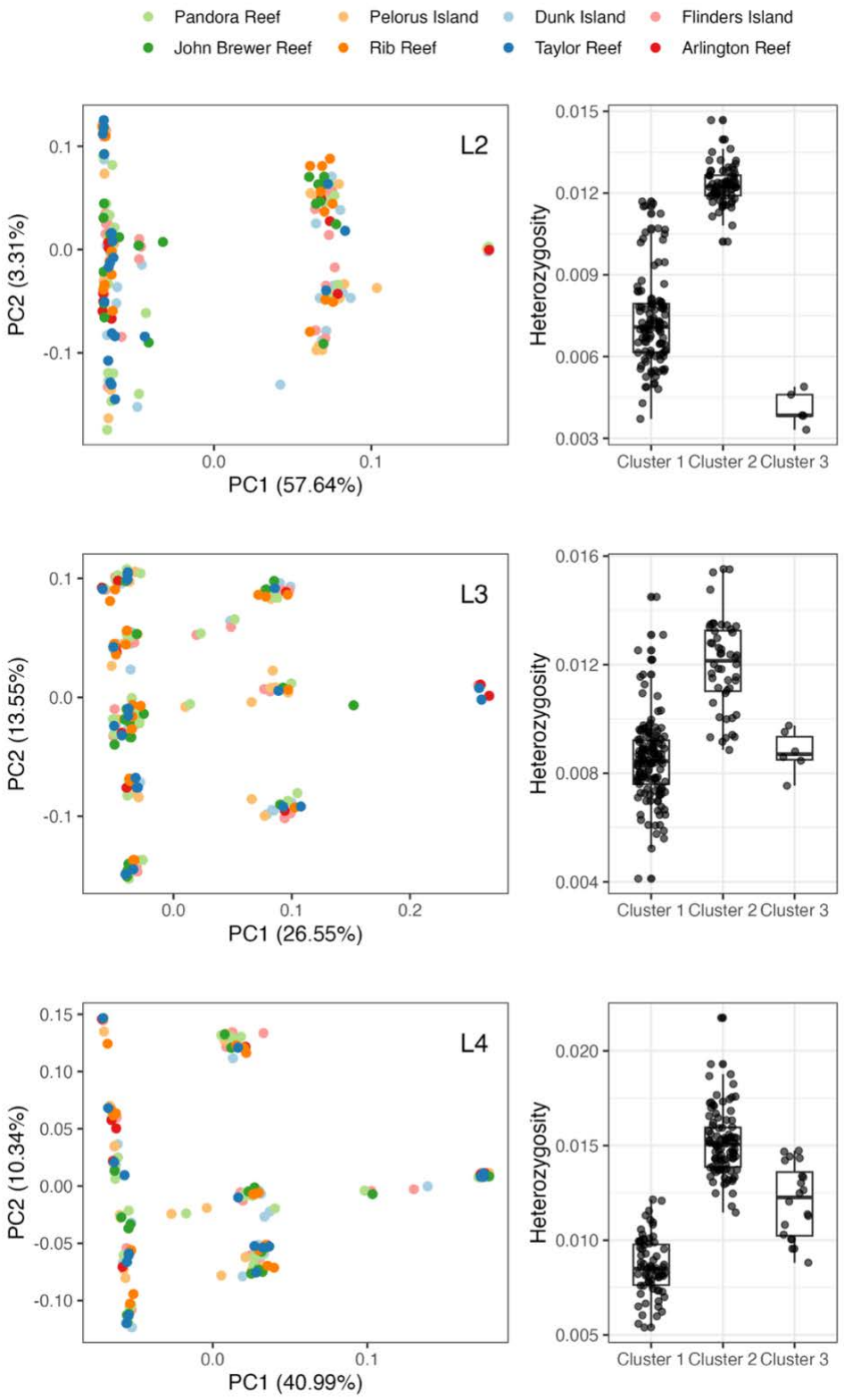


Figure 2.4 PCA plots depicting North GBR samples and boxplots of individual heterozygosity within each cluster of three loci (L2, L3, L4) as presented in Figure 2.3. Samples are colour-coded by

reef location, showing the first two components inferred with PCAngsd. The right-side plots display the individual heterozygosity levels of samples within each cluster as defined in the Figure 2.3 of corresponding locus.

2.3.5 Signatures of selection in the corals from Magnetic Island

The demographic history of both the north GBR and Magnetic Island *A. tenuis* has been thoroughly studied by Cooke et al 2020. Our 2D SFS plot also suggested the long divergence time between Magnetic Island and other reefs (Supplementary Fig 2.19). Under this secondary contact scenario, we also tested the hypothesis of genomic islands of speciation between Magnetic Island and the northern inshore reefs. We didn't observe elevated D_{XY} in genomic islands whereas the genetic diversity in the genomic island is decreased especially for the north inshore reefs (Figure 2.5a). We further found that those high F_{ST} regions are mainly characterised by the lower genetic diversity in the north inshore reefs (Figure 2.5b). It suggests that the highly differentiated regions are mainly shaped by linked selection which reduces genetic diversity while increasing the beneficial allele frequency in the population. Although the genome-wide high differentiation between Magnetic Island and the north inshore reefs makes it difficult to set a threshold for F_{ST} outliers, interestingly, the region with the maximum F_{ST} (>0.9) is composed of the epidermal growth factor (EGF) domain–containing gene cluster which was one of the top selective sweeps in the north inshore reefs identified by Cooke et al 2020. This again highlights the important roles of positive selection in shaping the coral genomic landscape of differentiation.

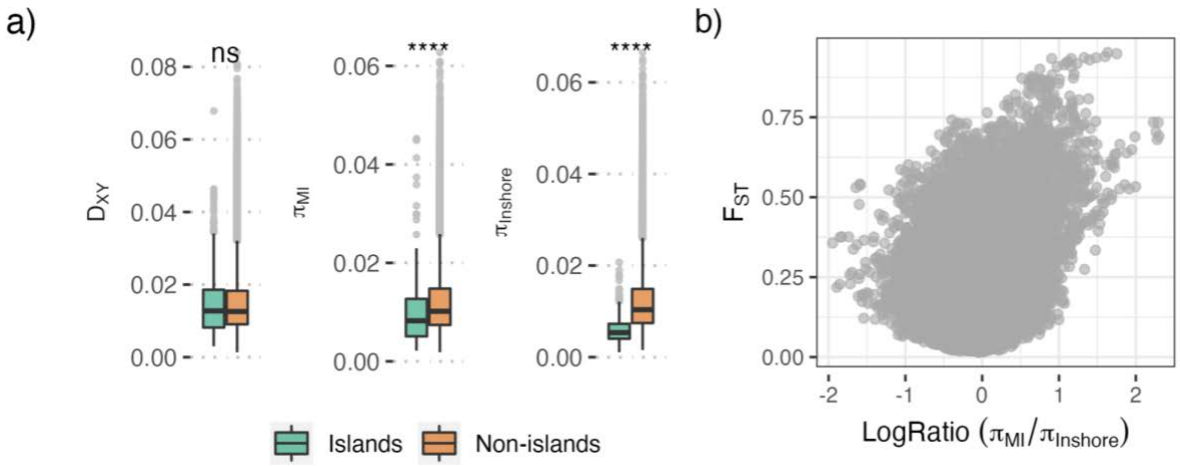


Figure 2.5 a) The distribution of D_{XY}, and genetic diversity in high F_{ST} regions. b) the relationship of log ratio($\pi_{MI}/\pi_{Inshore}$) versus the F_{ST}. A skew to the right at high F_{ST} values suggests that selective sweep mainly towards inshore reefs.

2.3.6 Environmental differentiation in symbiont communities

To quantify the genetic diversity of symbiont within and among each population, we implemented a *k*-mer based method to calculate the pairwise genetic distance among samples (See detailed method in Chapter 4). The symbiont composition between inshore and offshore forms two separated clusters based on pairwise D_2^S distance calculated from shared *k*-mer profiles (Figure 2.6). Samples within one cluster suggest a closer genetic relationship and may compose of a population. Since the distance is based on sequences from *Cladocopium*, the variation between clusters represents the with-in species variation. Within inshore corals, we observed the clustering of samples consistent with mitochondrial haplotype analyses in previous work (Cooke et al., 2020). The symbionts from MI form a tight cluster, while symbionts from Pandora Reef, Fitzroy Island, and Pelorus Island display much greater variation. Most samples within Dunk Island are clustered with MI samples except three individuals (Figure 2.6). As discussed by (Cooke et al., 2020), this pattern within inshore reefs may reflect the differentiation of symbiont communities between marine sites and plume sites as a result of different water quality regimes. The consistency between results based on D_2^S and previous mitochondrial haplotype analyses provides some validation for the D_2^S method which I use again in Chapter 3, and which I describe in detail in Chapter 4. The distinct pattern between inshore and offshore reefs potentially reflects the genetic diversity in symbionts associated with inshore-offshore environments, however, it is difficult to exclude the effect from the batched dataset.

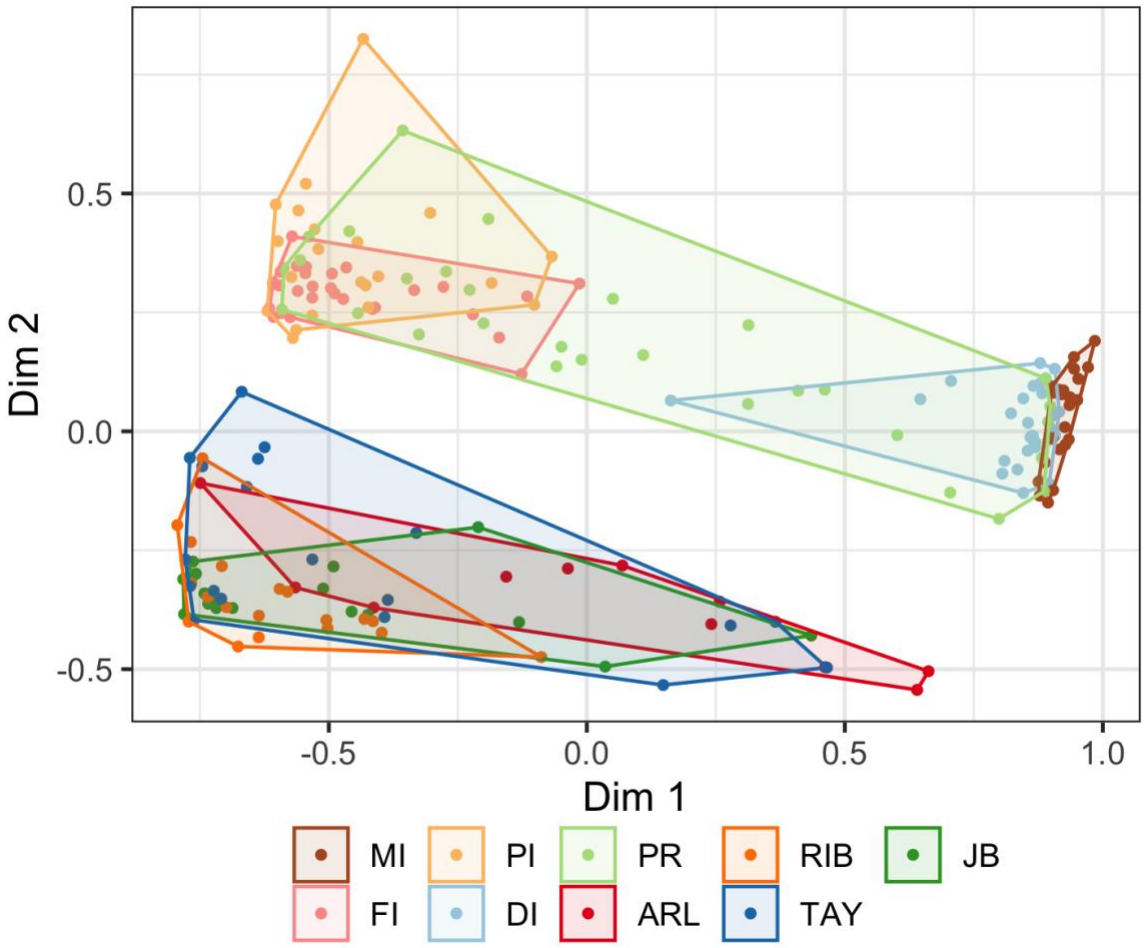


Figure 2.6 The MDS plot based on the D_2^S distances calculated using shared *k*-mer profiles. Points represent each sample coloured by location.

2.4 Materials and Methods

2.4.1 Experimental design, DNA extraction and sequencing

We combined the dataset of shallow whole-genome sequencing data of *Acropora tenuis* colonies at five locations from inshore Great Barrier Reef (GBR) (n=148; Ira Cooke et al.2020) and 4 locations from offshore GBR (n=80) along the approximately aligned latitudes (Figure 2.1). These samples were sequenced at a similar time with the same sequencing strategies (100bp paired-end) using the Illumina HiSeq platform. In total, we obtained 2-5x depth of coverage whole genome sequencing data for 228 samples (with the exception of two high-coverage samples). There was a slight difference in sequencing coverage between inshore and offshore reefs. Among inshore samples, they were sequenced at about 2-3X whereas offshore samples were sequenced at 4-5X.

Despite the same sequencing strategies were used for inshore (MI, FI, DI, PR, PI) and offshore (ARL, TAY, JB, RIB) samples, we were mindful of the potential for batch effects that might result in artifactual patterns distinguishing inshore and offshore samples. Such effects (if present) are expected to be most strongly observable as genome-wide population structure and systematic differences in individual heterozygosity (Lou & Therkildsen, 2022). Although no obvious population structure was present in principle component analyses (Supplementary Fig 2.5; Supplementary Fig 2.6) a neighbour-joining tree based on pairwise differences inferred using identity by the state did separate inshore from offshore samples (Supplementary Fig 2.8). A similarly subtle difference was observed in estimates of individual heterozygosity which were similar in magnitude across locations but much more variable in inshore reefs than offshore (Supplementary Fig 2.11). Taken together these results suggest that a small sequencing batch effect cannot be ruled out between inshore and offshore samples. Importantly, all of the key analyses presented here concern genomic variation that occurs across both inshore and offshore sites making it unlikely to be influenced by batch effects.

2.4.2 Data pre-processing and mapping

We followed the gatk germline variant calling best-practices workflow to generate the mapped bam files from raw reads for each sample. Reads passing quality checks from each sample and lane was converted to unmapped bam format (uBAM) files. Adapters were marked using MarkIlluminaAdapters (Picard) before mapping to the reference genome assembly using bwa (v0.7.17-r1188). Next, duplicate reads from polymerase chain reaction (PCR) during library preparation or artifacts from sequencer optical sensors were marked using MarkDuplicates (Picard). The mapping depth across the reference genome and mapping rate of each individual were then summarised using Bedtools (v2.30.0) (Quinlan & Hall, 2010) Coveragebed. Since two of our samples (FI-1-3, MI-1-4) were sequenced at much higher coverage (28X, 26X), we used sambamba (v0.8.2) (Tarasov et al., 2015) to downsample reads for these samples at about 3X to ensure all samples had approximately even coverage (Supplementary Fig 2.1).

2.4.3 Removal of clones and misidentified samples

To ensure no cryptic species or misidentified samples were in our data, we first reconstructed a mitochondrial genome sequence for each sample by aligning raw reads to the mitogenome sequence of *Acropora tenuis* (Genbank accession AF338425) and then extracting the most common base at each position using the -doFasta 2 option in ANGSD. These sequences of our samples were then used as queries to search the NCBI non-

redundant nucleotide sequence database (nt June 23 2022) using megablast (v0.8.2) (Morgulis et al., 2008) with the option to output a maximum of 5 best matches with taxonomy information (-outfmt '6 qseqid sseqid pident mismatch gapopen evalue staxids sscinames scomnames sskingdoms stitle'). The top hits for the most samples were the mitogenome of *A. tenuis* whereas nine samples from Arlington Reef and one sample from John Brewer Reef matched A. echinata the best. Next, we built a phylogenetic tree using IQtree (v1.6.4) (Nguyen et al., 2015) based on the alignments (mafft v7.394) (Katoh & Standley, 2013) of mitogenome sequences of all samples together with one *A.tenuis* mitochondrial sequence and the mitochondrial sequence *Acropora echinata* (LC201841.1). The resulting tree was then visualised using the R package ggtree (Supplementary Fig 2.2). This tree revealed that nine samples from Arlington Reef and one from John Brewer Reef formed a distinct monophyletic clade together with the A. echinata mitogenome. Since these same samples also had particularly low mapping rates and genome coverage (Supplementary Fig 2.1) it is highly likely that they were misidentified in the field. We, therefore, excluded these samples from all further analyses.

Clones and closely related samples could also bias the population-level inferences and thus need to be carefully excluded. To estimate pairwise relatedness, we used ngsRelate v2 (Korneliussen & Moltke, 2015) (https://github.com/ANGSD/NgsRelate) which calculates relatedness statistics based on genotype likelihoods. As input to ngsRelate, we used all SNPs passing quality filters (see below) and genotype likelihoods calculated using ANGSD (v0.928). Next, we followed the methodology described by (Waples et al., 2019) to identify closely related samples and clones based on three statistics of each pair of samples: R0, R1, and KING-robust kinship (all calculated by ngsRelate). This analysis revealed eight pairs of closely related (expected kinship of 0.125) samples all of which were from Magnetic Island) as shown in Supplementary Fig 2.3. For each pair of related samples, we kept the one with a higher sequencing depth and since some samples had multiple close kin this resulted in the removal of seven samples from Magnetic Island leaving a total of 21 from that location and 212 in total for further analysis.

2.4.4 Quality filtering on reference: mappability, simple short repeats, and sequencing depths

It is suggested to use all sites in population genetic statistics estimation in shallow whole genome sequencing because of the uncertainty of genotype (Lou et al., 2021). To differentiate these sites from missing sites, we filtered low-quality regions of the genome before running ANGSD. We applied several filters to remove regions prone to ambiguous mapping and errors. Firstly, we used GENMAP (v1.2.0) (Pockrandt et al., 2020) to estimate

the mappability of each site in the reference genome. The mappability scores were computed with 50bp *k*-mers, with a maximum of 2 mismatches (-K 50 -E 2). We only kept sites with a mappability score equal to one which suggests they can be uniquely mapped. We also used mdust (v2006.10.17; default parameters) to identify and exclude low-complexity regions in the genome. Since regions with very high or very low mapping depth are often associated with ambiguous mapping due to repeats, we removed sites with global depth across all samples with a depth higher or lower than the range's 1% percentile (minimum 17x, maximum 1102x) (Supplementary Fig 2.4). Finally, we excluded any sites from small scaffolds with a length of less than 1Mb to reduce the influence of artifacts at the ends of fragmented reference sequences. Applying all these filters left 258,421,687 accessible bases from the *A. tenuis* reference genome to be used for all further analyses unless otherwise noted.

2.4.5 Genotype likelihoods and single nucleotide polymorphisms (SNPs)

To account for the uncertainty of genotypes of each site due to low (2-5x) per-sample sequencing coverage, we used ANGSD to estimate genotype likelihoods. ANGSD was run using the genotype likelihood (GL) model from GATK (-gl 2), inferring major and minor alleles from GL data (-doMajorMinor 1), estimating allele frequencies from GL data (-doMaf 1) and was restricted to filtered sites we described above via the -sites option. In addition, we also limited our analysis to sites with a base quality score of at least 30 (-minQ 30), a read mapping quality score of at least 30 (-minMapQ) and no missing data in at least 100 individuals (-minInd 100). SNPs were further filtered to remove rare alleles (MAF >0.05) and keep only sites with p value<10-6 (-SNP_pval 1e-6) in the likelihood ratio test implemented in ANGSD.

We also generated an unlinked SNPs dataset and use them in PCA analysis and Admixture analysis to avoid the potential bias in analysis caused by strong linkage disequilibrium (LD). The pairwise LD scores of variation sites were calculated using ngsLD (v1.1.1) (Fox et al., 2019) which takes account of the uncertainty of genotype by using genotype likelihoods. LD pruning was then applied using the Perl script prune_graph.pl within ngsLD with a maximum distance of 5kb and a maximum weight of 0.5. After generating the list of unlinked sites, we computed the genotype likelihoods as a beagle format file (-doglf 2) of these loci again using ANGSD with the same parameters as before.

2.4.6 PCA analyses and admixture

With the genotype likelihoods from all samples in the beagle format we generated, we did a principal component analysis (PCA) using pcangsd (v1.10). The output covariance matrix estimated based on individual allele frequency was then used to compute eigenvalues and eigenvectors using the R package eigen and to generate PCA plots. In the same run computing the covariance matrix, pcangsd was also used to set the best number of clusters with a soft upper search bound of 10000 (-admix -admix_auto 10000) based on PC loadings. We also used NGSadmix (v33) to estimate individual admixture proportions based on genotype likelihoods of all SNPs across the genome and unlinked SNPs with both K=2 and K=3. The admixture results for both K were visually inspected to decide the best K and the run with the lowest log-likelihood values was presented. The PCA and Admixture analysis performed based on unlinked SNPs revealed qualitatively the same results with genome-wide SNPs, thus, the result from NGSadmix with all SNPs was presented in the results section. Additionally, four samples with high ancestral proportions were labelled as hybrids and were excluded in the following population-level analysis (Supplementary Table 2.1).

2.4.7 IBS and Neighbor-joining (NJ) tree

The pairwise genetic distances measured as identity-by-state were estimated by ANGSD (-doIBS 1) which randomly samples a single read from each position from each sample within filtered reference sites. A matrix was generated with the IBS distance of all samples for constructing a neighbour-joining (NJ) tree and visualising using the R package ape (Paradis & Schliep, 2018) and ggtree (Yu et al., 2018). Samples that were identified as containing a high level of admixture ancestry were excluded from the NJ tree (Supplementary Table 2.1).

2.4.8 Estimation of Effective Migration Surfaces (EEMS)

The IBS matrix was used as input for EEMS to visualise the relative migration rates among sampling locations. We manually extracted the centroid of geographical coordinates of each sampling location from google maps as the coordinate input needed by EEMS. Similarly, the outline of the habitat was manually drawn in Google Maps around our sampling locations. EEMS was run using the runeems_snps program and default settings for 5 million steps and a burn-in of 1 million steps, with 400 demes and the number of sites (n = 8,713,550) based on the IBS matrix. The results were visualized using rEEMSplots (https://github.com/dipetkov/eems).

2.4.9 Genome-wide estimates of genetic diversity within and between reefs

To calculate reef-specific diversity and divergence statistics, we first used the realSFS program within ANGSD to estimate one-dimensional (1D) folded site frequency spectra (SFS) for each of the nine reef locations separately, and two-dimensional folded SFS (2D) for each reef pair. Before estimating the SFS with realSFS, a 2-step procedure was first implemented to generate a saf (site allele frequency likelihood) file followed by an optimisation of the saf file using ANGSD (-dosaf 1). We also generated an SFS file with the north reefs as a whole (except Magnetic Island) considering the strong population structure between the north reefs and Magnetic Island. Pairwise nucleotide diversity (π), Watterson's θ, and Tajima's D were estimated from the 1D-SFS of each reef using the thetaStat function within ANGSD with a sliding-window size of 10kb and s step size of 2kb. Global estimates of F_{ST} for each pair of reefs were computed directly from the 2D-SFS using the Reich estimator implemented in realSFS (Reich et al., 2009). A hierarchical clustering based on pairwise F_{ST} was also generated using the R package ape.

2.4.10 Individual heterozygosity

The heterozygosity for each sample was estimated in ANGSD as the proportion of heterozygous sites in the 1D-SFS of each individual. A saf file was generated for each sample using ANGSD and used to estimate the 1D-SFS with the realSFS. The heterozygosity rate is calculated by dividing the number of variant sites by the total number of sites in R.

2.4.11 Calculating sliding-window population genetic statistics

We used ANGSD to estimate the genome-wide patterns of pairwise F_{ST}, genetic diversity, and Tajima's D. F_{ST} was calculated for both between north inshore and offshore reefs, and between north inshore reefs and Magnetic Island whereas genetic diversity and Tajima's D were calculated for north inshore reefs, offshore reefs, and Magnetic Island, separately.

Samples identified as hybrids before were excluded from this analysis. A saf file was generated for each grouped reef before using realSFS to estimate 1D SFS and 2D SFS. To ensure the same loci were used in all reefs, we first run ANGSD with all bam files first with filterings (-minQ 30 -minMapQ 30 -C 50 -uniqueOnly 1 -minInd 100 -sites) and extract the position. Next, ANGSD was used again to generate the saf files for each reef with no further filtering except site restriction (-sites). We then used the realSFS program (v0.928) within ANGSD to estimate the site frequency spectrum (SFS) and calculate the pairwise F_{ST} in sliding windows (realSFS fst stats2 -type 1) between north inshore and offshore reefs, and

between north inshore and Magnetic Island with the default settings, based on the saf files. The theta statistics were also estimated using thetaStat within ANGSD (thetaStat do_stat -type 1) with the same sliding windows for the north inshore, offshore reefs, and Magnetic Island, separately. We applied the sliding-window scan with three window sizes: 50kb, 20kb, and 10kb with a jump size of 10kb, 4kb, and 2kb, respectively. Eventually, we used a window size of 20kb with a step size of 4kb each which displays a good balance between reducing stochastic variation by averaging over the region and fine genomic resolution (Supplementary Fig 2.17). As the windowed theta statistic outputs are the sum of the per-site estimate, we divided the estimated values by the number of loci (both variant and invariant) in that window to obtain accurate estimates of diversity (Korunes & Samuk, 2021).To avoid false signals resulting from window-based statistics dominated by very little data in the window, we excluded windows with the number of pass filtering bases that constituted less than 10% of all sites (Supplementary Fig 2.18).

We used a Perl script getDxy.pl (https://github.com/mfumagalli/ngsPopGen/blob/9ee3a6d5e733c1e248e81bfc21514b0527da967b/scripts/getDxy.pl) provided by the ngsPopGen toolset to calculate the D_{XY} for every site in the mafs files generated by ANGSD, non-bi-allelic sites were removed in the calculation. Per-site D_{XY} values were then grouped into sliding windows from F_{ST} estimates and the average value was assigned as the value for each window using Bedtools intersect and groupby. Note that only the global distribution of D_{XY} was used in our results instead of actual values which were claimed to be overestimated.

To characterise the within and between population genetic diversity (π, D_{XY}) in islands and non-islands, we first defined regions with F_{ST} values greater than 99 quantile threshold as high F_{ST} regions and then compared the D_{XY} and π distribution in each category.

2.4.12 PC-based scan and patterns of heterozygosity

We performed PC-based genome-wide scans using PCAngsd (-selection) (Meisner et al., 2021) for outliers from neutral population structures within the north GBR samples. The PCAngsd statistics along the scaffolds were plotted as a Manhattan plot with log scaled p-value. The significance level of multiple testing correction was determined using Bonferroni (α=0.05). For each outlier locus, we inferred the population structure with PCAngsd and then binned the individual along the first principle component (PC1) into three clusters and calculated the heterozygosity of each sample within clusters using ANDSD to investigate the patterns.

2.4.13 Building a pseudo-chromosome reference

To facilitate the visualisation of genome-wide genetic statistics in Manhattan plots, we used ragtag v2.0.1 (Alonge et al., 2019) to align the *Acropora tenuis* genome to the *Acropora millepora* chromosome-level genome assembly (Fuller, Mocellin, et al., 2020) with default parameters. 488 out of 614 *A. tenuis* scaffolds were placed accordingly, comprising 94.3% of the assembly. The results were used to translate the base position in the original *A. tenuis* assembly into the pseudo-chromosome level assembly for visualisation purposes.

2.5 Discussion

Our results based on whole-genome sequencing data of *Acropora tenuis* provide evidence that inshore and offshore reefs on the central GBR form a single genetic cluster and confirmed the substantial separation between Magnetic Island and other reefs. Despite the high connectivity between inshore and offshore reefs we found localised genomic regions with extremely high F_{ST} but demonstrated that these are likely to be the result of linked selection and not a barrier to gene flow. Our local PCA-based scan across the genome of all inshore and offshore samples revealed population structure patterns that are indicative of chromosomal inversions. The analysis of associated symbiont communities demonstrated clustering based on local environments related to inshore and offshore locations and plume/marine sites.

A recent study has revealed the cryptic divergence within *Acropora tenuis* populations along the entire length of the GBR and found three sympatric genetic clusters (Matias et al., 2022). These clusters were shown to be separated back to approximately one million years ago with currently ongoing gene flow. Similarly, prior work using shallow-whole genome sequencing of the *A. tenuis* data also revealed the deep divergence between Magnetic Island and four other sites on the Central GBR (Cooke et al., 2020). This study also focused on the central GBR and extended the sampling sites to include four offshore reefs. Consistent with previous results, the distinct divergence between Magnetic Island and the other reefs was observed which aligns with the historical separation. The large divergence between Magnetic Island and the north GBR suggests that these might constitute separate species, however, the existence of hybrids indicates that ongoing gene flow between these genetic clusters is likely. Meanwhile, no structure between inshore and offshore reefs within the north group reflects high connectivity and a shared genetic origin among the 8 sampled reefs. The subtle differentiation among reefs across latitudes and at inshore-offshore locations (except MI) is not surprising as *Acropora* corals which are broadcast spawning corals often show great potential for long-distance dispersal and reefs from the GBR have

displayed a high level of connectivity in studies (Fuller, Mocellin, et al., 2020; Lukoschek et al., 2016; Matz et al., 2018). Although analysis with EEMS and pairwise F_{ST} suggests that gene flow between inshore and offshore may be reduced, this effect was small and we could not rule out the possibility that it is caused by batch effects in our data.

Within the central GBR, a clear environmental difference exists between offshore sites where water quality is high and inshore where rivers, tides and waves generate higher turbidity, salinity, temperature and nutrient loads (Browne et al., 2012). Despite the strong environmental differences, we did not observe strong differentiation between shore locations at the whole-genome level but a window-based scan of F_{ST} revealed regional spikes with elevated differentiation. Since these top differentiated regions by F_{ST} did not also have a consistent elevated absolute differentiation (D_{XY}) they are more likely to have arisen under directional selection rather than a genomic barrier (i.e. the hypothesis of genomic islands of differentiation). Although the identification of the causal gene is challenging due to linkage disequilibrium, two of the genes (GPSM2, CLPP) within these highly differentiated regions are associated with coral heat stress response and might therefore contribute to survival in inshore environments where temperatures are more variable (Poquita-Du et al., 2019; Connelly et al., 2022).

The role of structural variations has been underappreciated in genetic studies of non-model organisms as they remain challenging to detect. So far there have been no studies on the role of inversion polymorphisms in coral populations, yet this class of structural variant has been shown to play a major role in the spread of adaptive variation in other species (Harringmeyer & Hoekstra, 2022). Unlike insertion, deletion and other copy number variations, inversions that alter the orientation of chromosome segments almost completely suppress the local recombination rates in heterokaryotypes (Farré et al., 2013; Fuller, Koury, et al., 2020). Evidence has shown direct links between the generation and maintenance of the inversions and complex ecology and evolutionary processes (Kapun et al., 2016; Kirubakaran et al., 2016; Kunte et al., 2014; Küpper et al., 2016; Wellenreuther & Bernatchez, 2018). Recently, there have been an increasing number of studies emphasising the role of inversions for natural populations at heterogeneous habitats in local adaptation (Huang et al., 2020; Mérot et al., 2021) and differentiation (Harringmeyer & Hoekstra, 2022; X. Wang et al., 2022) whereas the role of inversions in corals remains unknown.

We attempted to detect inversions by uncovering the genetic structure caused by a subset of linked SNPs using local PCA (H. Li & Ralph, 2018). The principal component analysis (PCA) is designed for discovering and visualising the systematic patterns of genomic data sets. In this study, we applied a PCA-based scan to investigate the heterogeneity in patterns of

sample relatedness at intermediate genomic scales. The outlying loci which strongly affect the population structure form several clusters along the genome. Evidence from PCA visualisation and individual heterozygosity distribution suggests it is explained by chromosomal inversions. All of the inversions discovered by this process were common in the northern population, in Hardy Weinberg Equilibrium and showed no clear differences in frequency among reef locations. Although this lack of clear ecological association for inversions suggests that they may be selectively neutral it is not yet clear whether their maintenance in the population is driven by some selective mechanism or simply reflects the large Ne of coral populations allowing such variants to persist. Future work investigating inversions and testing the association with phenotypic variation using a larger sample size with a larger geographical distribution might answer these questions.

Direct identification of structural variation in genomes is challenging with short-read data, especially since inversion break-points tend to occur within repetitive regions (Corbett-Detig et al., 2019). While the shallow short-read data used here prevented us from precisely identifying the boundaries of inversions we were nevertheless able to identify them based purely on population genetic signatures. Future work using long-reads or deeper sequencing coverage will facilitate the validation of inversions and may reveal smaller or less frequently occurring inversions.

The zooxanthella distribution in *Acropora* species on the central GBR has been previously studied along with the host species and revealed composition variation within one species (van Oppen et al., 2001). Whether the patterns of coral-algae association are primarily shaped by genetics or environmental factors in different species remains unclear. Here, we found reefs from Magnetic Island show genetic distinction in both host and their algal symbionts compared with other inshore reefs. It is noteworthy that the succession of Symbiodiniaceae in *A. tenuis* juveniles can be dynamic for extended periods, this provides them the ability to establish with locally adapted Symbiodiniaceae species (Abrego et al., 2009). We also found symbiont genetic variation within a single reef which may contribute to the physiological diversity observed in corals (Beltrán et al., 2021).

In summary, we have confirmed the presence of previously known cryptic diversity between *A. tenuis* from Magnetic Island and other north reefs while also demonstrating a lack of genetic structure between inshore and offshore locations. Meanwhile, we uncovered the effects of selection and inversions in structuring the population at the intermediate genomic scale. Our discovery highlights the importance of detecting inversions specifically, and potentially other structural variations in future coral population genomics studies (Mérot, 2020; Mérot et al., 2020; Pokrovac & Pezer, 2022). Combined with diversity in symbiont

composition, our results address the importance of gene flow acting together with symbiont genetic composition to coral population differentiation.

CHAPTER 3

Evolutionary Responses of a Reef-building Coral to Climate Change at the End of the Last Glacial Maximum

This chapter is published in:

Jia Zhang, Zoe T Richards, Arne A S Adam, Cheong Xin Chan, Chuya Shinzato, James Gilmour, Luke Thomas, Jan M Strugnell, David J Miller, Ira Cooke (2022). Evolutionary Responses of a Reef-building Coral to Climate Change at the End of the Last Glacial Maximum, Molecular Biology and Evolution, Volume 39, Issue 10

3.1 Abstract

Climate change threatens the survival of coral reefs on a global scale, primarily through mass bleaching and mortality as a result of marine heat waves. While these short-term effects are clear, predicting the fate of coral reefs over the coming century is a major challenge. One way to understand the longer-term effects of rapid climate change is to examine the response of coral populations to past climate shifts. Coastal and shallow-water marine ecosystems such as coral reefs have been reshaped many times by sea-level changes during the Pleistocene, yet, few studies have directly linked this with its consequences on population demographics, dispersal, and adaptation. Here we use powerful analytical techniques, afforded by haplotype phased whole-genomes, to establish such links for the reef-building coral, *Acropora digitifera*. We show that three genetically distinct populations are present in northwestern Australia and that their rapid divergence since the last glacial maximum (LGM) can be explained by a combination of founder effects and restricted gene flow. Signatures of selective sweeps, too strong to be explained by demographic history, are present in all three populations and overlap with genes that show different patterns of functional enrichment between inshore and offshore habitats. In contrast to rapid divergence in the host, we find that photosymbiont communities are largely undifferentiated between corals from all three locations, spanning almost 1000 km, indicating that selection on host genes and not the acquisition of novel symbionts, has been the primary driver of adaptation for this species in northwestern Australia.

Keywords: *Acropora digitifera*, founder effects, glacial cycles, adaptive evolution, population genomics, selective sweeps

3.2 Introduction

Glacial cycling during the Pleistocene is thought to be a major driver of biodiversity dynamics (Hewitt 2000; Hofreiter and Stewart 2009) and its effects provide important lessons that can be used to help predict the impacts of future climate change (Hofreiter and Stewart 2009; Nogués-Bravo et al. 2018). Population genetics is a valuable tool to understand these past climate events because it can reveal historical changes in species' demography, connectivity, and diversity. Widespread application of population genetic tools to terrestrial (Hofreiter and Stewart 2009) and marine (Mattingsdal et al. 2019) species in the northern hemisphere has revealed a predominant picture of persistence in southern refugia followed by expansion and northward migration after the last glacial maximum (LGM), with more recent work describing differential species' responses depending on habitat requirements (Hofreiter and Stewart 2009) and patterns of dispersal (Mattingsdal et al. 2019). Much less is known about the impacts of past climate shifts on tropical marine systems such as coral reefs, despite the profound impacts that changes in temperature and sea level would have had on these shallow-water marine habitats (Wilson 2013; Ludt and Rocha 2015; Webster et al. 2018).

Throughout the tropics, the dominant effect of low sea levels during the last glacial maximum was a dramatic reduction in the amount of shallow water habitat (Kleypas 1997; Ludt et al. 2015). In broad agreement with this, many studies across a range of coral reef taxa have observed signatures of recent population expansion (Crandall et al. 2008; Crandall et al. 2012; Delrieu-Trottin et al. 2017), however not all populations follow this pattern. Genome-wide approaches are now revealing differential demographic histories of cryptic and recently diverged populations (Bierne et al. 2003; Cooke et al. 2020; Underwood et al. 2020; Bongaerts et al. 2021), some of which show signatures of recent isolation and decline (Moran et al. 2019). Moreover, the ranges of diverged populations in the marine environment are sometimes difficult to reconcile with modern geography and the potential for physical dispersal (Bierne et al. 2003; Cooke et al. 2020; Underwood et al. 2020; Bongaerts et al. 2021), and they may be better understood with reference to historical connectivity such as during past glacial maxima. A historical perspective may therefore be crucial to understanding gene flow and adaptation in extant populations. However, the value of this approach depends heavily on the temporal resolution of demographic analyses so that their timing can be linked to specific climate events, and with the ability to detect and characterise signatures of selection so that these can be used to assess modes of local adaptation.

Emerging techniques based on the sequentially Markovian coalescent (SMC) can be used to reconstruct demographic histories of species in unprecedented detail, potentially revealing

links with past climate (Nadachowska-Brzyska et al. 2015; Kozma et al. 2016; Chattopadhyay et al. 2019; Lucena-Perez et al. 2020). However, the most widely used variant of this technique, PSMC (Li and Durbin 2011), has limited power to infer recent events, a problem exacerbated by the large effective population size (Schiffels and Durbin 2014). Since corals and many other broadcast-spawning marine taxa have large effective population sizes, most studies so far have focussed on changes in the distant past that cover many glacial cycles (Prada et al. 2016; Mao et al. 2018; Fuller et al. 2020; Thomas et al. 2022). Inferences within the timeframe of the most recent glacial cycle require more sophisticated methods such as MSMC (Schiffels and Durbin 2014) and SMC++ (Terhorst et al. 2016) that make use of larger datasets (multiple whole genomes) to improve the sampling of haplotypes that share a recent common ancestor.

Even in systems where the effects of past climate change on biodiversity are relatively well understood, the role of natural selection and adaptation in response to climate change remains uncertain (Nogués-Bravo et al. 2018). Addressing this gap for climate-sensitive taxa such as corals is a pressing issue (Torda et al. 2017) directly relevant to their conservation and management in the Anthropocene. Adaptive evolution in corals is complex because it is likely to involve selection on the coral hosts themselves, as well as selections on and/or exchange of their dinoflagellate photosymbionts. Symbiont exchange is of particular interest because it may enable corals to adapt rapidly to anthropogenic climate change (Berkelmans and van Oppen 2006; Torda et al. 2017). Numerous studies have observed variation in host-symbiont associations along environmental gradients (Bongaerts et al. 2013; Camp et al. 2020; Ros et al. 2021), and experiments have demonstrated that a switch in symbiont partnership can be induced by stress (Matsuda et al. 2022). Another potential mode of climate adaptation in corals is the selection of the coral host. A range of studies examining population genetic, and gene expression differences between heat-adapted and naive corals all suggest that adaptation to heat is likely to involve many loci (Palumbi et al. 2014; Dixon et al. 2015; Fuller et al. 2020; Thomas et al. 2022). Modelling efforts have also attempted to describe the envelope of population genetic parameters, and the rate of climate change under which corals could adapt based on natural selection (Matz et al. 2018). So far, however, there are few studies (see Smith et al. 2022) that identify signatures of selection in relation to adaptation and survival over a sustained period of warming, such as the transition from the LGM to today.

In this study, we used a population whole-genome sequencing approach to understand the impacts of past climate change on the widespread reef-building coral, *A. digitifera* in northwestern Australia. In this region, *A. digitifera* is common on offshore atolls at the shelf-

edge and also forms part of a diverse inshore community (in the Kimberley region) that thrives despite extreme heat, frequent aerial exposure and highly variable turbidity (Richards et al. 2015; Richards et al. 2019). Modern coral reefs in the Kimberley were extirpated during the last glacial maximum (LGM), while those offshore may have persisted but would have experienced a period of much reduced shallow-water habitat and been much closer to the coast (Wilson 2013; Solihuddin, O'Leary, et al. 2016; McCaffrey et al. 2020). The contrasting biogeography of these sites provides an ideal case study of the effects of climate change during the last glacial cycle, and our analytical approach is designed to investigate this comprehensively. We do so through demographic modelling based on multiple whole genomes providing accurate inferences in the window leading up to and following the LGM (1kya - 100kya), and through sensitive detection of signatures of recent selection via extended haplotype homozygosity and population branch statistics. In addition, we use non-host reads to profile the dinoflagellate symbionts inhabiting each coral colony based on standard markers such as the ITS2 region of ribosomal RNA as well as via mitochondrial sequences and a novel *k*-mer-based distance metric. This combination of approaches allows us to examine the interplay between demographic change, connectivity, selection and shifts in symbiont community composition during a rapid climate change event for the first time in a coral.

3.3 Results

Whole-genome sequencing of 75 *Acropora digitifera* colonies from three reef systems in northwestern Australia yielded a mean per-sample coverage of 19.5X that we used to call approximately 9.6 million high-quality biallelic single nucleotide polymorphisms (SNPs) with GATK (Supplementary Fig 3.1; Supplementary Table 3.1, Supplementary Table 3.2). Of the few coral whole-genome studies conducted to date, most (Shinzato et al. 2015; Cooke et al. 2020; Thomas et al. 2022) adopted a shallow sequencing approach (except see Fuller et al. 2020). The relatively high sequencing depth in our study allowed us to reliably call genotypes at more than 95% of sites in 90% of samples (Supplementary Fig 3.1) supporting population-based haplotype phasing with SHAPEIT (Delaneau et al. 2012). As SHAPEIT infers missing genotypes based on phasing information, we tested its accuracy by removing genotypes with high-quality calls and then comparing their original value with that imputed by SHAPEIT. This confirmed that imputation (and by extension phasing) was generally highly accurate, relatively unaffected by minor allele frequency, but slightly better for sites with fewer missing values and for homozygous genotypes (Supplementary Fig 3.2).

3.3.1 Population structure in the coral host

PCA, and fineSTRUCTURE analysis (Figure 3.1C, Supplementary Fig 3.3), showed a clear genetic structure that divided corals from the six sampled reefs into three geographically separated groups, hereafter called North Offshore (NO) which includes Ashmore Reef, South Offshore (SO) which includes all reefs from the Rowley Shoals, and Inshore (IN) which includes two locations within macrotidal coral communities in the Kimberley (Adele Island, Beagle Reef). Using fineSTRUCTURE we also identified the substructure within the inshore population between samples from Adele Island (AI) and Beagle Reef (BR) (Supplementary Fig 3.3), however, the very tight clustering of all inshore samples in PCA analyses (PCs 1-3) indicated that this comprised a relatively minor component of genetic variation, and we, therefore, focussed on the three major clusters for our remaining analyses. Pairwise relatedness estimates based on shared genomic regions that were identical by descent (IBD) clearly partitioned samples into the three major clusters but failed to identify a distinction between BR and AI locations (Supplementary Fig 3.4).

The relative distance between PCA clusters, a tree inferred by fineSTRUCTURE (Supplementary Fig 3.3), another tree based on allele counts at established phylogenetic markers (Supplementary Fig 3.5) and relative amounts of IBD segments indicated a closer relationship between the two offshore populations than between offshore and inshore. Consistent with this, genome-wide estimates of F_{ST} were markedly lower (F_{ST} ~ 0.007) between offshore populations than between north-offshore and inshore (F_{ST} ~ 0.02) and south-offshore and inshore (F_{ST} ~ 0.02) (Supplementary Table 3.3). Despite low overall divergence (as measured with genome-wide F_{ST}) between these populations, admixture coefficients (calculated using ADMIXTURE; Alexander et al. 2009) showed complete assignment (>99%) of each individual to its parent cluster (Figure 3.1B), suggesting that migration is rare or non-existent between locations. Demographic modelling with fastsimcoal2 (see below) confirmed this as it supported a model with recent gene flow but with very low migration coefficients (probability of migration/individual/generation ~1e-4; Supplementary Table 3.9, Supplementary Table 3.10). Analysis of simulated data under this model with ADMIXTURE produced the same complete assignment to locations as observed for the real data (Supplementary Fig 3.16).

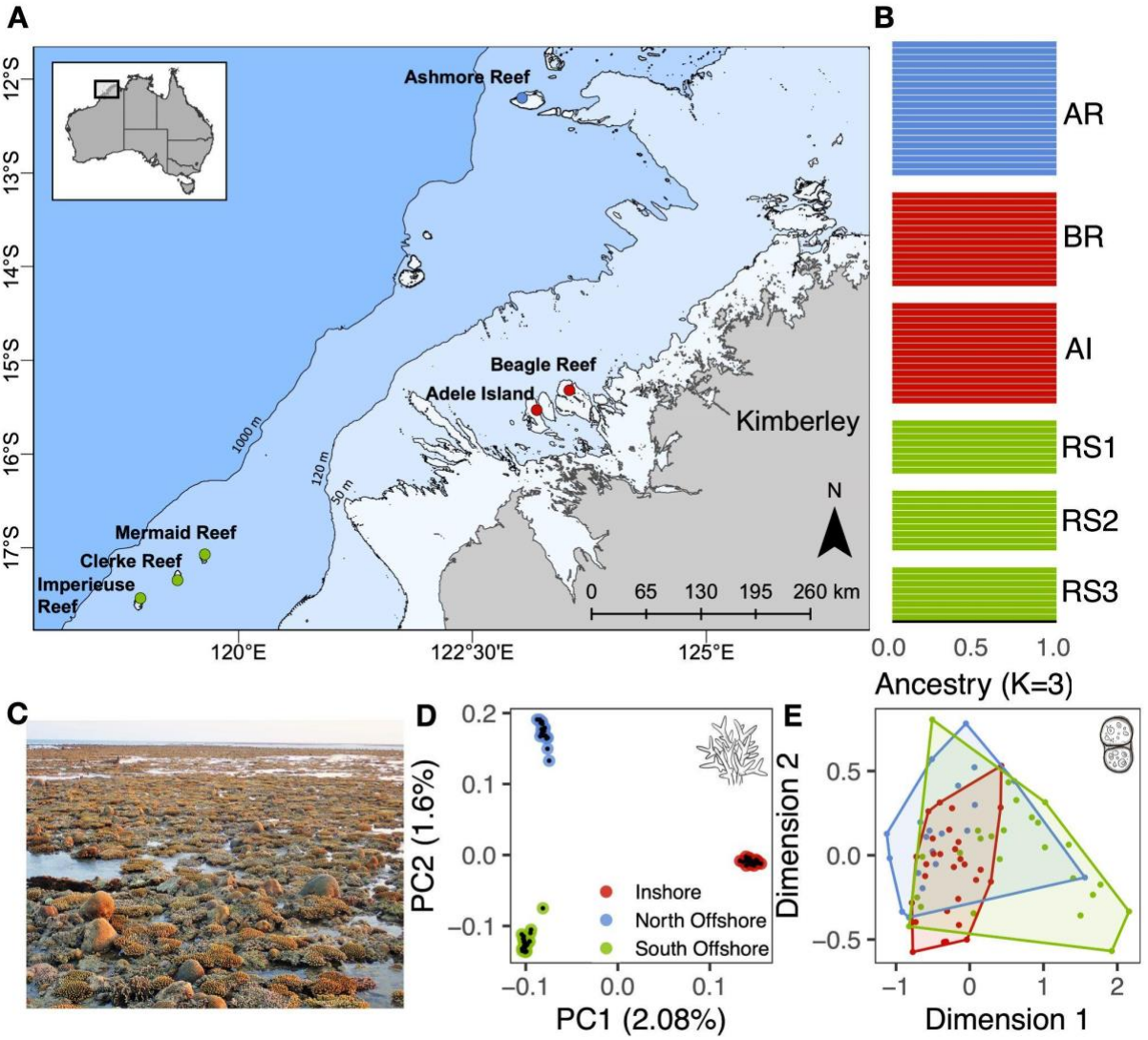


Figure 3.1 Sampling locations and genetic structure for the coral host and symbionts. All plots use the same colour scheme for locations as follows; North offshore, Ashmore Reef (AR) is shown in blue, inshore locations, Adele Island (AI) and Beagle Reef (BR) are shown in red, south offshore locations, Rowley Shoals (RS1: Mermaid Reef, RS2: Clerke Reef, RS3: Imperieuse Reef) are shown in green. A. Sampling locations in the Kimberley region, northwestern Australia. Bathymetric contours are shown at 50, 120, and 1000m depth with the present-day landmass shown in grey. B. Admixture proportions for each colony were calculated using ADMIXTURE with K=3 and coloured by the dominant cluster in each location. Each horizontal bar represents a single coral colony. C. Photograph of the reef flat at Adele Island showing corals exposed at low tide. Subaerial exposure for up to three hours during spring low tide is a characteristic feature of the inshore locations, AI and BR in this study. D. PCA showing the first and second principal components of genetic variation in the coral host. Points represent individual samples and are coloured by location. E. Multidimensional scaling plot showing relative pairwise distances between samples based on shared *k*-mers ($\boldsymbol{D}_2^S$ distance) from reads mapping to the dominant symbiont genus, *Cladocopium*. Convex hulls enclose points representing samples from the same location.

To place these Western Australian populations in a broader context we downloaded publicly available whole genome sequencing data from 5 *A. digitifera* colonies sampled from

Okinawa, Japan (NCBI Bioproject PRJDB4188; Shinzato et al. 2015) and for which the sequencing depth was similar to that of our study (16- 19x). Using allele counts at established genome-wide markers for phylogenetic inference in *Acropora* (Cowman et al. 2020) we built a phylogenetic tree (using a polymorphism-aware model, HKY+P, in IQ-TREE) that included Western Australian and Japanese *Acropora digitifera* as well as outgroup species *A. millepora* and *A. tenuis* (Supplementary Fig 3.5). This placed all *A. digitifera* populations within the same clade and placed the Japanese samples outside those from Western Australia. The longest branch lengths within the *A. digitifera* clade were around 40-fold shorter than between *A. digitifera* and *A. millepora*. Consistent with this relatively low divergence between *A. digitifera* populations, we also found that all four shared a single dominant mitochondrial haplotype (Supplementary Fig 3.6) with few samples showing any variation from it. We also found that when a conventional phylogenetic approach (ignoring allele frequency shifts) was used for the same markers it was unable to resolve differences between Western Australian or Japanese populations, or the published *A. digitifera* reference genome (Supplementary Fig 3.7). All four populations are therefore likely to be conspecific and congruent with the published *A. digitifera* genome.

3.3.2 Symbiont profiles

Based on the relative proportion of reads classified as Symbiodiniaceae by Kraken (Wood and Salzberg 2014) all samples from all locations were dominated by symbionts from the genus *Cladocopium* (Supplementary Fig 3.8) which is the most common and diverse genus of symbiont in Indo-Pacific corals (LaJeunesse et al. 2018). To investigate the symbiont diversity within *Cladocopium*, we used three complementary approaches, all of which indicated that there was little difference in symbiont composition between locations. Firstly, a haplotype network based on consensus mitochondrial sequences (Supplementary Fig 3.9B) for 41 samples where there was sufficient data (at least 20X mapping depth at mappable sites) revealed that all but one of the 41 samples were dominated by a single haplotype. This represents a much lower level of diversity than was observed in a previous study using the same approach to profile symbionts in *A. tenuis* on the GBR (Cooke et al. 2020). Since mitochondrial genomes are rarely used to profile Symbiodiniaceae (Waller and Jackson 2009; Gagat et al. 2017), and cannot easily be linked to known types, we also mapped the putative symbiont reads to the more commonly used phylogenetic marker of ITS2 sequences, using the SymPortal database (Hume et al. 2019). This revealed a single ITS2-type profile comprising C40c, C72, C40, and C40e which occurred in most coral samples (Supplementary Fig 3.9A). Finally, in order to minimise inherent biases in ITS2 or mitochondrial markers, we adopted an alignment-free approach based on analysis of shared

k-mers (i.e. short sub-sequences of defined length k) (Reinert et al. 2009; Chan et al. 2014) in the symbiont reads to calculate a distance measure between all possible pairs of samples (see methods). An MDS plot based on this metric (Figure 3.1E) revealed similar levels of within-location to between-location diversity, confirming that there were no consistent differences in symbiont composition between locations.

3.3.3 Demographic history and divergence times

To explore changes in effective population size (Ne) and to estimate divergence times among the coral populations identified above, we performed demographic modelling using two complementary approaches, SMC++ (Terhorst et al. 2016) and fastsimcoal2 (Excofffier et al. 2021). Translating demographic parameters to real timescales for both approaches requires a mutation rate and generation time. Our chosen value of 5 years for generation time is widely used for *Acropora* (Mao et al. 2018; Matz et al. 2018; Cooke et al. 2020) and reflects its fast growth rate combined with the high mechanical vulnerability of older colonies (Madin et al. 2014). For the mutation rate, we calculated an updated value (μ=1.2e-8 per base per generation) based on recently published divergence times (Shinzato et al. 2020). To capture uncertainty in both parameters we ran demographic analyses with SMC++ using alternative published values for the mutation rate (μ=1.86e-8, 2.98e-8 per base per generation) and alternative plausible values for generation time (3y, 7y). Variation in these parameters did not result in qualitative changes to the shape of Ne curves, but generally led to more- recent estimates for key events such as bottlenecks and population splits (Supplementary Fig 3.11).

Changes in effective population size (Ne) during the past 1My inferred by SMC++ revealed qualitatively similar trajectories for the three populations identified in population structure analyses. All experienced a strong bottleneck some time between 7 and 15 Kya followed by expansion and stabilisation. The timing of these bottlenecks coincides with a period of rapid sea level rise at the end of the last glacial maximum (Figure 3.2B). In agreement with the existence of a bottleneck and subsequent population expansion, genome-wide estimates of Tajima's D for all three populations were negative (Supplementary Fig 3.12).

Populations differed in the timing and severity of the bottleneck, with the strongest and most recent effects seen inshore. This was evident in the SMC++ trajectory as well as the much higher prevalence of homozygous-by-descent (HBD) segments in inshore (Figure 3.2F) along with elevated inbreeding coefficients (Figure 3.2E) and linkage disequilibrium (Figure 3.2D). Differences between the two offshore populations were less pronounced than between offshore and inshore, however, it was clear that the north offshore population

retained the highest overall levels of diversity as it had the lowest inbreeding coefficient, the smallest proportion of HBD segments and the highest SMC++ estimated Ne during the recent stable period (2-5Kya).

Divergence time estimates from both SMC++ and fastsimcoal2 indicate a recent split for all three populations that coincides with the same post-glacial time window as bottlenecks observed in SMC++ analyses. Bootstrap estimates for the inshore-offshore split based on the best-fitting model in fastsimcoal2 (Figure 3.2C; Supplementary Table 3.10) were older (5-8Kya) than those between offshore locations (4-5Kya), matching our expectations based on pairwise F_{ST} values and population structure analyses (see above). Estimates from SMC++ were in approximate agreement with this (9Kya) but did not differentiate between inshore-offshore and offshore-offshore splits.

In addition to estimating split times, we used fastsimcoal2 to test a range of competing demographic scenarios (Supplementary Fig 3.14). The results indicate that a model IMc (Figure 3.2C inset) with constant migration between offshore populations and secondary contact between inshore and offshore provides a better fit to the SFS than competing models with strict isolation (SI), ancient migration (AM) or continuous migration (IM) (Supplementary Table 3.10). Support for a model (IMc) with contemporary migration was surprising given the lack of evidence for gene flow in admixture analyses but is reconciled by the fact that estimated migration rates from the IMc model were extremely low (~1e-4) (Supplementary Table 3.10). To confirm that the IMc model is consistent with this and other key features of our data we calculated summary statistics and performed admixture analyses for simulated data under this model. These analyses (summarised in Supplementary Fig S16) showed similar patterns of HBD, inbreeding coefficient and admixture to our results based on sequencing (Figure 3.1) but produced positive values for Tajima's D (negative in our real data). This discrepancy in Tajima's D likely reflects the fact that our simple IMc model was unable to perfectly fit the shape of the 2D SFS at low-medium MAF values (Supplementary Fig 3.18, Supplementary Fig 3.19), a region that has strong effects on Tajima's D. It also highlights the fact that our demographic models did not capture all factors influencing the SFS, potentially including selection across many linked loci or unmodelled bottleneck effects (Gattepaille et al. 2013).

As our estimates of gene flow assume a constant rate across the genome, we also considered the possibility that gene flow was much higher than estimated and that the observed strong population structure was due to barrier loci that (a) maintained ancient divergence (Tine et al. 2014) or (b) enabled divergence under gene flow via spatially or ecologically variable selection (Malinsky et al. 2015; Rippe et al. 2021). We failed to find

evidence for either scenario. The first (barrier loci maintaining ancient divergence) is inconsistent with recent divergence times estimated independently by SMC++ and fastsimcoal2, extremely low admixture coefficients (Figure 3.2B) and the relative rarity of strongly segregating loci in pairwise SFS plots (Supplementary Fig 3.18). Under the second scenario, putative barrier loci should be associated with both high relative divergence (F_{ST}) and elevated absolute divergence D_{XY} (Cruickshank and Hahn 2014; Malinsky et al. 2015). Although we did find a slight increase in D_{XY} in regions of high F_{ST} for inshore vs offshore comparisons the magnitude of this change was small (Supplementary Fig 3.20) indicating that genomic islands were unlikely to be the primary driver of population structure in *A. digitifera* from Western Australia.

Strong bottlenecks and low migration are both potential contributors to population differentiation. To estimate the relative contribution from these factors, we ran simulations based on the IMc model, but with bottlenecks removed by setting a constant effective population size (equal to the ancestral value) and other parameters, including split times and migration rates set to their best-fitted values. Compared with simulations under the full model, removing the bottleneck dramatically reduced pairwise F_{ST}; by fivefold for the inshore-offshore split and 2.5-fold for the split between offshore locations (Supplementary Fig 3.17A).

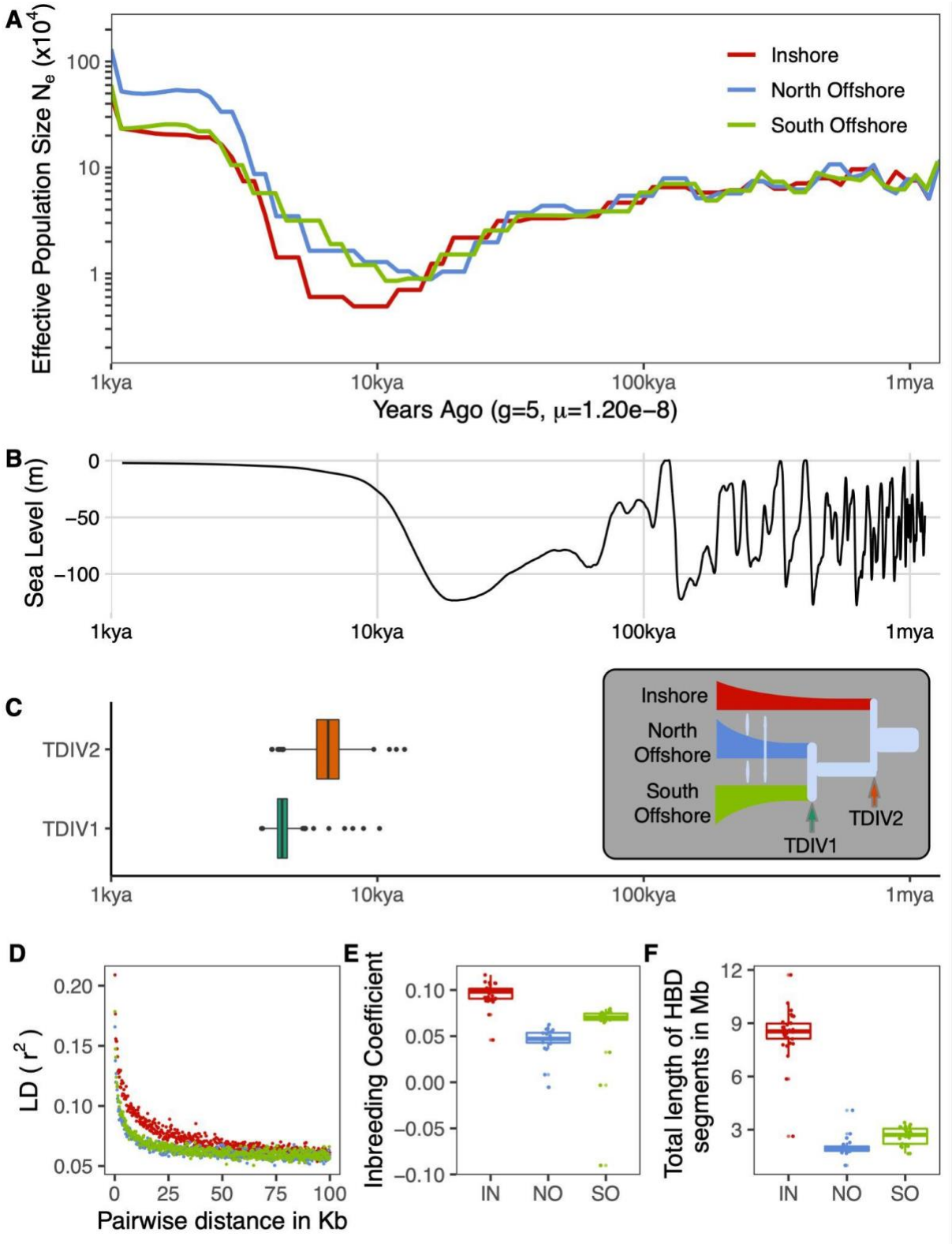


Figure 3.2 Demographic history of *A. digitifera* in Western Australia during the past 1 million years. Locations are denoted by two-letter codes, inshore (IN), north offshore (NO), south offshore (SO) and coloured as shown in A. A. Changes in effective population size (Ne) inferred by SMC++. B. Change in global sea level over the same timescale as depicted in A (data from Bintanja and Wal 2008). C. Estimated divergence times for the inshore- offshore split (TDIV2) and offshore split (TDIV1) obtained using fastsimcoal2. Inset shows the best model; also used to fit bootstrap parameter estimates. D. linkage disequilibrium (LD) decay calculated using plink. E.

Boxplot of the inbreeding coefficient calculated using plink2 for each sample. F. Total length of genomic regions within each individual that were homozygous by descent (HBD) was calculated using ibdseq (Brian L. Browning and Browning 2013). All demographic parameter estimates for both SMC++ and fastsimcoal2 were scaled to real times based on a generation time of five years and an estimated mutation rate of 1.2x10-8 per base per generation.

3.3.4 Genome-wide scan for selective sweeps

To investigate the effects of natural selection on the *A. digitifera* populations identified above we performed a genome-wide scan for signatures of selective sweeps (regions of low diversity arising due to positive selection and linkage to a beneficial allele). As the primary basis for this scan, we used three statistics (iHS, XP-EHH, XP-nSL) that summarise patterns of extended haplotype homozygosity (EHH) because these have high power to detect selective sweeps within independent populations (iHS) (Voight et al. 2006) or as a contrast between pairs (XP-EHH; XP-nSL) (Sabeti et al. 2007; Szpiech et al. 2021). Following standard binning and normalisation practice (see methods; Szpiech and Hernandez 2014) we identified a total of 231 loci (50kb windows) in which at least one of these three statistics was significant (top 1%) based on the frequency of occurrence of SNPs with extreme values. These putative sweep loci were spread throughout the genome (Figure 3.3A; Supplementary Table 3.4, Supplementary Table 3.5) and included 72 specific to inshore, 80 to south offshore, and 79 to north offshore. They were also enriched in SNPs for which the allele-frequency-based indicator of selection, population branch statistic (PBS), had extremely high values (Figure 3.3A).

To control for demographic effects such as bottlenecks we used simulated data under the best-fitting (IMc) demographic model to calculate threshold values for the PBS that would result in fewer than 1% false positives. As expected, given its more severe bottleneck, this threshold was higher for inshore (0.76) compared with offshore populations (NO:0.48, SO:0.44). Even at this higher threshold however, the inshore population had more sweep regions identified by EHH statistics that also overlapped SNPs with significant PBS values (33/72, 45%) compared with north offshore (18/79, 23%) and south offshore (25/80, 31%).

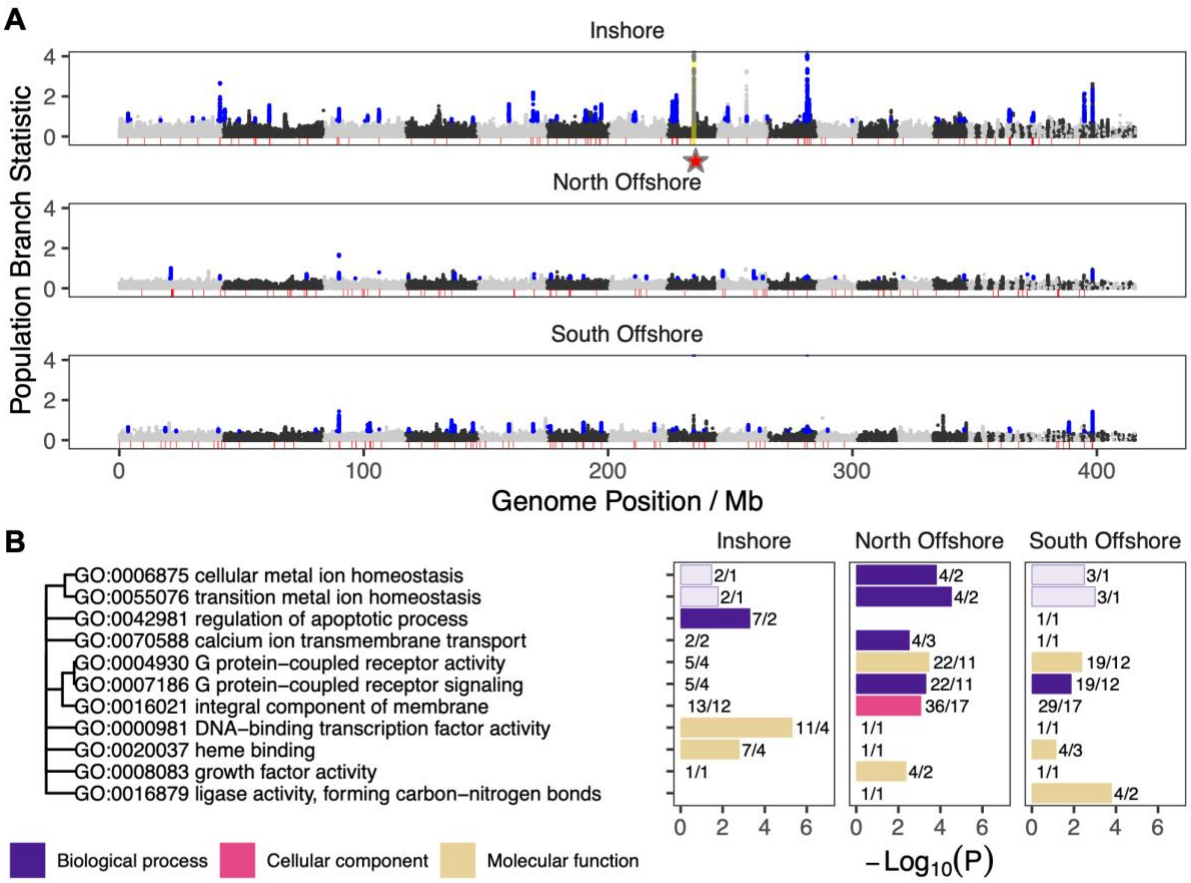


Figure 3.3 Genome-wide distribution of signatures of selection and functional enrichment for overlapping genes. A. Manhattan plots showing values of the population branch statistic (PBS) and regions under selection identified by EHH-based scans. PBS estimates are shown as points for each population and represent allele frequency change since its divergence from the other two. Points are shown in black and grey to indicate transitions between alternating pseudo-chromosomes via mapping to the A. millepora assembly from (Fuller et al. 2020). The red-shaded baseline shows the location of regions identified as candidates for positive selection using EHH-based scans. Blue points indicate PBS values with a probability of false discovery of less than 1% under the best-fitting demographic model, and which are coincident with EHH scans. The yellow highlighted region (also indicated by a red star) in Inshore shows the location of the peroxinectin locus. B. GO term enrichment for regions under selection in inshore and offshore populations. Bar colour indicates one of three broad ontologies, BP: Biological Process, CC: Cellular Compartment, and MF: Molecular Function. Relationships between enriched terms based on numbers of shared genes are shown as a dendrogram (left). The length of the bar indicates the log odds of enrichment (-Log10(p)) based on p-values calculated from Fisher's exact test. Numerical labels indicate the number of genes putatively under selection followed by the number of loci intersected by those genes. Dark-shaded bars show significant enrichment based on the number of genes and the number of independent sweeps while light-shaded bars are significant based on the number of genes but not sweeps.

Of the 1015 genes that overlapped with loci putatively under selection (231 loci identified via EHH-stats; see above), 515 could be assigned a GO term using InterProScan 5 (Jones et al. 2014) based on gene family membership inferred from the presence of conserved domains. Analysis with topGO revealed a total of 11 GO terms across all three ontologies (6 MF;5 BP; 1 CC) that were enriched ($p<0.005$; at least 2 distinct sweep regions) in these genes

(Supplementary Table 3.6) compared with the background in one or more of the three populations (Figure 3.3B). Since multiple genes often overlapped with each sweep region, we also calculated enrichment statistics based on sweep regions rather than genes as independent units and found that all these terms were also enriched (Fisher's exact test $p<0.005$) in at least one population under this criterion (Figure 3.3B).

Three groups of GO terms showed exclusive enrichment in either inshore or offshore locations, potentially reflecting broad patterns of selection related to contrasting environmental conditions. Terms related to membrane G protein-coupled receptors (GPCRs) (GO:0004930, GO:0007186, GO:0016021) were strongly enriched in both offshore populations but not in the inshore, with genes underpinning this pattern distributed across 23 independent sweep regions. Exclusive enrichment in inshore was observed for the GO terms, transcription factor activity (GO:0000981) and regulation of apoptotic process (GO: 0042981). Genes supporting enrichment of transcription factor activity in inshore included a diverse range of transcription factors including those containing homeobox, C2H2 zinc finger, T-box, and forkhead domains, all of which are involved in regulating early development. Enrichment for the GO term, the apoptotic process was supported by two independent sweeps, one containing a Bcl-2-like protein (IPR026298) and another that hosted a cluster of 6 genes each containing a single death effector domain (IPR001875).

3.3.5 Selective sweep at the peroxinectin locus

To investigate the link between selection, climate change, and gene function in additional detail we chose to focus on one of the strongest signatures of selection in the inshore population. This locus was associated with the highest PBS values (yellow highlight and red star in Figure 3.3A), low Tajima's D (Figure 3.4A), and had a clear differentiation between selected and background haplotypes (Figure 3.4B). It also contained by far the largest number (84; next-highest, 7) of near privately fixed SNPs (>90% allele frequency in inshore, absent in offshore), and of these, over 90% were contained within a single gene, s0150.g24 (peroxinectin).

Unlike many other sweep loci where the diversity of genes makes it difficult to associate gene function with selection, four of the five genes overlapping this 50kb sweep region encoded peroxinectin-like proteins (Panther subfamily PTHR11475:SF4; CDD cd09823) and these formed part of a cluster of 8 peroxinectin genes found within 200kb of the sweep. A genome-wide search for haem peroxidases (IPR019791), the parent superfamily that contains peroxinectins, revealed a total of 15 in *A. digitifera*, however only one additional peroxinectin-like gene was found outside the peroxinectin locus. All remaining haem

peroxidases were scattered on different scaffolds throughout the genome indicating that peroxinectins, but not haem peroxidases in general are co-located. Orthologous genomic clusters of peroxinectins were also present in other *Acropora* species (*A. millepora, A. tenuis*; Supplementary Fig 3.21) indicating that the arrangement is at least as old as the crown age of this genus (~50Mya; Shinzato et al. 2021).

The strongest statistical indicators of selection at the peroxinectin locus are centred on the gene, s0150.g24 (Figure 3.4A). An estimate for the timing of selection on this gene based on the inferred time to the most recent common ancestor for selected haplotypes (8.0-8.3Kya; starTMRCA Smith et al. 2018) approximately matches the divergence time for inshore corals. Examination of the age of individual alleles at SNPs in this gene inferred by GEVA (Albers and McVean 2020) showed a pattern consistent with recent selection on ancestral variation. Young alleles (aged less than 15 Kya) had low frequencies in both selected and background haplotypes, consistent with their emergence after the sweep, whereas alleles older than 15 Kya showed a strong shift toward high frequencies in selected haplotypes compared with background (Figure 3.4D, Supplementary Fig 3.22). GEVA estimates the age of a mutation event giving rise to an allele by comparing TMRCA estimates for haplotype pairs where the allele is shared (concordant; younger than the mutation) versus those where it is present in one haplotype and not the other (discordant; older). Although this has been shown to give accurate estimates in humans (Albers and McVean 2020) we expect higher error rates in our study due to a relatively low sample size and uncertainty in input parameters such as the effective population size (Ne).

Examination of the consequences of variants within the gene, s0150.g24 suggests that selected haplotypes may encode a change in exon usage. We identified a total of 10 missense variants in the third exon in selected haplotypes compared with just one at low frequency in the background. Such accumulation of variation in an otherwise conserved region suggests that this exon may no longer be expressed. Although more work is required to confirm this we note that several variants that might encode the change are present, including a change in the splice region between the third intron and fourth exon as well as five variants in the first intron, a region that often contains gene regulatory elements (Chorev and Carmel 2012).

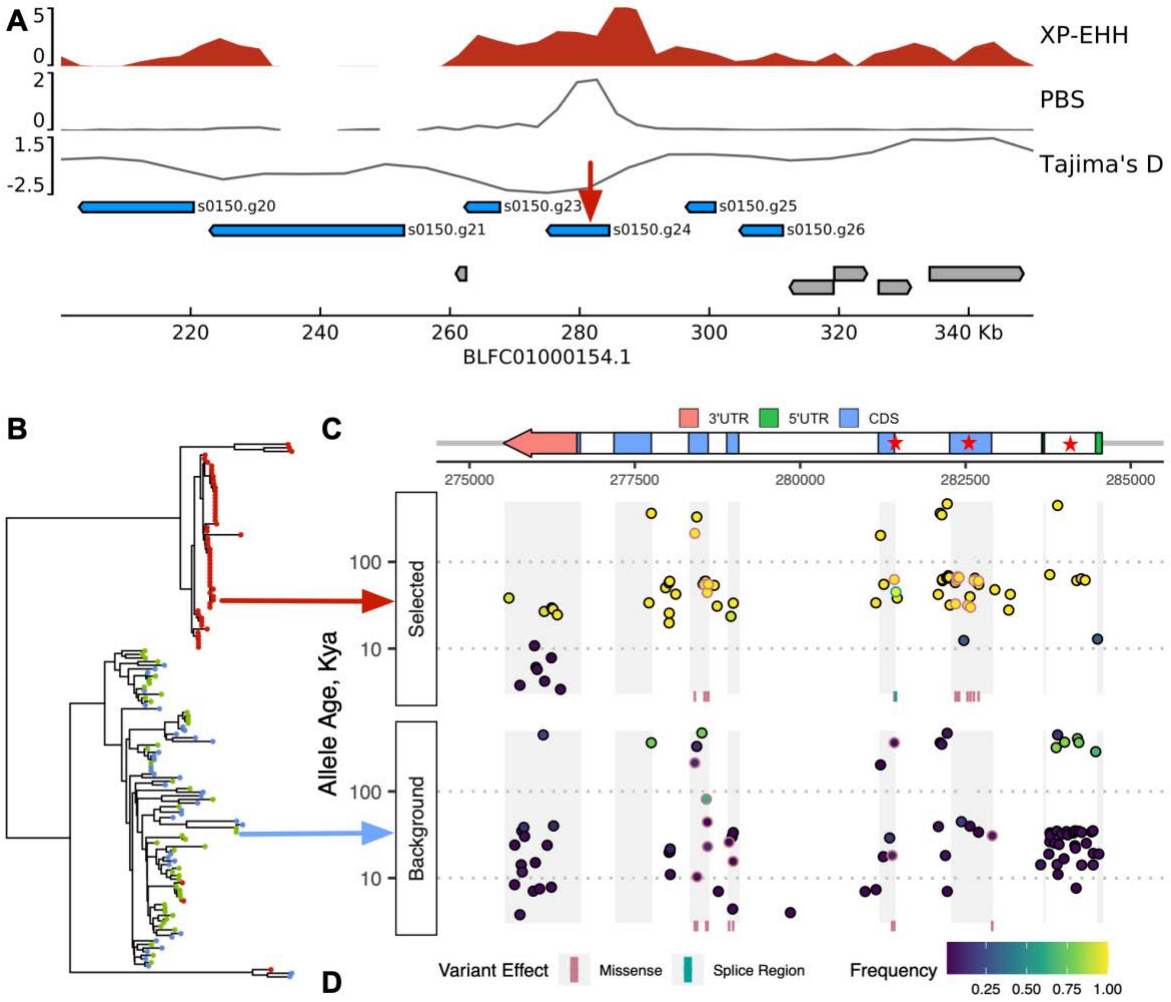


Figure 3.4 Gene arrangement, haplotype structure and timing of selection for a selective sweep at the peroxinectin locus A. Zoomed detail at the locus highlighted in yellow in Figure 3.3A. Tracks show values for XP- EHH, PBS and Tajima's-D for the inshore population. Horizontal bars show the location of genes with peroxinectins in blue and all other genes in grey. B. Neighbour-joining tree (left) based on core haplotypes. Core haplotypes include 200 phased variant sites centred on position 281245 on scaffold BLFC01000154.1 (shown with a red arrow in A). Each haplotype is shown as a terminal branch in the tree and coloured according to sample location. Haplotypes with the derived allele at the focal SNP all partition into the top clade (selected haplotypes) and those with the ancestral allele into the bottom clade (background). C. Gene structure of s0150.g24 showing exons, cds and untranslated regions. Stars indicate key regions of the gene mentioned in the text. From right to left, they are; the first intron, third exon and splice region D. Age, consequence and frequency of variants overlapping the gene s0150.g24. Scatterplots show variants on selected haplotypes (top) and background haplotypes (bottom). Point positions reflect genomic coordinates (x-axis) and age (y-axis). Point fill color shows allele frequency calculated as the proportion of haplotypes with the derived allele in the given population grouping, i.e. selected or background. The position of missense and splice region variants is shown with vertical lines in a strip beneath each scatterplot. Both vertical lines and scatterplot borders for these variants are coloured according to the variant effect category. Grey vertical bars serve as guides to indicate the position of exons.

3.4 Discussion

Our results demonstrate rapid divergence in *A. digitifera* from northwestern Australia resulting in three genetically distinct populations separated by location. Estimated split times of 5-10Kya and similarly timed bottlenecks in all three populations coincide with geological evidence for the post-glacial reestablishment of reef growth on the tops of atolls (Collins et al. 2011) and inshore reefs (Solihuddin, Bufarale, et al. 2016) in this region. Simulations based on our best-fitting demographic model showed that population size changes were a major contributor to overall levels of population differentiation, most likely through increased genetic drift at small population sizes. Limited dispersal indicates that these bottlenecks are likely to represent founder effects arising from post-glacial colonisation, and the two factors (low dispersal and bottlenecks) are the main neutral drivers of divergence.

Since many marine taxa have pelagic larvae and large species ranges it was initially thought that they should exhibit limited or weak population structure (Palumbi 1992; Cowen and Sponaugle 2009). Recent advances in our understanding of larval dispersal in corals and reef fishes have shown that both can be highly variable (Jones et al. 2009), indicating that in specific settings strong population structure may be present (Underwood et al. 2020). In agreement with this, population structure has now been observed for a range of coral reef taxa (Warner et al. 2015; Lukoschek et al. 2016; Underwood et al. 2018; Thomas et al. 2020; Adam et al. 2022) but the mechanisms given rise to this diversity remain poorly understood. Our study demonstrates that population structure can arise rapidly (<10Kya) when dispersal is low, especially if this is combined with the colonisation of new habitats thereby inducing founder effects that enhance drift. Strong selection (as observed in our study) might also contribute to population structure, however, our neutral simulations show that this is not required to account for rapid divergence.

The limited connectivity inferred between locations in northwestern Australia agrees with a growing consensus based on strong genetic structure (Underwood et al. 2009; Thomas et al. 2020; Adam et al. 2022), local recruitment (Gilmour et al. 2013), and limits to the larval movement (Gilmour et al. 2009) that reefs in this region are largely self-seeded. This represents a stark contrast to studies of acroporid species on the Great Barrier Reef (GBR) (Lukoschek et al. 2016; Cooke et al. 2020; Fuller et al. 2020), and the Ryukyu Archipelago (Shinzato et al. 2015). Both *A. tenuis* and *A. millepora* on the GBR form highly connected populations with weak isolation by distance structure over hundreds to thousands of kilometres along north-south stretches of the reef (Lukoschek et al. 2016; Matias et al. 2022). Where highly differentiated populations do exist (eg *A. tenuis*; Cooke et al. 2020) they show signs of recent admixture and likely reflect ancient splits that are now in secondary contact.

This high level of connectivity most likely reflects the fact that reefs in the GBR form a continuous chain with spacing between 1 and 50km (Almany et al. 2009), and those in the Ryukyu are connected by the Kuroshio current (Shinzato et al. 2015). In contrast, reefs in Western Australia are relatively isolated on offshore atolls or inshore islands separated by distances of 100's km (Wilson 2013). The results of this study, therefore, highlight the potential for physical distances combined with a lack of intermediate habitats to act as a barrier to gene flow, even in a broadcast spawning marine species with a pelagic larval stage. It also underscores the importance of historical context and demographic modelling when interpreting measures of genetic differentiation such as F_{ST}. In this case, low F_{ST} did not mean high connectivity as in Wright's Island model (Wright 1931) but was revealed to be due to recent divergence via demographic modelling.

Recent work has also shown that the low levels of divergence between northwestern Australian *A. digitifera* populations also extend to southern inshore sites (Ningaloo Reef) (Adam et al. 2022), which suggests that *A. digitifera* recolonised Western Australia from a single refuge population after the LGM. Low inbreeding coefficients and higher effective population size estimates for the north offshore population are consistent with a refuge at Ashmore Reef or recolonisation via Ashmore Reef from neighbouring Indonesia.

Coral spawning in Western Australia takes place primarily in autumn, with a second smaller event in spring. *Acropora digitifera* is among the majority of corals that spawn in autumn (Gilmour et al. 2016), a time when the Leeuwin Current (LC), a poleward flowing ocean boundary current, is at its strongest, and the potential for current-mediated larval dispersal is at its highest (Feng et al. 2003). This suggests that although the levels of gene flow in our study are low relative to highly connected environments such as the GBR, they may be at the upper end of the spectrum of gene flow for corals in Western Australia. A recent study on the spring spawning lineage of *A. tenuis* identified a strong population structure (F_{ST} >0.25) separating Rowley Shoals and Scott Reef. Although divergence times have not been estimated for *A. tenuis* populations in WA, previous microsatellite work has shown that the species comprises two deeply diverged spawning lineages (Gilmour et al 2016; Rosser et al. 2020). Shallower divergences between sites including between inshore and offshore locations exist within lineages and have been interpreted as arising due to recolonization after the LGM (Rosser et al. 2020). This suggests that the high F_{ST} dividing Rowley Shoals and Scott Reef (Thomas et al. 2022) has arisen rapidly (since the LGM), which points toward even lower levels of gene flow in the spring spawning *A. tenuis* lineage than in autumn spawning *A. digitifera*.

3.4.1 Contrasting selection between inshore and offshore habitats

We identified clear evidence for selection across a wide diversity of loci in all three populations but with the strongest signals observed in the inshore. The inshore reefs of northwestern Australia are notable for their extreme temperatures (short-term maxima of 37 °C), frequent aerial exposure at low tide and highly variable turbidity (Wilson 2013; Solihuddin et al. 2015). The complex, polygenic nature of these stressors, combined with the fact that signatures of selection often cover many genes (due to linkage) make it difficult to identify causal alleles or genes (Dixon et al. 2015; Fuller et al. 2020; Smith et al. 2022; Thomas et al. 2022). As more studies document the effects of natural selection on coral populations it may be possible to identify gene families or pathways that are frequent targets of directional or balancing selection. Our finding that genes involved in the regulation of apoptosis were enriched in selective sweeps unique to the inshore population is similar to a pattern observed by Thomas et al. (2022) where genes encoding NACHT and TNF receptor domain-containing proteins were identified on two of four linkage groups under balancing selection between reef slope (cooler) and lagoon (warmer) habitats in *Acropora tenuis* populations at the Rowley Shoals. Much remains unknown about the complex apoptotic pathways of corals (Moya et al. 2016), however, there is evidence that they play a role in bleaching (Tchernov et al. 2011) and responding to stress (Cziesielski et al. 2019). In the context of inshore corals in the Kimberley however, the fact that we also observed enrichment for transcription factors involved in early development suggests that co-enrichment for apoptotic regulators might also be part of a broader suite of selective pressures related to larval development, metamorphosis, and early growth.

In our study, we identified a highly localised signal on a gene (s0150.g24) within a locus dominated by other genes from the same family (peroxinectin-like haem peroxidases). This provides a rare instance in which a gene family targeted by selection is relatively unambiguous. Peroxinectins are best characterised in arthropods where they mediate the immune response via cell adhesion (Johansson et al. 1995) and prostaglandin synthesis (Park et al. 2014). Heat stress experiments in molluscs (Lang et al. 2009), and corals (Voolstra et al. 2009; Shinzato et al. 2021; Traylor-Knowles et al. 2021) consistently identify peroxinectin- like proteins as differentially expressed, and there is evidence that they have undergone recent expansion in some heat-tolerant coral lineages (Shinzato et al. 2021). Unfortunately, the role of peroxinectins in corals has been obscured because many peroxinectin-like proteins are annotated as peroxidasin homologues in the NCBI nr database. For three key publications (Voolstra et al. 2009; Shinzato et al. 2021; Traylor-Knowles et al. 2021) we manually checked sequences annotated as peroxidasin-like, and that were

differentially expressed in response to heat stress and found that in all cases the corresponding protein sequences had a similar domain structure to the peroxinectins identified in this paper. All contained one or more characteristic conserved domains of peroxinectins (Panther subfamily PTHR11475:SF4 or CDD cd09823) but lacked the N-terminal leucine-rich repeats and immunoglobulin domains found in peroxidasins.

Our results highlight the potential importance of peroxinectins in adaptation to the extreme conditions experienced by inshore corals and invite future work to characterise the evolution and function of co-located peroxinectins in *Acropora* and related taxa. Since the selected haplotypes differ in amino acid sequence to the background, further functional genetic work has a strong chance of identifying the precise nature of the beneficial change, thereby providing a rare opportunity to associate gene function with local adaptive benefit in a wild population.

3.4.2 Implications for coral reefs under future climate change

Our results document the dynamic population responses of *Acropora digitifera* to past climate change. They suggest that this species was likely extirpated throughout much of Western Australia during the last glacial maximum, but recolonised and underwent rapid population expansion when conditions became favourable. Signatures of selection in all three populations indicate that dispersal and diversification were also accompanied by local adaptation via selective pressure on many loci. Of particular interest in the context of future climate change are the inshore Kimberley populations as these corals are known for their ability to survive extreme heat, turbidity and exposure (Richards et al. 2015; Richards et al. 2019). The complex selective pressures resulting from future climate change are difficult to predict, however, there is little uncertainty about the fact that corals will need to adapt to higher temperatures. Understanding the genetic basis for this trait is a key prerequisite for assessing the capacity of corals to adapt. Our finding of strong selection on a peroxinectin gene in the inshore adds weight to existing evidence (Voolstra et al. 2009; Shinzato et al. 2021; Traylor-Knowles et al. 2021) that this may be a key gene family in adapting to heat stress. Moreover, we found that peroxinectins are located in a conserved cluster in corals and therefore expect that variation at this locus may be important in determining the capacity of corals to adapt to climate change.

Identifying the origins of population structure is an essential precondition for understanding the relationship between simple measures of divergence such as F_{ST} and connectivity. We found that *A. digitifera* populations in northwestern Australia diverged recently and that gene flow was particularly low between inshore and offshore sites. Connectivity (and gene flow) in

coral populations is a key deciding factor in their ability to adapt to climate change (Matz et al. 2018) because it allows natural selection to act on a larger overall gene pool, and because it mitigates against local losses. This combination of risk factors (bottlenecks and low connectivity), seen in our study may also be present in other coral reef systems with similar biogeography such as widely spaced offshore atolls and island chains. Our results, therefore, suggest that corals from northwestern Australia and other similar systems may be at a higher risk from climate-related losses than in highly connected systems such as the Great Barrier Reef.

3.5 Materials and Methods

3.5.1 Sample collection and sequencing

Small nubbins of *A. digitifera*, approximately 1-6 cm3 were collected in November 2017 (Rowley Shoals, Ashmore Reef, Adele Island and Beagle Reef) and March 2018 (Rowley Shoals) across our three study locations. DNA extractions were performed by Diversity Array Technology Pty Ltd. (DArT P/L) and extracted DNA was then sent to the QB3 UC Berkeley sequencing centre for whole genome sequencing. Initial sequencing was performed on a single NovaSeq S4 flowcell to obtain ~3 billion 2x150bp paired-end reads across all samples. Additional sequencing was then performed on a second NovaSeq S4 flowcell for 33 samples because they failed to achieve the target depth of 10x in the first batch. Samples included in the second batch of sequencing were spread across all sites in the study (Supplementary Table 3.1) and we did not observe any population structure attributable to the batch in fineSTRUCTURE analyses (Supplementary Fig 3.3). One sample from Inshore (BR_5_121) was likely mislabelled (see supplementary methods) and we excluded it from population structure, demography, and selection analyses.

3.5.2 Variant calling, quality control and haplotype phasing

After verifying that all samples passed read quality checks with FastQC version 0.11.9 and multiQC version 1.6 (Ewels et al. 2016) we then followed the GATK4 (4.1.9) (McKenna et al. 2010) best practice workflow for germline variant calling. Key workflow steps were as follows; raw reads were first aligned to the *Acropora digitifera* reference genome (Shinzato et al. 2011; Shinzato et al. 2020) using BWA version 0.7.17 (Li and Durbin 2009) with the BWA-MEM algorithm; duplicated reads were removed using the MarkDuplicates function in GATK. Next, HaplotypeCaller was used to call variants in each dataset and generate a file in the GVCF format. The GVCFs from all samples were consolidated into a GenomicsDB datastore using GenomicsDBImport and passed to the joint genotyping tools GenotypeGVCFs.

The initial variant call set was filtered with the objective of minimising bias while maintaining quality biallelic SNPs suitable for the population genomic analysis. Filtering steps involved the removal of sites that; (a) were within 5bp of InDels, (b) failed recommended GATK hard filtering quality thresholds, (c) were located within simple repeats, (d) had more than 10% missing genotype calls, (e) had read coverage outside expected bounds. After filtering, we obtained 9,656,554 high-quality biallelic SNPs from 75 samples. A summary of the number of missing genotypes in all samples after filtering is provided in Supplementary Fig 3.1B. The read-aware phasing mode of SHAPEIT v2 (Delaneau et al. 2012) was used to phase all segregating sites in the filtered VCF file. Additional details are provided in supplementary methods.

3.5.3 Genome-wide population genetic statistics

Nucleotide diversity(π), Tajima's D, linkage disequilibrium, and heterozygosity were calculated genome-wide using the unphased, filtered variant set. The het function in PLINK2 (v2.00a3) (Chang et al. 2015) was used to calculate heterozygosity in each sample. Nucleotide diversity and Tajima's D were both calculated in 10kb windows with a 2kb overlap using VCFtools and VCF-kit (Cook and Andersen 2017) respectively. To avoid bias from gaps and masked regions in these window-based estimates, we used BEDTools v2.29.2 (Quinlan and Hall 2010) to remove windows that have less than 70% of bases covered, leaving 136,435 windows. Pairwise linkage disequilibrium (r2) was calculated in 1Mb windows using plink v1.9 (Purcell et al. 2007) based on an equal number (20) of samples from each location. Pairwise F_{ST} for all SNPs was calculated using the weir-fst-pop function in VCFtools.

3.5.4 Population structure

PCA and ADMIXTURE analysis was performed on the unphased, filtered variant set after further filtering to remove sites with minor allele count less than or equal to one, or that deviated from Hardy-Weinberg equilibrium (p-value < 1e-4). SNPs in high linkage disequilibrium (LD) were removed using PLINK v1.9 (—indep -pairwise 50 10 0.1). PCA analysis was performed using smartpca from EIGENSOFT v6.1.4 (Price et al. 2006) with LD-pruned SNPs. Admixture analysis was performed on the same LD pruned data using ADMIXTURE v1.3.0 (Alexander et al. 2009), varying the number of clusters from 1 through to 6. Although the cross-validation error was lowest for K=1, we chose to use K=3 because it reflected the number of clusters seen in PCA and because inference of K=1 is common in situations where overall divergence between clusters is low (Lawson et al. 2012).

We also performed a fineSTRUCTURE (version 4.1.0) analysis (Lawson et al. 2012) on the phased dataset. Inputs were generated by converting SHAPEIT phase files with impute2chromopainter.pl. We assumed a uniform genome-wide recombination rate and allowed the Markov Chain Monte-Carlo (MCMC) to run for 2,000,000 iterations with a burn-in of 1,000,000. Tree inference was performed with 10,000 maximization steps.

Genomic regions inherited by descent (IBD) we identified using the package Refined IBD (Brian L. Browning and Browning 2013). Breaks and short gaps in segments were removed using merge-ibd-segments and pairwise relatedness was calculated based on the total length of shared haplotypes as a proportion of total genome size (Browning and Browning 2011).

3.5.5 Phylogenetic inference based on UCE and Exon probes

To place the *A. digitifera* populations from this study within a broader phylogenetic context we extracted established phylogenetic markers (ultra-conserved-element and exon sequences from Cowman et al. 2020) from our Western Australian samples, previously published data from Japanese samples (Shinzato et al. 2015) (Bioproject PRJDB4188), and published reference genomes for *Acropora millepora* (Ying et al. 2019) and *Acropora tenuis* (Cooke et al. 2020). First, we mapped the hexa-v2 probeset (Cowman et al. 2020) to the genomes of all three species (*A. digitifera, A. tenuis, A. millepora*) using BWA (v0.7.17). A consensus sequence corresponding to a 1000bp interval around the central base of each probe was then called using BCFtools (1.11), with ambiguous bases arising from heterozygous sites encoded using their corresponding IUPAC codes. Consensus sequences for Western Australian samples were called based on bam files generated for variant calling. For Japanese samples raw reads were mapped to the genome using BWA MEM and duplicates were marked using GATK as was done for our own samples. After mapping a total of 16 Japanese samples we selected 5 with coverage >15x (DRR099286, DRR099287, DRR099291, DRR099303, DRR099351). After extracting consensus sequences for all samples we then used MAFFT (v7.394) (Katoh et al. 2002) to align sequences for each (~1000bp) locus separately.

Phylogenetic inference was performed using IQ-TREE (v2.0.3; Nguyen et al. 2015) using; (1) a polymorphism (PoMo) aware approach (Schrempf et al. 2016), and (2) a traditional maximum-likelihood approach that ignores allele frequency changes. The allele count file for PoMo was generated using the Fasta2Counts script https://github.com/pomo-dev/cflib based on alignments across all UCE/Exon loci and the inference was performed using the HKY+P model with 1000 ultrafast bootstraps. For the traditional phylogenetic approach, we used the

same alignments as for PoMo and created a partition file in Nexus format listing them. Using modelfinder (Kalyaanamoorthy et al. 2017) we identified the best model for each partition and used this optimised partition scheme to build a tree with 1000 ultrafast bootstraps (Hoang et al. 2018).

3.5.6 Demographic history with SMC++

SMC++ analysis was performed based on the unphased vcf callset, including only scaffolds with a length greater than N90 (107,903bp). The vcf files of each scaffold were converted into SMC++ input format using the vcf2smc script while masking large uncalled regions. Multiple SMC files were generated for each scaffold by varying the choice of “distinguished individual” over all samples. To estimate population size histories, all SMC++ input files were used together in a single run with the options, thinning 3000, 50 EM iterations, 40 knots, mutation rate 1.20e-8 per base per generation, and starting and ending time points set to 20-200000 generations. Divergence times for each population pair were inferred using the SMC++ split command with marginal estimates produced by using the estimate option. To address the uncertainty in SMC++ analysis from mutation rate and generation time parameters, we tested two additional mutation rates: 1.86e-8 (Cooke et al. 2020); 2.98e-8 (Mao et al. 2018) and three generation times 3, 5, and 7 years (van Oppen et al. 2000; Baria et al. 2012; Matz et al. 2018).

3.5.7 Demographic history with fastsimcoal2

To prepare data for fastsimcoal2 (Excofffier et al. 2021) we used BCFtools to remove sites located in genic regions and performed LD pruning in 1000bp windows with a cut-off of r2>0.3. After removing sites with missing genotypes we used easySFS (https://github.com/isaacovercast/easySFS) to generate a joint three-dimensional folded SFS with 257,314 SNPs. To utilise the mutation rate in branch length calculations, we estimated the number of monomorphic sites based on the proportion of mappable sites defined by the SNPable pipeline.

First, we tested four alternative topologies indicating alternative splitting modes among three populations (Supplementary Table 3.9). For each model, fastsimcoal2 (version 2705) was used to fit parameters to the joint SFS with 50 ECM optimization cycles and 200,000 coalescent simulations. Model fitting was repeated 100 times based on different randomly sampled starting parameter values. We report the best AIC and likelihood values for all four models (across the 100 runs) in Supplementary Table 3.9. Based on the best fitting tree topology ((NO, SO), IN), we then tested six competing models all with exponential population

size change (Supplementary Fig 3.13). Model normalised relative likelihoods (Excoffier et al. 2013) (Supplementary Fig 3.15) supported one of these models (IMc; secondary contact for offshore-inshore but isolation with migration for offshore-offshore). Extended details of the model selection process are provided in supplementary methods. Confidence intervals for the parameters of the best model were estimated using 100 non-parametric bootstraps, each of which was generated by sampling 257,314 SNPs with replacement from the original set of SNPs. For each bootstrapping data set, we performed 20 independent runs. The final results are shown in Supplementary Table 3.10.

3.5.8 Analysis of simulated data under the best fitting model

We generated simulated data under the best fitting parameter set for the IMc model using fastsimcoal2 with an identical model specification file to that used for SFS fitting. We performed 50 independent simulations, each of which used parameters drawn randomly from a uniform distribution across a 90% confidence interval based on our bootstrap estimates (see above). Each simulation generated 20 scaffolds of length 2Mb. Based on this data we then calculated; (1) the length of HBD segments using ibdseq, (2) inbreeding coefficient using plink2, (3) Tajima's D using vk tajima, (4) admixture coefficients using ADMIXTURE, (5) population branch statistics using plink. All calculations were performed using identical settings to those used for real data. The results are shown in Supplementary Fig 3.16.

Simulations based on a modified version of the IMc model were used to assess the contribution of population size changes (ie the bottleneck) to population differentiation. The IMc model was modified so that the total population was conserved at its ancestral size, dividing this at population splits to achieve equal populations in the most recent time period. All other parameters were unmodified. We ran 10 independent simulations using the same process described above with parameter draws allowing variation in divergence times and migration rates but not population sizes. Based on this data we calculated pairwise F_{ST} and performed PCA using plink2. Results are shown in Supplementary Fig 3.17.

3.5.9 Signatures of selection

We used selscan v1.3.0 (Szpiech and Hernandez 2014) with default parameters to calculate test statistics (iHS, XP-EHH, and XP-nSL) based on extended haplotype homozygosity (EHH). Normalization was performed in 50 separate allele frequency bins using the companion program norm. After normalization, SNPs with extreme values were identified genome-wide based on the following criteria (|iHS|>2, XP- EHH/XP-nSL > upper first

percentile). We then calculated the proportion of SNPs with extreme values within 50kb windows and identified windows as candidates for selective sweeps as those in the top 1% based on this proportion. This process was performed separately for each of the three test statistics (iHS, XP-EHH, XP-nSL) and multiIntersectBed (Quinlan and Hall 2010) was used to report the overlapping candidate regions of all tests. Since our goal was to identify sweeps unique to each population we removed those that were significant based on iHS in more than one population. This was not required for the cross-population tests since those already target regions that differ between populations.

We also calculated population branch statistics (PBS) which measure the change in allele frequency in a focal population since its divergence from two other populations. First, we used the --fst function in PLINK to calculate F_{ST} statistics genome-wide for all pairs of populations, using the default F_{ST} calculation (Hudson). These F_{ST} values were then used to calculate the population branch statistic as described in its original paper (Yi et al. 2010). We then used coalescent simulations based on the best-fitting demographic model to determine separate threshold significance values for PBS in each population (see supplementary methods). Our approach differs slightly from the original usage of PBS since we follow Wang et al. 2018 by allowing the outgroup (inshore in this case) to be the focal population and use simulations to control for false positives.

3.5.10 GO enrichment analysis

To support GO enrichment analysis we performed functional annotation of *A. digitifera* genes, assigning GO terms via blast and Interproscan searches (see supplementary methods). The R package topGO v2.42 (Alexa et al. 2006) with the default “weight01” algorithm was used to test for enrichment of GO terms assigned to genes within sweep regions. In this analysis, all genes overlapping with putative selective sweeps were assigned to the target set and the complete set of all annotated genes was assigned as the background set. Since genes are not randomly distributed across the genome we also performed a second test where GO terms were assigned to sweep regions and not to individual genes. As this test was used as a complement to the first we performed it only for GO terms that were significant at the gene level. For the second test, we first assigned GO terms to all 50kb regions in the genome based on the GO terms assigned to overlapping genes. We then calculated a p-value based on Fisher's exact test by counting the number of sweep regions (a subset of all 50kb regions) with a given term and comparing this to the background count across all regions.

3.5.11 Symbiont analysis

Using a custom database composed of the genomes of five common coral associating Symbiodiniaceae genera and the *Acropora digitifera* genome assembly we classified raw reads from all samples using Kraken v1.0 (Wood and Salzberg 2014). This confirmed the dominance of *Cladocopium* in all samples and identified between 4k and 1.7M (median 260k) reads originating from Symbiodiniaceae. Next, we mapped the reads to the mitochondrial genome of *Cladocopium goreaui* and built a haplotype network using PopART (Leigh and Bryant 2015) with the consensus sequences of 41 samples after removing samples with less than 20X average mapping depth (excluding regions with no reads mapped). We also mapped non-host reads to ITS2 sequences from the symportal (Hume et al. 2019) database and quantified their abundance by counting the number of unique mapping reads to each ITS2 reference sequence. Finally, we used an alignment-free method (https://github.com/chanlab-genomics/alignment-free-tools) to calculate the D_2^S metric based on shared *k*-mers in sequencing reads from each pair of samples. This produced a set of pairwise distances which we visualised using an MDS plot (Figure 3.1E).

Although the D_2^S metric has previously been shown to discriminate between whole genome sequences of different Symbiodiniaceae species (Dougan et al. 2022) its power to distinguish differences based on low coverage whole genome sequencing has not previously been established. To establish such a benchmark we used D_2^S statistics to analyse data from a study of *Acropora tenuis* samples in the great barrier reef. Although the overall sequencing depth in that study was much lower than ours (approx 2-3x per sample) we found that D_2^S statistics successfully recapitulated observed differences identified through a mitochondrial haplotype network (Figure 2 in Cooke et al. 2020 vs supplementary Fig 3.10). This power to detect differences despite low overall coverage is surprising if one considers genome coverage to be uniform. We found, however, that despite having a coverage of less than 0.4x there were over 3.5 million sites covered by at least one read in at least 40 samples. These regions (likely repeats) provide for shared *k*-mers between samples and thereby provide power even at low overall coverage.

3.5.12 Estimating the timing of selection at the peroxinectin locus

We used the R package starTMRCA (commit cf9f021 from GitHub) (Smith et al. 2018) to estimate the timing of selection at the peroxinectin locus. Since we did not know the beneficial allele (required by starTMRCA), we instead identified alleles likely to be in complete linkage with it to serve as its proxy. We did this by choosing sites for which the derived allele was nearly fixed (on all but 3 haplotypes) in the inshore population and

completely absent offshore. There were 84 such SNPs within the sweep locus, of which 75 were found within the gene s0150.g24 that overlapped with the strongest statistical indicators of selection (Figure 3.4A). Of these 75 sites, we chose 3 spanning the length of the gene (at positions 278594, 281245, 282923). After performing visual checks of haplotype structure (see supplementary methods) we then ran starTMRCA separately for each of the 3 chosen SNPs using a 1Mb phased region around the centred on s0150.g24. Other parameters were; a mutation rate of 1.2e-8 per base per generation, a recombination rate of 3.2e-8 per base per generation (see supplementary methods), chain length of 10000, proposal standard deviation of 20, the initial value of TMRCA drawn from a uniform distribution from 0-10000 generations. Convergence was checked by running 10 independent chains and calculating the Gelman diagnostic using the coda package in R. For each SNP we recorded the median value of the posterior estimates of the TMRCA after discarding the first half as burn-in. Our final estimate for the time of selection on the locus is reported as the range of estimated values across these three SNPs.

3.5.13 Estimating allele age with GEVA

To estimate the time of origin for derived alleles in the peroxinectin locus we used Genealogical Estimation of Variant Age (GEVA) (Albers and McVean 2020). First, ancestral and derived alleles were polarised using est-sfs (Keightley and Jackson 2018) (see supplementary methods). GEVA was run assuming an effective population size of 30000, mutation rate used of 1.2e-8 per base per generation, and recombination rate ($3.2e^{-8}$ per base per generation) as used for starTMRCA.

3.5.14 Phylogenetic analyses of haem peroxidases

To investigate the evolutionary origins of the peroxinectin locus we used blastp to search for homologous genes in four other coral species, *Acropora millepora*, *Acropora tenuis*, *Porites lutea* and *Pachyseris speciosa*. Protein sequences for all genes identified as belonging to the haem peroxidase family (IPR019791) by Interproscan were extracted from *Acropora digitifera*. Using these as query sequences we identified all close homologs (e-value < $1e^{-10}$) from the protein sets of all other species using blastp. These were then aligned using the MAFFT (v7.394) (Katoh et al. 2002) with the algorithm set to auto. After masking positions with more than 50 missingness, IQ-TREE (v2.0.3; Nguyen et al. 2015) was used to perform tree inference based on this alignment with 1000 ultrafast bootstraps and automatic model selection using modelfinder.

CHAPTER 4

A workflow and *k*-mer-based approach to dissect the symbiont diversity in coral whole-genome sequencing data

---

4.1 Abstract

Understanding the genetics of coral-associated Symbiodiniaceae is as essential as studying that of their coral hosts. While whole-genome sequencing is increasingly being used in coral population genomics the symbiont data incidentally obtained by these studies are rarely used. As the symbiont fraction of reads from whole coral tissue typically comprises only a few percent of each sample, metagenomic approaches relying on deep coverage such as binning and assembly cannot be applied. Alignment-free approaches, however, provide a straightforward way to investigate symbiont reads without the need for reference genomes or deep coverage. *k*-mer-based distance metrics such as D_2^S statistics, based on shared substring (*k*-mer) occurrences in sequences, provide an intuitive way to evaluate sequence differences and have been widely used in phylogenomic studies. Here we applied this method directly to short reads of symbiont origin in whole genome sequencing datasets of corals. Using both simulated and empirical data, we show that *k*-mer-based D_2^S statistics are not only able to differentiate datasets from different Symbiodiniaceae genera but can also resolve differences between samples with symbiont communities dominated by a single genus (*Cladocopium*). Our Python implementation of this method can quickly generate pairwise distances based on D_2^S statistics directly from sequencing data. It is suitable for investigating the inter-sample diversity of metagenomic data in population genomic studies with large sample sizes.

Keywords: Symbiodiniaceae, alignment-free methods, genetic diversity, coral

4.2 Introduction

The symbiotic relationship between corals and their photosynthetic microalgae is essential for the health of reef ecosystems. Through the development and widespread application of marker gene and whole genome sequencing approaches it has become clear that the algal partners in this symbiosis are diverse, comprising dinoflagellates of the family Symbiodiniaceae. Importantly, some coral-algal partnerships appear to be more stable under heat stress (Lien et al., 2007; Silverstein et al., 2015) while others may be important for survival across a range of environmental conditions including depth, turbidity and temperatures (Hoadley et al., 2021; Morikawa & Palumbi, 2019; Tonk et al., 2013). Understanding the implications of diversity in coral-algal symbioses is a major focus of research because it provides potential for coral to adapt to climate change related stressors (Baker, 2003; Sampayo et al., 2008), but progress is hindered by limitations of current marker gene approaches (ITS2, and psbA^{ncr}) (B. C. C. Hume et al., 2019; Stat et al., 2011; Wham et al., 2017; Ziegler et al., 2018) as well as the fact that relatively few studies collect sequencing data for both coral host and symbiont across the same samples. Sequencing data from coral samples contains a significant (but highly variable 1-50%) fraction of reads of symbiont origin, however, these reads are usually discarded as contaminants without further analysis. The increasing number of coral population genomic studies adopting a whole-genome sequencing (WGS) approach and an even larger number of RNA sequencing studies therefore presents a largely untapped opportunity to learn more about coral-algal symbioses.

Alignment-free (AF) sequence methods allow comparisons to be made between individual sequences or groups of sequences without the need to align them to each other or to a reference. The main advantages of AF-based approaches are their computational efficiency (which increases with the number and length of sequences), and their robustness to genome rearrangement events (Chan et al., 2014). A widely used class of AF approaches involves calculating statistical measures of similarity/dissimilarity based on *k*-mer profiles (Zielezinski et al., 2017; Luczak et al., 2017). This is exemplified by the D_2 statistic and its variations which summarise the similarity between two sequences or groups of sequences using the number of shared *k*-mers (Wan et al., 2010). Normalisation modifications to D_2 statistics, such as D_2^S, have been developed to avoid bias from different sequence lengths and nucleotide content by self-standardising based on the probability of the occurrence of a specific *k*-mer in the sequences (Reinert et al., 2009).

Recently, D_2^S statistics have been used in phylogenetic analyses of Symbiodiniaceae based on data from whole-genomes or large genomic regions, revealing congruence with alternative multi-gene alignment-based phylogenies as well as distinct phylogenetic signals caused by differential selective pressure (Dougan, González-Pech, et al., 2022; González-Pech et al., 2021; Lo et al., 2022). In this study, we report an adaptation of this method to analyse raw unassembled sequencing reads arising from a (potentially) mixed community of symbionts in coral whole-genome sequencing data. We also report a Python implementation of this framework which can efficiently calculate the pairwise distances based on D_2^S statistics from *k*-mer counts derived from high-throughput sequence data. Using both simulated and empirical data we show that D_2^S statistics can differentiate different Symbiodiniaceae genera with as few as 10K short reads. When applied to real data from the GBR and the Kimberley, Western Australia, the method recovered patterns of variation in symbiont composition congruent with those identified via alternative methods.

4.3 Materials and Methods

4.3.1 Simulated and empirical sequencing data

To simulate whole-genome short-read sequencing data, we used InSilicoSeq v1.5.4 (Gourlé et al., 2019) to produce datasets representing each of the five available Symbiodiniaceae genera. Each genus is represented by the genome of one of the following species: *Symbiodinium microadriaticum* (Nand et al., 2021), *Breviolum minutum* (Shoguchi et al., 2013), *Cladocopium goreaui* (Liu et al., 2018), *Durusdinium trenchii* (Dougan, Bellantuono, et al., 2022), and *Fugacium kawagutii* (Liu et al., 2018). We generated three datsets for each species with varied amounts of simulated 125bp paired-end Illumina short reads (10K, 100K, and 500K).

To complement simulated datasets, we used whole-genome sequencing data of coral populations from the Great Barrier Reef (Chapter 2). *Cladocopium* was identified to be the dominant genus across all samples (except one from the GBR reefs dataset in Chapter 2 where *Durusdinium* was dominant). We thus implemented this method to infer the within-genus genetic diversity using samples dominated by *Cladocopium* spp. Using bwa (v0.7.17) and samtools (v1.16.1), we extracted reads that did not map to the coral host genome assemblies of *Acropora tenuis* (Cooke et al., 2020; Shinzato et al., 2011) but could be aligned to the genome assembly of *C. goreaui* ((Liu et al., 2018) or ITS2 sequences (downloaded from https://symportal.org/ published named sequences) classified as *Cladocopium* from Symportal (B. C. C. Hume et al., 2019). To avoid biases due to the different numbers of available reads in each sample, we downsampled samples with more

than 160 thousand reads and excluded any samples with fewer reads than 160 thousand. 198 out of 226 samples were used in the analysis after removing coral samples that are not *A. tenuis* and samples with less than 160k reads from symbionts.

4.3.2 Preprocessing

Sequences from each sample can be provided in FASTQ or FASTA format. Jellyfish count v2.3.0 (Marçais & Kingsford, 2011) was first used to count the occurrence of all substrings (*k*-mers) in samples. The hash size parameter "-s " used in jellyfish count should be chosen to fit as many *k*-mer as possible across samples to ensure the same hash size is used for all samples. The choice of k affects memory usage and the sensitivity of the analysis since a large k will result in fewer shared *k*-mers. The optimal choice of k was determined by comparing the statistics of a series of values (from 13 to 33, incrementing by 2) and choosing the smallest value that achieves a high proportion of distinct and unique *k*-mers in the dataset (Supplementary Figure 4.1) (Greenfield & Roehm, 2013). For each dataset, we used jellyfish stats to compute the total numbers of distinct and unique *k*-mers, where distinct *k*-mer refers to the number of all *k*-mers without multiplicity and unique *k*-mer refers to the number of *k*-mer present exactly once (singleton) in the dataset. Based on this analysis, a single k (k=25 for simulated data and k=17 for empirical data) was used for subsequent analyses.

4.3.2 Normalised D_2^S statistics and D_2^S distance

To quantify the similarity between sequences, the D_2 statistic is defined as the count of the exact word (*w*) matches of length *k* shared between two sequences (Marr et al., 2018). As a variation of the D_2 word count statistic, the D_2^S statistic (Reinert et al., 2009; Shepp, 1964; Song et al., 2014) accounts for any differences in background noise by normalising a D_2 score based on the expected number of the specific *k*-mer in the sequences. Let X_w and Y_w be the number of occurrences of a specific word *w* in the sequences of sample X and sample Y. The normalised $\tilde{X}_w$ and $\tilde{Y}_w$ are defined as

$$\tilde{X}_w = X_w - np_w^X \text{ and } \tilde{Y}_w = Y_w - mp_w^Y$$

where n and m are the numbers of all possible *k*-mers in sequences of sample X and sample Y, respectively. p_w^X and p_w^Y represent the probability of the word *w* in sequences of sample X and sample Y which are calculated by the multiplication of frequencies of nucleotide bases in the dataset with the combination of the *k*-mer. Then the D_2^S score is calculated as

$$D_2^S(XY) = \sum_{w \in A^k} \frac{\tilde{X}_w \tilde{Y}_w}{\sqrt{\tilde{X}_w^2 + \tilde{Y}_w^2}}$$

Eventually, for each D_2^S score for one pair of samples, the score was transformed to a distance, *S* following Chan et al. (2014) via taking the absolute logarithmic representation of the geometric mean, as defined:

$$S = \left| \ln \frac{D_2^S(XY)}{D_2^S(XX) D_2^S(YY)} \right|$$

4.3.3 Implementation and benchmark

Previous implementations of D_2^S statistics such as jD2Stat (Chan et al., 2014) and scripts available at https://github.com/chanlab-genomics/alignment-free-tools were primarily developed with a focus on phylogenomics, where datasets are typically limited by genome size and where the number of pairwise comparisons is typically small (few available genomes). To apply the $\mathrm{D}_2^{\mathrm{S}}$ method in whole-genome sequencing data from population genomics studies we found that performance improvements were required to cope with the large number of pairwise comparisons that arise for datasets with many hundreds of samples.

Our new implementation of D_2^S statistics was developed as a Python command-line tool. The sequence information, such as frequency of each DNA base, number of sequences and total length were required for the D_2^S score calculation of each *k*-mer, our tool contains a get_seqinfo() function to calculate and save values in a list of tuples. For the normalisation and transformation required to calculate the distance *S*, a self-matching D_2^S score was first calculated for each sample with function cal_d2s(). The cross-sample D_2^S scores were then calculated from the shared *k*-mer table and the pairwise distance, *S* can be obtained for the generation of output as a distance matrix. To visualise the results, we applied multidimensional scaling (MDS) to the distance matrix and plotted it with ggplot2. Based on the distance matrix, we built a neighbour-joining (NJ) tree using the R package ape (Paradis & Schliep, 2018) and rooted it with the genus *Symbiodinium* (LaJeunesse et al., 2018).

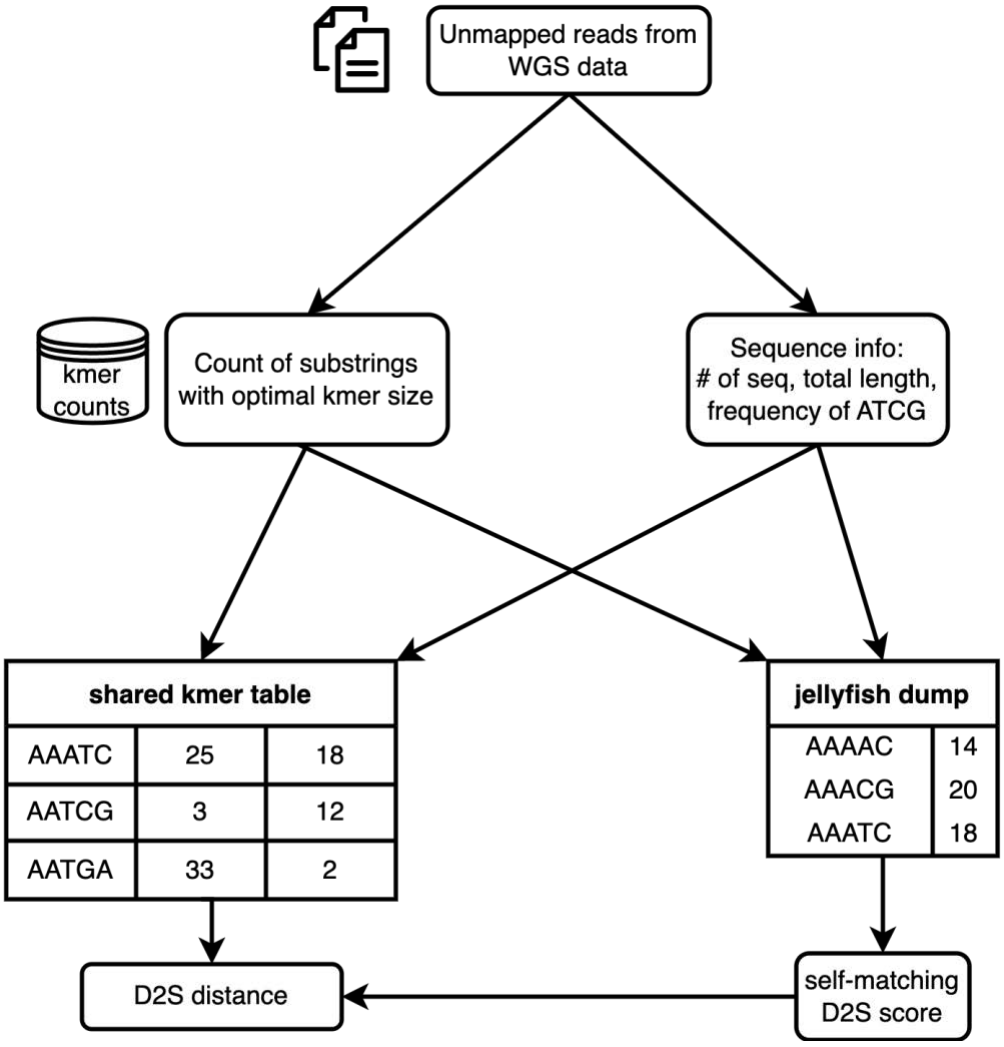


Figure 4.1 The implementation of D_2^S statistics calculation. The expected input files are output from the jellyfish count with the same hash sizes. The summary of sequences, such as the number of sequences, the total length of sequences, and the frequency of nucleotides was calculated. The self-matching D_2^S scores were then calculated and used in the normalisation of the final cross-sample D_2^S score calculation and transformation to *S* distances .

To improve the scalability and speed, we used the Python multiprocessing module to parallelise the calculation of *S* distances across multiple sample pairs. Runtime benchmarks were estimated using the Rust tool Hyperfine (https://github.com/sharkdp/hyperfine), averaging three independent runs with the system cache warming up. To evaluate the speed improvement of the core calculation without the confounding influence of core count, we compared the running time of d2ssect versus the previous implementation (https://github.com/chanlab-genomics/alignment-free-tools) for calculating the distance, *S* between two genome assemblies, *S. microadriaticum* (Nand et al., 2021) and *D. trenchii* (Dougan, Bellantuono, et al., 2022). A single thread was used to run d2ssect and the alignment-free-tools for the purposes of benchmarking performance independent of the parallelisation capability.

4.4 Results

We conducted two computational experiments to assess the performance of this D_2^S based method in differentiating sequences from different symbionts with both simulated data and real-life data from coral whole-genome sequencing data. For both datasets, we also benchmarked the computational efficiency of our tool d2ssect (https://github.com/bakeronit/d2ssect) based on the total run time using different numbers of threads.

4.4.1 Simulated data

Multidimensional scaling (MDS) plots based on the *S* distance matrix between simulated samples reveal clear clustering into groups representing the five Symbiodiniaceae genera we used (Figure 4.1). With as few as 10 thousand short reads, the *S* distances, are able to clearly resolve the sequence differences at the genus level. As expected, the variation between replicates in one genus was smaller when using more data (Figure 4.2, also Supplementary Fig 4.2). Sequences from *Durusdinium* were consistently more distant from the other four genera, which might reflect a putative whole-genome duplication event in *Durusdinium trenchii* (Dougan, Bellantuono, et al., 2022). Nevertheless, a neighbour-joining tree built from the distance matrix of all datasets, and rooted with *Symbiodinium*, reflects the known phylogenetic structure (LaJeunesse et al., 2018)) (Supplementary Fig 4.3).

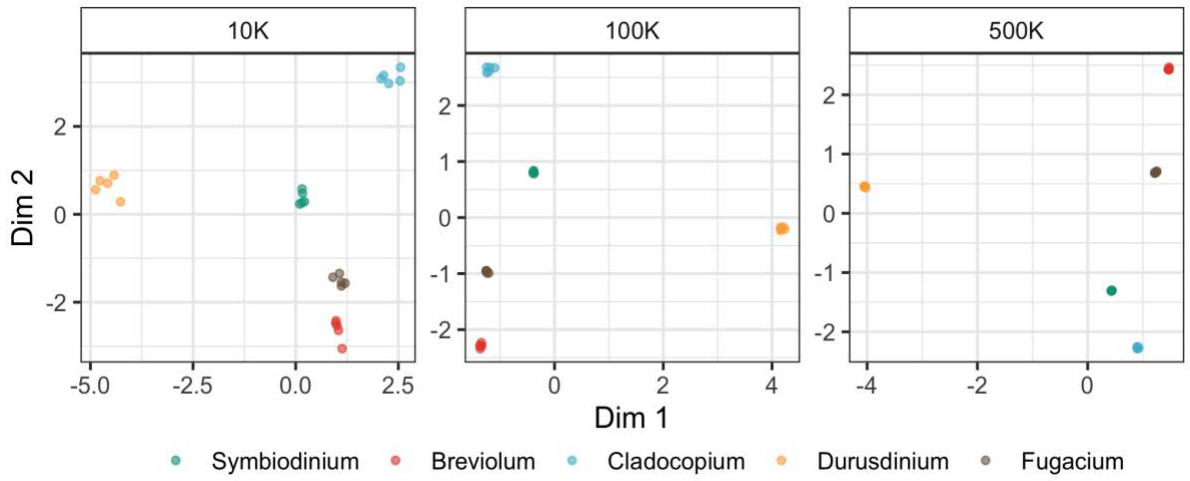


Figure 4.2 Multidimensional scaling (MDS) plots based on a matrix of *S* distances, calculated from simulated read sets (samples) generated from five Symbiodiniaceae genera.

4.4.2 Empirical data: Symbionts of *Acropora tenuis* from the GBR

Reads originating from *Cladocopium* spp. in *Acropora tenuis* whole genome sequencing from the GBR (See Chapter 2) form two clusters representing the inshore and offshore

samples (Figure 4.3). While the clear separation between inshore and offshore samples likely reflects the strong genetic-environment association with symbionts, it is not possible to rule out the possibility that is driven by a batch effect as discussed in Chapter 2.

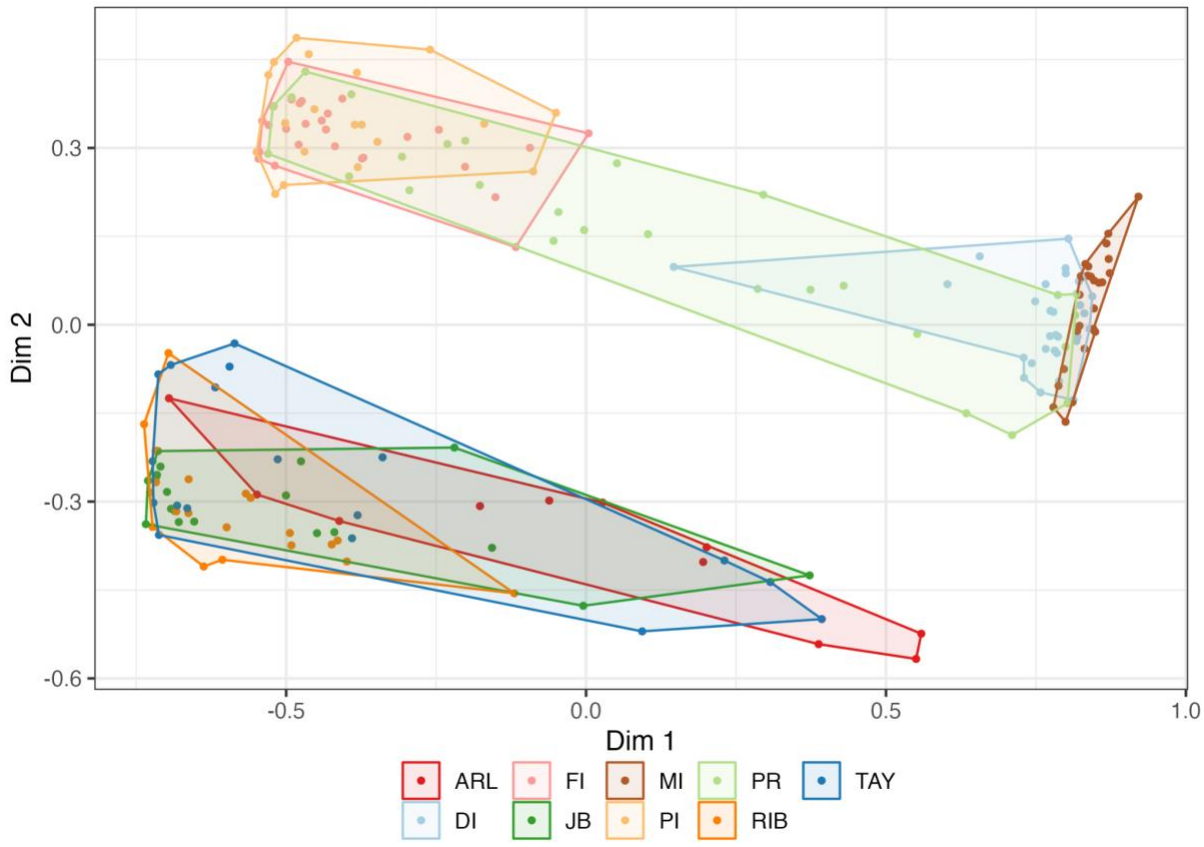


Figure 4.3 MDS plot of *S* distance matrix based on reads of *Cladocopium* origin in *Acropora tenuis* samples from the Great Barrier Reef in Chapter 2.

Within inshore samples (all sequenced in the same batch), the clustering pattern based on *S* distances closely mirrors that observed in the mitochondrial haplotype network in (Cooke et al., 2020) (Figure 4.4). The marine sites composed of PI and FI samples contain very different symbiont compositions as in the plume sites (DI, MI). Corals from Magnetic Island and Dunk Island show greater homogeneity in symbionts than those in Pelorus Island, Fitzroy Island and Pandora Reef. Furthermore, within inshore samples, those from DI and PR, both not clustered with MI in the MDS plot (Figure 4.4b), match the individuals that contain haplotypes different from those of MI in the mitochondrial haplotype network (Figure 4.4a).

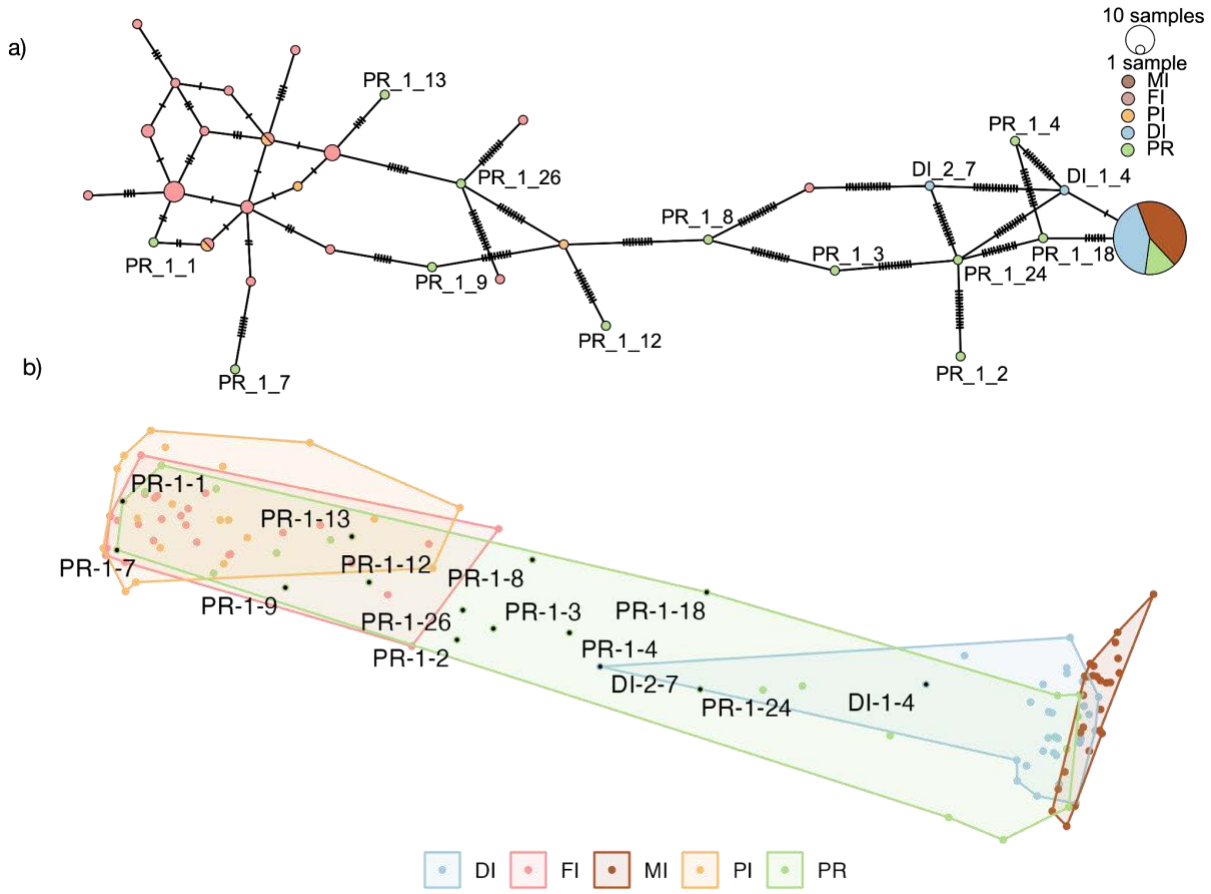


Figure 4.4 a) The mitochondrial haplotype network generated with multi-mitogenome alignment from (Cooke et al., 2020), samples from DI and PR that contain haplotypes different to those of MI were labelled. b) MDS plot of $\boldsymbol{S}$ distance matrix based on reads of *Cladocopium* origin in inshore *A. tenuis* samples. Samples from DI and PR that contain haplotypes different to those of MI in a) were labelled.

4.4.4 Computation time

Our Python implementation uses the Jellyfish library to output a shared *k*-mer table between samples and calculate self-matching and cross-sample D_2^S scores separately. This single change accounts for the majority of speed improvement compared with previous pure-python implementations of D_2^S statistics. It also provides a faster computation for whole-genome data by more than 22.21 times (Supplementary Table 4.1) with a comparable run with a single thread. Expectedly, more samples and more data will take more time to process, however, it will largely benefit from using multiple cores (Figure 4.4; Supplementary Table 4.2). With the support of the multiprocessing library, analysis of multiple samples can be conducted in parallel, thus, it is scalable and time-saving for projects with hundreds of samples as in population genomics study.

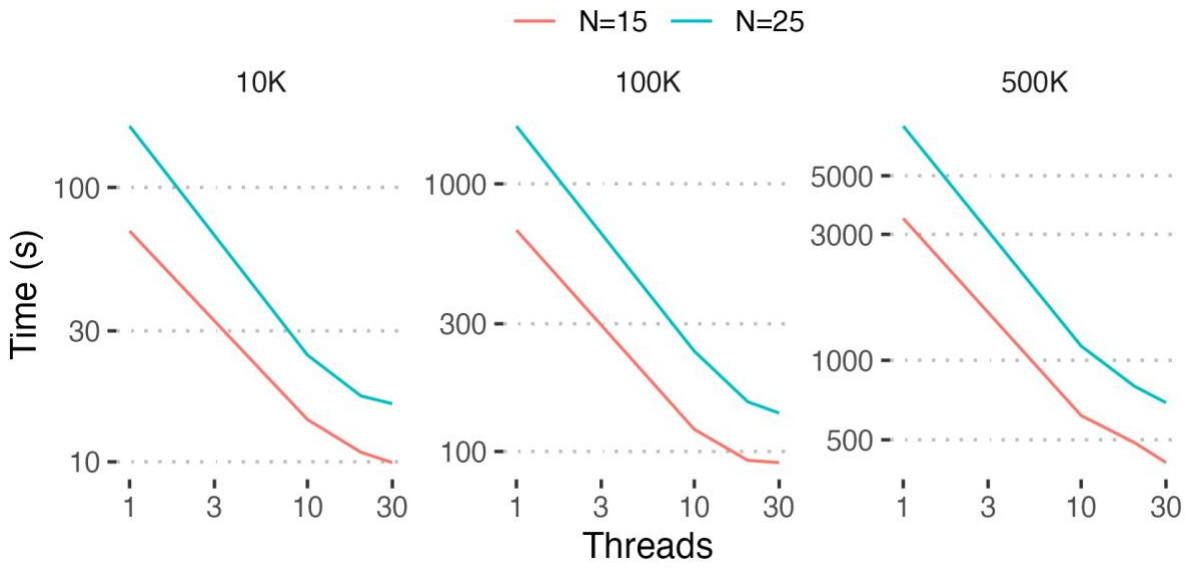


Figure 4.5 The running time in seconds for each run with 15 and 25 datasets and different sequencing coverage using different threads. The y and x axes are in log-scale, such that a straight line indicates perfect scaling (constant performance per core) and slightly bent curves reflect the diminishing returns for the highest thread counts.

4.5 Conclusion and Discussion

The d2ssect tool allows the quantitative evaluation of the genetic diversity of sequencing data across various types, from whole genome assemblies to raw reads. It is particularly useful as a way to make use of unmapped reads in population genomics studies for metagenomic inference without extra amplicon sequencing. This method has been used to classify lineages of microbial eukaryotes at the phylogenetic level (Dougan, González-Pech, et al., 2022). Here, we show that it can resolve the differentiation of mixed communities with the same dominant genus of symbiont. Subsampling to obtain even coverage is an important preprocessing step to minimise unwanted signal, however, the side effect of this is a loss of data in datasets with highly uneven sequencing coverage. Meanwhile, the significant improvement in speed achieved by the d2ssect package benefits existing applications of pairwise *S* distance for phylogenetic inference as well as the proposed new application here for sample clustering from whole genome sequencing data.

The effectiveness and convenience of this tool would greatly benefit from integrated automation of optimising *k*-mer size. *k*-mer size is a critical parameter in alignment-free phylogenetic analysis and the optimal *k*-mer size depends on the length of sequences and the extent of sequence divergence (Bernard et al., 2017). Thus, for sequences with unknown divergence rates, tuning the optimal *k*-mer sizes to the target sequence datasets is necessary to achieve the appropriate resolution. For instance, in empirical data, heuristic

methods were proposed to determine the optimal *k* based on the plateau reached in the cumulative proportions of distinct versus unique *k*-mers in the overall dataset (Chikhi & Medvedev, 2014; Greenfield & Roehm, 2013). This is the method I adopted in this chapter, however, it needs a preliminary *k*-mer counting for datasets with ranges of *k* sizes to generate the accumulative curves and it is challenging to automate the *k*-mer selection.

Alignment-free methods bypass the biological context of comparing two molecular sequences. Although they outperform alignment-based methods in speed and simplicity (Zielezinski et al., 2017, 2019) the underlying biological interpretation of the output of the alignment-free sequence comparison, which is a similarity score, is challenging. In general, the top-ranked scores predict the close sequence homology, however, the length of sequences and *k*-mer size could affect the score distributions (Cattaneo et al., 2021; Forêt et al., 2009; Lippert et al., 2002). Here, we transformed the D_2^S scores into *S* distances that represent the sequence dissimilarity. The matrix composed of *S* distances resolved the consistent phylogeny and diversity as in alignment-based methods. However, extra investigation is needed when matching the distance values to evolutionary distances. For example, jackknife resampling was used to generate node supports of AF trees which appear to be biologically meaningful (Bernard et al., 2016).

Together, our d2ssect tool can be used to quickly analyse non-coral sequences in coral genomic data and assess the differentiation at the sample level by clustering. Although the taxonomic identification of sequences in clustered samples still relies on ITS2 sequences. This tool will therefore be of most value in contexts where at least a subset of samples can be inferred by methods, such as ITS2, and thus the symbiont sequences can be anchored to a known taxonomy.

CHAPTER 5

General Discussion

5.1 Overview

My thesis provides a greater understanding of the population genetics of reef-building corals. It focussed on the widely studied genus, *Acropora* and examined patterns of genomic diversity and demographic history in two case studies. The first of these applied shallow whole genome sequencing to study *Acropora tenuis* populations in the central Great Barrier Reef (Chapter 2), while the second applied deep whole genome sequencing to *Acropora digitifera* from reefs in the Kimberley region of northwestern Australia (Chapter 3). These two studies identified contrasting patterns of population connectivity reflecting the differences in local biogeography and demographic history. Additionally, I investigated the genetic composition of symbiont communities within the same datasets, using the method proposed in Chapter 4, which shows variation in the strength of association between symbiont compositions and habitats. My thesis underscores the importance of local and historical factors as drivers of evolutionary processes in corals and demonstrates the power of whole genome approaches (both shallow and deep) to disentangle the complex factors that contribute to coral evolution and diversity.

As reviewed in Chapter 1, advances in sequencing technologies are facilitating genomic studies of non-model organisms, primarily via improved reference assemblies and reduced costs for genomic studies. My thesis integrated new and existing whole-genome sequencing data from almost 300 *Acropora* coral colonies from two species. Through analyses of these data using a variety of computational tools, I studied the genetic patterns, evolutionary histories, and genomic signatures of selection in *Acropora tenuis* and *Acropora digitifera*. Separately, each *A. tenuis* coral used in Chapter 2 was sequenced at a shallow depth per sample (~3x). This approach, known as lcWGS (low coverage whole genome sequencing) has many advantages over reduced representation (e.g. ddRADseq) or pooled sequencing approaches, however, it must be analysed with care due to the high uncertainty in individual genotype calls at low coverage. By capturing the uncertainty of individual genotypes in a genotype likelihood framework, I was able to examine variation in genetic diversity across the genome and identify several putative structural variants (likely inversions). Deeper sequencing coverage (10-20x) was applied for *A. digitifera* corals in Chapter 3, enabling me

to call accurate genotypes across the whole genome. This type of data allowed me to infer demographic history using sequentially markovian coalescent (SMC) based methods as well as identify patterns of heterozygosity, linkage disequilibrium and regions inherited by descent (IBD). Furthermore, I used read-backed and population-based methods to phase individual genotypes into long-range haplotypes, thereby allowing me to use sophisticated tools that require haplotype information to detect and characterise signatures of selection. As one of the very few studies currently using long-range haplotype information to study evolutionary processes in non-model organisms, my thesis provides a valuable roadmap for future work. Specifically, it demonstrates how to build the required genomic resources, and how to apply quality control checks so that these computational tools developed for humans can be applied to corals and other organisms. Additionally in Chapter 4, I proposed a novel approach to use unmapped reads in coral population WGS data for exploring metagenomic diversity. I also demonstrated the utility of the approach through case studies and provided a fast computational implementation so that it can readily be adopted by other researchers. Thus, my thesis collectively displays the great power and potential of genomic approaches in solving genetic questions of corals and their associated symbionts.

5.2 Population structure and connectivity

Coral populations are structured by a complex array of physical and biological processes, such as larval movement, reproduction strategies, geographic or genetic barriers and demographic history (Goodbody-Gringley et al., 2012; Rippe et al., 2017; Whitaker, 2006). From the two major reef systems that I studied here, I observed contrasting patterns of population structure and reef connectivity based on genome-wide genomic data.

In the central GBR, genetic admixture was detected in *A. tenuis* populations not only between long-distance reefs but also between reefs with deep divergence (Magnetic Island) in close (<50km) proximity. While a strong population structure was observed between Magnetic Island and the other reefs in the north reflecting their deep historical divergence, no structure was found between inshore and offshore reefs suggesting a lack of barriers to gene flow for these reefs both across latitude gradients and distance from shore. This observation of cryptic population structure despite a lack of apparent physical barriers to gene flow has also been previously seen in *A. tenuis* and other taxa based on whole-genome sequencing data (Bongaerts et al., 2021; Cooke et al., 2020; Fuller, Mocellin, et al., 2020; Matias et al., 2022). In all of these previous examples, their cryptic population structure was shown to be the result of ancient divergence (hundreds of thousands to millions of years ago) making it difficult to deconstruct its evolutionary drivers. My thesis uncovered a different story for *A. digitifera* in the Kimberley reefs in northwestern Australia

where migration rates are highly restricted among three reef locations, inshore, north and south offshore. I showed that the split time between these reefs is very recent, reflecting changes to seascapes associated with the end of the last glacial maximum (around 10-20kya). This allowed me to model the drivers of population structure in corals in unprecedented detail. Although the strong population structure I observed can be partially explained by large geographic distances among reefs coupled with a lack of adjacent intermediate reefs, my thesis emphasises that changes in population size, such as founder effects, are also major drivers of this fast separation.

The contrasting patterns of population structure between the East and West Coast highlight the importance of bringing both current and historical context into the interpretation of population differentiation and connectivity. Identifying the sources of population structure is essential for assessing genetic differentiation and connectivity between reefs. My thesis underscores the potential high risks of reduced resilience in less connected Kimberley reefs in northwestern Australia than in the highly connected Great Barrier Reef for future coral conservation planning.

5.3 Evolutionary history

Throughout this thesis, I demonstrate the strong influence of past climate events on demographic history and genetic divergence in the last glacial cycle. Deep divergences with potential periodical gene flow between cryptic populations back to 0.27-1.5 Mya were suggested along the GBR (Cooke et al., 2020; Matias et al., 2022). This coincides approximately with the formation of the GBR and sea-level fluctuations during the late Pleistocene (Hewitt, 2000; Webster & Davies, 2003). A period most *Acropora* species show a pattern of decline in population size with strong introgression (Fuller et al., 2020; van Oppen et al., 2001). I included data from four locations on the central GBR in Chapter 2 and suggest the same origin of inshore and offshore reefs and reflect their recolonisation history from the north refugia on the Queensland submarine plateau (Betzler et al., 1995; Dechnik et al., 2017).

The evolutionary history was more thoroughly studied for *A. digitifera* in the Kimberley (Chapter 3), in which a strong population structure was identified across the sea shelf in northwestern Australia. Demographic history inferences suggest recent bottlenecks followed by divergence in three *A. digitifera* populations. Throughout the chapter, I demonstrate that the recent separations among reefs result from low dispersal and founder effects. Large geographical distances between Western Australian reefs of several hundred kilometres limit the potential for larval dispersal in the area and previous studies suggest high levels of local

recruitment in the Kimberley (Thomas et al., 2020; Underwood et al., 2013, 2020). While such geographical limits to dispersal are well known, a novel finding from my thesis is that recent bottlenecks in reefs after the last glacial maximum (LGM) could speed up the differentiation via drift. The theory of speciation led by founder effects with enhanced drift has been a topic of interest to geneticists for many decades (Matute, 2013; Moya et al., 1995; Templeton, 2008) but has few examples in wild populations. In my thesis, simulated demographic models suggest that the combination of both factors of migration and population size changes is essential for the current population differentiation. Particularly, this study set an example of how marine organisms usually assumed to have weak population structures could separate in a short time.

Although non-neutral processes, such as positive selection, are often the primary focus for many studies, this thesis demonstrates that it is crucially important to first understand and quantify the effects of neutral processes such as demographic change. In Chapter 3 I provide a practical example of how population genetic simulations can be used to quantify the limits of neutral processes on statistical measures used to search for signatures of positive selection.

5.4 Local population structure points to potential inversion polymorphism

Throughout this thesis, I used various genetic statistics to quantify the genome-wide variation between populations. In Chapter 2 I applied this to investigate the sources of islands of differentiation between populations under gene flow. Examining genetic differentiation across the genome between inshore and offshore *A. tenuis* corals on the central GBR (Chapter 2) I identified regions of high and low genetic divergence. Based on both indexes of relative divergence (F_{ST}) and absolute divergence (D_{XY}), the results suggest that islands of strong localised differentiation are likely driven by selective sweeps instead of barriers to gene flow (Cruickshank & Hahn, 2014). In addition to these outliers of strong divergence between inshore and offshore reefs, my thesis characterised the local population structure within north reefs excluding and identifying 6 regions with very strong structures. Within these genomic regions, individuals were roughly clustered into three groups along the first principal component which suggests that these outlier regions could represent the three possible genotypes of a recombination-inhibiting structural variant such as an inversion (Figure 5.1).

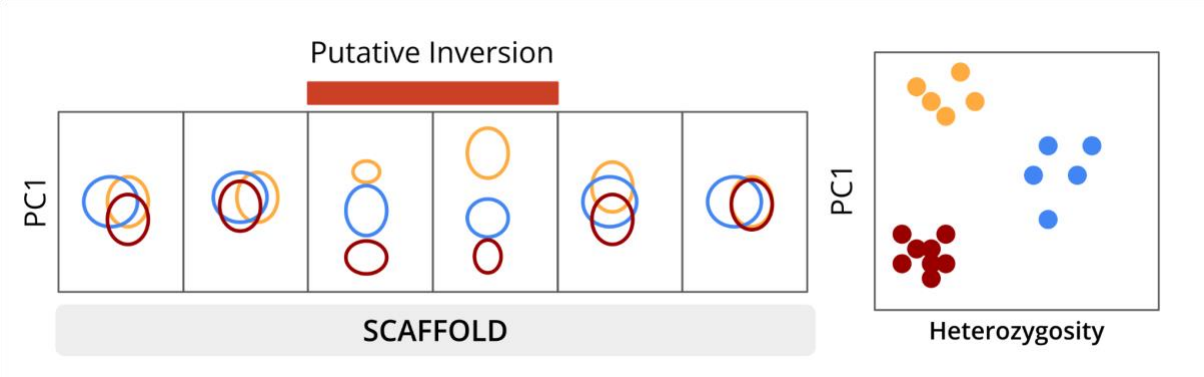


Figure 5.1 Illustration of detection of putative inversion based on local principal component analysis (PCA). The right plot shows further supporting evidence that three PCA clusters are the potential three inversion genotypes.

The ability to use population-level whole-genome resequencing data to detect large structural variations, like inversions, has been shown in studies of seaweed flies (Mérot et al., 2021), European eels (Enbody et al., 2021), and mammals (Harringmeyer & Hoekstra, 2022; Huang et al., 2020) but not previously for corals. Among all types of structural variations, chromosome inversions are widely studied and are suggested to play important roles in intraspecies divergence and speciation as they create derived alleles, decrease local recombination and dramatically increase the accumulation of new mutations (Faria et al., 2019; Huddleston et al., 2017). Evidence has accumulated that many variations happened at large inversions in populations and these chromosomal inversions can suppress the recombination which potentially protects the co-adapted alleles and lead to adaptation and speciation (Mérot, 2020; Wellenreuther & Bernatchez, 2018). Detecting inversions using PCA provides a direct way to genotype each sample and this link highlights the necessity of accounting for linkage disequilibrium (LD) in population structure analysis. Although, in my thesis, the role of the outlier regions based on local PCA in adaptation and speciation has not been fully investigated because of a lack of colony-level environmental and phenotypic data to test for associations with inversion polymorphisms. These within-population inversion polymorphisms I identified could be the putative loci that limit the gene flow among individuals and are potentially favoured by selection.

5.5 Signature of selection from past climate events

Increased frequency of heat-induced mass bleaching and mortality in corals has led to a strong interest in the associations between heat stress adaptation and evolutionary selection in many tropical corals. So far, some of the most important insights have been obtained from studies that compare naturally heat-tolerant corals with counterparts in cooler environments (Palumbi et al., 2014; E. G. Smith et al., 2022; Thomas et al., 2018). However, the

complexity of heat adaptation, diverse selective signals, and uncertain evolutionary scale of the selection mean that the molecular basis of heat resilience remains unclear. In addition, some studies have revealed the key roles of coral-associated symbiont composition and microbiome in physiological performance (B. Hume et al., 2013; B. C. C. Hume et al., 2015; van Oppen & Blackall, 2019). The Kimberley reef system introduced in Chapter 3 provides a valuable case study for evolutionary selection. Inshore corals that live in extreme macrotidal environments show resistance to high temperatures (Richards et al., 2015). The recent separation and strong population structure between inshore and offshore reefs allow us to detect the unique signals in the inshore. Furthermore, the lack of heterogeneity in symbiont composition makes it possible to interpret the findings unambiguously as the genetic changes in coral.

Using whole-genome sequencing data of *A. digitifera* from the Kimberley, my thesis provides the potential genetic mechanisms underpinning the extreme tolerance in inshore intertidal corals (Solihuddin et al., 2016). In Chapter 3, I combined several summary statistics to identify the selective sweeps in inshore and offshore reefs. Scattered signatures were observed across the genome suggesting polygenic effects are responsible for the stress responses; a finding that agrees with previous studies (Dixon et al., 2015; Fuller, Mocellin, et al., 2020; E. G. Smith et al., 2022; Thomas et al., 2022). However, inshore reefs display more intense selection signals and the enriched functions of selected sequences differ between inshore and offshore. The different patterns of selective sweeps highlight the role of local adaptation in population differentiation and reflect the contrasting local environment in inshore-offshore habitats. Importantly, a highly localised signal was identified within those strongest sweeps in inshore reefs and notably, this locus contains a tandem repeat of genes from the peroxinectin-like haem peroxidase gene family. This homogeneous biological function within a sweep provides a clear candidate for a molecular mechanism that may underpin coral thermal resilience. Additionally, this selective sweep identified in inshore corals displays unique long haplotype homozygosity and was suggested to arise from less than 10Kya. Thus, my thesis links the inshore coral resilience to Holocene climate change adaptation and identified the molecular mechanism behind it.

5.6 Symbiont diversity and structure in coral populations

The interplays between corals and their symbiotic algae (Symbiodiniaceae) have been the subjects of considerable scientific investigation, but most previous studies have used separate sequencing techniques to target the host (coral) and symbionts. In my thesis, I developed methods to infer between-sample variation in symbiont compositions from the same whole-genome data used to genotype the coral host. This was possible because the

nature of the endosymbiotic relationship inevitably produces mixed sequences from corals and Symbiodiniaceae. My thesis uses alignment-free methods and D_2^S statistics based on the non-coral reads as a metric to compare samples. Previous studies applied this method to whole-genome data of dinoflagellates to infer phylogenetic structure and argue that whole-genome data is the most comprehensive way to reconstruct evolutionary history since it captures signals from divergent regions (Dougan, González-Pech, et al., 2022; González-Pech et al., 2021; Lo et al., 2022). Although the numbers of symbiont reads per sample comprised only very shallow sequencing coverage I found that it was nevertheless possible to resolve known differences between samples in test data. The use of alignment-free techniques was key here since much of the signal in the data is likely to arise from higher coverage regions (likely repeats) that are difficult to align. My results of D_2^S distances based on whole-genome sequencing data of five common coral-associated Symbiodiniaceae genera demonstrate that D_2^S statistics have the ability to differentiate the genus-level sequence difference consistently with very little data (<0.01X). For re-sequencing data of corals, the non-coral reads are various in both the source and the relative proportions. Although using single values to quantify the difference between two composite sequences dataset is challenging, my thesis provides a different proxy to quickly evaluate symbiont diversity within corals without extra sampling and sequencing. When limiting the reads that are from *Cladocopium* spp., I observed different patterns again in the GBR and Kimberley reefs. The lack of population structure in the coral hosts of the GBR turns out to have a very distinct symbiont composition between inshore and offshore. While corals in the Kimberley are very distinct and disconnected, they have homogenous symbiont profiles. These findings highlight the complexity of the coral-algae symbiosis relationship, the weight of each side can vary under different ecological setups.

5.7 Future directions for coral population genomics

My thesis uncovers the evolutionary history of corals and provides a link between biogeography and molecular adaptation using shallow- and deep-coverage whole-genome sequencing data. The techniques used in this thesis, especially those powered by deep whole-genome sequencing data, are not yet widely used in coral genomics. In the future, I expect that this will change as sequencing costs continue to fall, and more coral species benefit from improved reference genomes. A high-quality reference will largely increase the number of variations one can detect to perform genomic analysis and facilitate the biological interpretation of variation. Meanwhile, factors like ecological importance, conservation concern and potential for scientific research should be taken into consideration when making a priority list for sequencing. I outline here some specific issues in population genomics

research that need to be addressed in the future and that will need to be addressed for the field to take full advantage of the potential that whole-genome-sequencing approaches offer.

5.7.1 Cryptic species in sampling and batch sampling

Sympatric cryptic species have been found to be very common in corals, especially in *Acropora* of which cryptic species have been found distinguishable both morphologically and reproductively (Hayashibara & Shimoike, 2002). While sampling is an important step before implementing DNA extraction and sequencing in genomic studies, misidentified samples are difficult to avoid due to the broad distribution of coral species and the number of colonies that must be sampled for population genomic work. Furthermore, sampling of *Acropora* corals usually takes a small segment of the branch tip and this (alone) may not be sufficient for morphological-based species identification. While the skeleton specimen can be a reference for taxonomic revisions, molecular markers serve as quick tools to differentiate species from population-level samples (Voolstra, Quigley, et al., 2021). Depending on the study designs, it is critical to avoid making inferences about the target species without knowing the existence of cryptic species (Chenuil et al., 2019). Barcoding sequencing of genetic markers such as mtDNA genes 12S rRNA and ITS sequence allows quick exclusion of non-target species before applying the real high-throughput sequencing to all samples.

Sampling and sequencing in several batches are also inevitable in high-throughput genomic studies for practical reasons. While consistent bioinformatic tools can be applied to the data, sequencing service providers and protocol change from time to time (Leek et al., 2010). These batch effects affect the mapping and variant calling process because of different read quality distributions and thus also the following genetic analyses. Ideally, the same samples should be sequenced across batches, as this allows the effect on genetic data to be assessed (e.g. with principal component analysis (PCA)) and potentially removed. Batch effects can be mitigated by using stringent quality controls such as high mapping quality thresholds and trimming poly-G tails, etc (Lou & Therkildsen, 2022). Removing batch effects often leads to loss of data and whether to apply these mitigation measures should always take experiment design into consideration.

Sometimes, batch effects may be difficult to quantify when including sequencing data from previous studies as the samples in each batch do not overlap. In Chapter 2, I combined two datasets from different batches sequenced at a similar time period and with the same sequencing technology. The batch effect did not manifest in the PCA, however, I observed different variations in individual heterozygosity across samples from different batches. Since the batches also correspond to different populations, it is difficult to justify the source of the

difference. In this case, I focussed on genomic loci with statistical measures indicative of strong selection and potential inversions. Since batch effects would be expected to affect the entire genome, such inferences at individual loci should be resilient to batch.

5.7.2 Variant calling and quality control of genomic data

My thesis emphasises the huge potential of using whole-genome sequencing in coral population genomics and demonstrates how tools that were originally designed for human genetic studies can be used for this purpose. The Genome Analysis Toolkit (GATK) is the most used tool in various organisms for genotype calling (McKenna et al., 2010). It was first designed and optimised for human genetics to generate VCF files from BAM files. When applying this tool to coral data, modifications of the GATK best practice germline variant calling pipeline are essential. For example, the HaplotypeCaller uses a default heterozygosity level of 0.001 to determine the likelihood of genotypes whereas corals have much higher heterozygosity (~2%) (Ying et al., 2019). Base quality score recalibration (BQSR) that detects the systematic errors from sequencing machines is recommended by GATK, however, it builds a model based on known variants that are usually unavailable for the non-model organisms. Some suggest generating high-quality variant calls and implementing the bootstrapping BQSR for organisms without known variant data (McCormick et al., 2015; Ottenburghs et al., 2023), however, extra caution is needed since miscalibrated base quality scores could also bring batch effects in data (Lou & Therkildsen, 2022). Most coral genomic studies applied basic filters (based on hard thresholds of mapping quality, read depth, etc) on SNP datasets (Cooke et al., 2020; Fuller, Mocellin, et al., 2020; Shinzato et al., 2015; Thomas et al., 2022; J. Zhang et al., 2022), Matz *et al* took advantage of clones in samples and used reproductive SNPs across clones as “true set” to implement variant quality score recalibration (VQSR) (Matz et al., 2018),. However, the use of a self-generated call set in VQSR model training has not been fully evaluated. Furthermore, a recent study shows that the BCFtools mpileup variant calling pipeline performs better than GATK for non-human data as the HaplotypeCaller tool in GATK is prone to call false heterozygous SNPs at sites with homozygous alternative alleles (Lefouili & Nam, 2022). Overall, variant calling is an essential step in population genetics analysis, the inadequate genomic data processing introduces bias into downstream analyses. With the striking increase in coral genomic data, additional care and customised pipelines are needed, especially for low-coverage data, to make accurate inferences about coral evolutionary scenarios.

5.7.3 Validation of molecular changes under selection

In many coral population genomics studies, lists of selected genes were identified in heat-resilient corals with biological functions of stress response proteins, oxidative and iron binding, etc (Cooke et al., 2020; Dixon et al., 2015; Thomas et al., 2022). These genes are candidates for coral adaptation since they are located in the genomic regions with signatures of positive selection. However, the causal genes or variants are difficult to be determined due to the widespread linkage disequilibrium (LD). The method that uses a composite of multiple signals (CMS) was proposed in human genetics to narrow down the signature region of selection further and potentially identify the causal mutations (Grossman et al., 2010; Lotterhos et al., 2017; Ma et al., 2015). Meanwhile, this method is rarely used in non-human species since it requires the ability to simulate data based on calibrated demographic models which are not available for most species (Ma et al., 2015). In addition, the accuracy of localised signals also relies on the known ancestral allele states and regional recombination rates. In my thesis, a key result worth further exploration is the molecular function of selected genes such as Peroxinectin in Kimberley inshore corals. In Chapter 3, unique selected sweeps were detected in inshore corals which exhibit extreme thermal tolerance. The localised signal identified in a tandem repeat of genes from the peroxinectin-like gene family provides valuable biological insight. Furthermore, missense mutations identified in exons and splicing regions suggest modifications in transcripts and proteins. A test using RNAseq data of inshore corals to confirm the presence of alternative splicing is an important next step. As more coral population genomics data are generated, platforms of molecular validation in corals are needed to thoroughly understand the molecular basis of coral adaptation.

5.7.4 Investigation of coral structural variations using long-read sequencing

Structural variations account for a greater number of nucleotide changes than single nucleotide polymorphisms (SNPs) due to the consequences of segmental losses or gains (Huddleston et al., 2017). The impact of structural variation in population divergence, speciation and rapid adaptation has been studied increasingly with the advancement of long-read sequencing (Coster et al., 2021; Weissensteiner et al., 2020). Structural variations within coral populations remain largely unknown. Our low-coverage short-read data, while adequate for the analyses performed in Chapter 2, is not robust enough for calling large structural variations in populations. However, local PCA analysis revealed genomic regions with increased LD that structured samples into three clusters containing homozygous with reference allele, heterozygous, and homozygous with alternative allele. These patterns fit the characteristics of chromosome inversions (H. Li & Ralph, 2018) and were confirmed in

the deer mouse genome with long-read sequencing data (Harringmeyer & Hoekstra, 2022). Performing long-read sequencing to whole-genome or genome regions of low recombination rates in *A. tenuis* samples in Chapter 3 will be an essential next step for confirming the inversion mutation. Meanwhile, a more in-depth investigation of the inversions and other types of structural variation is warranted in the coral reef system considering their significant impact on recombination, genome divergence, and local adaptation.

5.7.5 Population genomics of coral holobionts

The fact that corals exist as part of complex assemblages, called coral holobionts, has necessitated an updated perspective in coral genetic studies (Cooke et al., 2019; Jaspers et al., 2019; Voolstra, Quigley, et al., 2021). The physiological performance of coral reefs under stress is a interaction of the genotypes of the host, symbionts (photosynthetic algae and microbes) and the environment (Rädecker et al., 2021; Voolstra, Suggett, et al., 2021). It is also suggested that the dynamic changes of microorganism genetics in the host, either through horizontal transfer or mutation, might also drive rapid changes in the host genome, and thus could enhance the adaptation to heat stress (Rosenberg & Zilber-Rosenberg, 2018; Zilber-Rosenberg & Rosenberg, 2008). The divergence of symbiont composition was evaluated using whole-genome data of algae within corals. However, the shallow and uneven coverages across coral samples limited our ability to explore the diversity in high resolution. While the alignment-free methods (Chapter 4) are suitable for estimating sequence divergences despite the lineage, our data contains very few microorganism genomic sequences which also restricts the investigation of the microbial communities in corals. Since the coral holobiont functions as a distinct biological entity and should be treated as a united operational unit for natural selection, the data from coral-associated microbial communities including symbiotic algae and bacteria could provide insight into the within-species diversity of coral holobiont.

5.8 Final thoughts

How coral will adapt to future climate change is determined by the interactions of an untold myriad of factors. Whether “life finds a way” depends on the standing genetic variation for natural selection to act upon. Genetic adaptation enables corals to develop traits and characteristics that enhance their resilience to environmental challenges is crucial for the long-term survival of corals. My thesis offers a glimpse into the genetic landscapes of two critical coral reef systems in Australia and explores the genetic patterns associated with past and current environments. These insights have the potential to enhance the reef management plan. Specifically, I identified novel adaptive genetic variations that can serve

as indicators of coral adaptation potential. When combined with the insights into genetic connectivity patterns and coral-algae associations, these findings allow us to pinpoint coral populations that are resilient to climate change and prioritise them for protection and restoration.

Reference

Abrego, D., van Oppen, M. J. H., & Willis, B. L. (2009). Onset of algal endosymbiont specificity varies among closely related species of Acropora corals during early ontogeny. *Molecular Ecology*, *18*(16), 3532–3543. https://doi.org/10.1111/j.1365-294x.2009.04276.x

Abrego, D., Ulstrup, K. E., Willis, B. L., & van Oppen, M. J. H. (2008). Speciesspecific interactions between algal endosymbionts and coral hosts define their bleaching response to heat and light stress. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1648), 2273–2282. https://doi.org/10.1098/rspb.2008.0180

Adam, A. A. S., Thomas, L., Underwood, J., Gilmour, J., & Richards, Z. T. (2022). Population connectivity and genetic offset in the spawning coral *Acropora digitifera* in Western Australia. Molecular Ecology, 31(13), 3533–3547. https://doi.org/10.1111/mec.16498

Ainsworth, T. D., Heron, S. F., Ortiz, J. C., Mumby, P. J., Grech, A., Ogawa, D., Eakin, C. M., & Leggat, W. (2016). Climate change disables coral bleaching protection on the Great Barrier Reef. Science, 352(6283), 338–342. https://doi.org/10.1126/science.aac7125

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biology, 20(1), 224. https://doi.org/10.1186/s13059-019-1829-6

Anand, S., Mangano, E., Barizzone, N., Bordoni, R., Sorosina, M., Clarelli, F., Corrado, L., Boneschi, F. M., D'Alfonso, S., & Bellis, G. D. (2016). Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. Scientific Reports, 6(1), 33735. https://doi.org/10.1038/srep33735

Anderson, E. C., Skaug, H. J., & Barshis, D. J. (2014). Next‐generation sequencing for molecular ecology: a caveat regarding pooled samples. Molecular Ecology, 23(3), 502‐512. https://doi.org/10.1111/mec.12609

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Andrews, K. R., Hohenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., & Luikart, G. (2014). Trade‐offs and utility of alternative RADseq methods: Reply to Puritz et al. Molecular Ecology, 23(24), 5943‐5946. https://doi.org/10.1111/mec.12964

Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. Molecular Ecology, 23(7), 1661–1667. https://doi.org/10.1111/mec.12686

Ayre, D. J., & Hughes, T. P. (2004). Climate change, genotypic diversity and gene flow in reef‐building corals. Ecology Letters, 7(4), 273‐278. https://doi.org/10.1111/j.1461-0248.2004.00585.x

Baker, A. C. (2003). FLEXIBILITY AND SPECIFICITY IN CORAL-ALGAL SYMBIOSIS: Diversity, Ecology, and Biogeography of Symbiodinium. Annual Review of Ecology, Evolution, and Systematics, 34(1), 661–689. https://doi.org/10.1146/annurev.ecolsys.34.011802.132417

Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. Heredity, 57(3), 357–376. https://doi.org/10.1038/hdy.1986.135

Baskett, M. L., Gaines, S. D., & Nisbet, R. M. (2009). Symbiont diversity may help coral reefs survive moderate climate change. Ecological Applications, 19(1), 3–17. https://doi.org/10.1890/08-0139.1

Bay, R. A., Rose, N. H., Logan, C. A., & Palumbi, S. R. (2017). Genomic models predict successful coral adaptation if future ocean warming rates are reduced. Science Advances, 3(11), e1701413. https://doi.org/10.1126/sciadv.1701413
Beltrán, V. H., Puill-Stephan, E., Howells, E., Flores-Moya, A., Doblin, M., Núñez-Lara, E., Escamilla, V., López, T., & van Oppen, M. J. H. (2021). Physiological diversity among sympatric, conspecific endosymbionts of coral (Cladocopium C1acro) from the Great Barrier Reef. Coral Reefs, 40(4), 985–997. https://doi.org/10.1007/s00338-021-02092-z
Berkelmans, R., & van Oppen, M. J. H. (2006). The role of zooxanthellae in the thermal tolerance of corals: a nugget of hope for coral reefs in an era of climate change. Proceedings of the Royal Society B: Biological Sciences, 273(1599), 2305–2312. https://doi.org/10.1098/rspb.2006.3567
Bernard, G., Chan, C. X., Chan, Y., Chua, X.-Y., Cong, Y., Hogan, J. M., Maetschke, S. R., & Ragan, M. A. (2017). Alignment-free inference of hierarchical and reticulate phylogenomic relationships. Briefings in Bioinformatics, 20(2), 426–435. https://doi.org/10.1093/bib/bbx067
Bernard, G., Chan, C. X., & Ragan, M. A. (2016). Alignment-free microbial phylogenomics under scenarios of sequence divergence, genome rearrangement and lateral genetic transfer. Scientific Reports, 6(1), 28970. https://doi.org/10.1038/srep28970
Betancourt, A. J., Welch, J. J., & Charlesworth, B. (2009). Reduced Effectiveness of Selection Caused by a Lack of Recombination. Current Biology, 19(8), 655–660. https://doi.org/10.1016/j.cub.2009.02.039
Betzler, C., Brachert, T. C., & Kroon, D. (1995). Role of climate in partial drowning of the Queensland Plateau carbonate platform (northeastern Australia). Marine Geology, 123(1–2), 11–32. https://doi.org/10.1016/0025-3227(95)80002-s
Bintanja, R., & Wal, R. S. W. van de. (2008). North American ice-sheet dynamics and the onset of 100,000-year glacial cycles. Nature, 454(7206), 869–872. https://doi.org/10.1038/nature07158
Bongaerts, P., Cooke, I. R., Ying, H., Wels, D., Haan, S. den, Hernandez-Agreda, A., Brunner, C. A., Dove, S., Englebert, N., Eyal, G., Forêt, S., Grinblat, M., Hay, K. B., Harii, S., Hayward, D. C., Lin, Y., Mihaljević, M., Moya, A., Muir, P., … Hoegh-Guldberg, O. (2021). Morphological stasis masks ecologically divergent coral species on tropical reefs. Current Biology, 31(11), 2286-2298.e8. https://doi.org/10.1016/j.cub.2021.03.028
Bongaerts, P., Frade, P. R., Ogier, J. J., Hay, K. B., Bleijswijk, J. van, Englebert, N., Vermeij, M. J., Bak, R. P., Visser, P. M., & Hoegh-Guldberg, O. (2013). Sharing the slope: depth partitioning of agariciid corals and associated Symbiodinium across shallow and mesophotic habitats (2-60 m) on a Caribbean reef. BMC Evolutionary Biology, 13(1), 205–205. https://doi.org/10.1186/1471-2148-13-205
Brodie, K. L., & Cohn, N. T. (2020). Reference Module in Earth Systems and Environmental Sciences. 894–905. https://doi.org/10.1016/b978-0-12-409548-9.12483-2
Browne, N. K., Smithers, S. G., & Perry, C. T. (2012). Coral reefs of the turbid inner-shelf of the Great Barrier Reef, Australia: An environmental and geomorphic perspective on their occurrence, composition and growth. Earth-Science Reviews, 115(1–2), 1–20. https://doi.org/10.1016/j.earscirev.2012.06.006
Browning, B. L., Tian, X., Zhou, Y., & Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. The American Journal of Human Genetics, 108(10), 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005
Browning, S. R., & Browning, B. L. (2011). Haplotype phasing: existing methods and new developments. Nature Reviews Genetics, 12(10), 703–714. https://doi.org/10.1038/nrg3054
Browning, S. R., Browning, B. L., Daviglus, M. L., Durazo-Arvizu, R. A., Schneiderman, N., Kaplan, R. C., & Laurie, C. C. (2018). Ancestry-specific recent effective population size in the Americas. PLoS Genetics, 14(5), e1007385. https://doi.org/10.1371/journal.pgen.1007385

Buerkle, C. A., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: how low should we go? Molecular Ecology, 22(11), 3028–3035. https://doi.org/10.1111/mec.12105
Burgess, S. C., Johnston, E. C., Wyatt, A. S. J., Leichter, J. J., & Edmunds, P. J. (2021). Response diversity in corals: hidden differences in bleaching mortality among cryptic Pocillopora species. Ecology, 102(6), e03324. https://doi.org/10.1002/ecy.3324
Burgon, J. D., Vieites, D. R., Jacobs, A., Weidt, S. K., Gunter, H. M., Steinfartz, S., Burgess, K., Mable, B. K., & Elmer, K. R. (2020). Functional colour genes and signals of selection in colour ‐ polymorphic salamanders. Molecular Ecology, 29(7), 1284 ‐ 1299. https://doi.org/10.1111/mec.15411
Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnstrӧm, A., Ružić, M., Sæther, S.-A., Sætre, G.-P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Research, 25(11), 1656–1665. https://doi.org/10.1101/gr.196485.115
Camp, E. F., Schoepf, V., Mumby, P. J., Hardtke, L. A., Rodolfo-Metalpa, R., Smith, D. J., & Suggett, D. J. (2018). The Future of Coral Reefs Subject to Rapid Climate Change: Lessons from Natural Extreme Environments. Frontiers in Marine Science, 5, 4. https://doi.org/10.3389/fmars.2018.00004
Camp, E. F., Suggett, D. J., Pogoreutz, C., Nitschke, M. R., Houlbreque, F., Hume, B. C. C., Gardner, S. G., Zampighi, M., Rodolfo-Metalpa, R., & Voolstra, C. R. (2020). Corals exhibit distinct patterns of microbial reorganisation to thrive in an extreme inshore environment. Coral Reefs, 39(3), 701–716. https://doi.org/10.1007/s00338-019-01889-3
Cattaneo, G., Petrillo, U. F., Giancarlo, R., Palini, F., & Romualdi, C. (2021). The power of word-frequency-based alignment-free functions: a comprehensive large-scale experimental analysis. Bioinformatics, 38(4), 925–932. https://doi.org/10.1093/bioinformatics/btab747
Chan, C. X., Bernard, G., Poirion, O., Hogan, J. M., & Ragan, M. A. (2014). Inferring phylogenies of evolving sequences without multiple sequence alignment. Scientific Reports, 4(1), 6504. https://doi.org/10.1038/srep06504
Charlesworth, B. (2012). The Effects of Deleterious Mutations on Evolution at Linked Sites. Genetics, 190(1), 5–22. https://doi.org/10.1534/genetics.111.134288
Cheal, A. J., MacNeil, M. A., Emslie, M. J., & Sweatman, H. (2017). The threat to coral reefs from more intense cyclones under climate change. Global Change Biology, 23(4), 1511–1524. https://doi.org/10.1111/gcb.13593
Cheng, A. Y., Teo, Y.-Y., & Ong, R. T.-H. (2014). Assessing single nucleotide variant detection and genotype calling on whole-genome sequenced individuals. Bioinformatics, 30(12), 1707–1713. https://doi.org/10.1093/bioinformatics/btu067
Chenuil, A., Cahill, A. E., Délémontey, N., Luc, E. D. S. du, & Fanton, H. (2019). From Assessing to Conserving Biodiversity, Conceptual and Practical Challenges. History, Philosophy and Theory of the Life Sciences, 77–106. https://doi.org/10.1007/978-3-030-10991-2_4
Chikhi, R., & Medvedev, P. (2014). Informed and automated *k*-mer size selection for genome assembly. Bioinformatics, 30(1), 31–37. https://doi.org/10.1093/bioinformatics/btt310
Cleves, P. A., Shumaker, A., Lee, J., Putnam, H. M., & Bhattacharya, D. (2019). Unknown to Known: Advancing Knowledge of Coral Gene Function. Trends in Genetics, 36(2), 93–104. https://doi.org/10.1016/j.tig.2019.11.001
Colton, M. A., McManus, L. C., Schindler, D. E., Mumby, P. J., Palumbi, S. R., Webster, M. M., Essington, T. E., Fox, H. E., Forrest, D. L., Schill, S. R., Pollock, F. J., DeFilippo, L. B., Tekwa, E. W., Walsworth, T. E., & Pinsky, M. L. (2022). Coral conservation in a warming world must harness evolutionary adaptation. Nature Ecology & Evolution, 6(10), 1405–1407. https://doi.org/10.1038/s41559-022-01854-4

Combosch, D. J., & Vollmer, S. V. (2015). Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific Pocillopora corals. Molecular Phylogenetics and Evolution, 88, 154–162. https://doi.org/10.1016/j.ympev.2015.03.022
Connelly, M. T., McRae, C. J., Liu, P.-J., Martin, C. E., & Traylor-Knowles, N. (2022). Antibiotics Alter Pocillopora Coral-Symbiodiniaceae-Bacteria Interactions and Cause Microbial Dysbiosis During Heat Stress. Frontiers in Marine Science, 08, 814124. https://doi.org/10.3389/fmars.2021.814124
Cooke, I., Mead, O., Whalen, C., Boote, C., Moya, A., Ying, H., Robbins, S., Strugnell, J., Darling, A., Miller, D., Voolstra, C. R., Adamska, M., & Participants, C. of A. A. of S. B. R. C. (2019). Molecular techniques and their limitations shape our view of the holobiont. Zoology, 137, 125695. https://doi.org/10.1016/j.zool.2019.125695
Cooke, I., Ying, H., Forêt, S., Bongaerts, P., Strugnell, J. M., Simakov, O., Zhang, J., Field, M. A., Rodriguez-Lanetty, M., Bell, S. C., Bourne, D. G., van Oppen, M. J. H, Ragan, M. A., & Miller, D. J. (2020). Genomic signatures in the coral holobiont reveal host adaptations driven by Holocene climate change and reef specific symbionts. Science Advances, 6(48), eabc6318. https://doi.org/10.1126/sciadv.abc6318
Corbett-Detig, R. B., Hartl, D. L., & Sackton, T. B. (2015). Natural Selection Constrains Neutral Diversity across A Wide Range of Species. PLoS Biology, 13(4), e1002112. https://doi.org/10.1371/journal.pbio.1002112
Corbett-Detig, R. B., Said, I., Calzetta, M., Genetti, M., McBroome, J., Maurer, N. W., Petrarca, V., Torre, A. della, & Besansky, N. J. (2019). Fine-Mapping Complex Inversion Breakpoints and Investigating Somatic Pairing in the Anopheles gambiae Species Complex Using Proximity-Ligation Sequencing. Genetics, 213(4), 1495–1511. https://doi.org/10.1534/genetics.119.302385
Coster, W. D., Weissensteiner, M. H., & Sedlazeck, F. J. (2021). Towards population-scale long-read sequencing. Nature Reviews Genetics, 22(9), 572–587. https://doi.org/10.1038/s41576-021-00367-3
Cowen, R. K., & Sponaugle, S. (2009). Larval Dispersal and Marine Population Connectivity. Annual Review of Marine Science, 1(1), 443–466. https://doi.org/10.1146/annurev.marine.010908.163757
Crawford, D. L., & Oleksiak, M. F. (2016). Ecological population genomics in the marine environment. Briefings in Functional Genomics, 15(5), 342–351. https://doi.org/10.1093/bfgp/elw008
Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology, 23(13), 3133–3157. https://doi.org/10.1111/mec.12796
Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. Nature Reviews Genetics, 14(4), 262–274. https://doi.org/10.1038/nrg3425
Dandan, S. S., Falter, J. L., Lowe, R. J., & McCulloch, M. T. (2015). Resilience of coral calcification to extreme temperature variations in the Kimberley region, northwest Australia. Coral Reefs, 34(4), 1151–1163. https://doi.org/10.1007/s00338-015-1335-6
Davies, S. W., Moreland, K. N., Wham, D. C., Kanke, M. R., & Matz, M. V. (2020). Cladocopium community divergence in two Acropora coral hosts across multiple spatial scales. Molecular Ecology, 29(23), 4559–4572. https://doi.org/10.1111/mec.15668
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, 109(44), 17995–17999. https://doi.org/10.1073/pnas.1208909109
Dechnik, B., Webster, J. M., Webb, G. E., Nothdurft, L., Dutton, A., Braga, J.-C., Zhao, J., Duce, S., & Sadler, J. (2017). The evolution of the Great Barrier Reef during the Last Interglacial Period. Global and Planetary Change, 149, 53–71. https://doi.org/10.1016/j.gloplacha.2016.11.018

DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). SweepFinder 2: increased sensitivity, robustness and flexibility. Bioinformatics, 32(12), 1895–1897. https://doi.org/10.1093/bioinformatics/btw051

Delaneau, O., Zagury, J.-F., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nature Methods, 10(1), 5–6. https://doi.org/10.1038/nmeth.2307

DeVantier, L., Turak, E., & Szava-Kovats, R. (2020). Species Richness and Abundance of Reef-Building Corals in the Indo-West Pacific: The Local–Regional Relation Revisited. Frontiers in Marine Science, 7, 487. https://doi.org/10.3389/fmars.2020.00487

Dixon, G. B., Davies, S. W., Aglyamova, G. V., Meyer, E., Bay, L. K., & Matz, M. V. (2015). Genomic determinants of coral heat tolerance across latitudes. Science, 348(6242), 1460–1462. https://doi.org/10.1126/science.1261224

Domanska, D., Kanduri, C., Simovski, B., & Sandve, G. K. (2018). Mind the gaps: overlooking inaccessible regions confounds statistical testing in genome analysis. BMC Bioinformatics, 19(1), 481. https://doi.org/10.1186/s12859-018-2438-1

Dougan, K. E., Bellantuono, A. J., Kahlke, T., Abbriano, R. M., Chen, Y., Shah, S., Granados-Cifuentes, C., van Oppen, M. J. H., Bhattacharya, D., Suggett, D. J., Chan, C. X., & Rodriguez-Lanetty, M. (2022). Whole-genome duplication in an algal symbiont serendipitously confers thermal tolerance to corals. BioRxiv, 2022.04.10.487810. https://doi.org/10.1101/2022.04.10.487810

Dougan, K. E., González-Pech, R. A., Stephens, T. G., Shah, S., Chen, Y., Ragan, M. A., Bhattacharya, D., & Chan, C. X. (2022). Genome-powered classification of microbial eukaryotes: focus on coral algal symbionts. Trends in Microbiology. https://doi.org/10.1016/j.tim.2022.02.001

Drury, C., Dale, K. E., Panlilio, J. M., Miller, S. V., Lirman, D., Larson, E. A., Bartels, E., Crawford, D. L., & Oleksiak, M. F. (2016). Genomic variation among populations of threatened coral: *Acropora cervicornis*. BMC Genomics, 17(1), 286. https://doi.org/10.1186/s12864-016-2583-8

Dustan, P. (1996). Corals in Space and Time: The Biogeography and Evolution of the Scleractinia. J. E. N. Veron. The Quarterly Review of Biology, 71(2), 285–285. https://doi.org/10.1086/419409

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

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. Trends in Ecology & Evolution, 29(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Enbody, E. D., Pettersson, M. E., Sprehn, C. G., Palm, S., Wickström, H., & Andersson, L. (2021). Ecological adaptation in European eels is based on phenotypic plasticity. Proceedings of the National Academy of Sciences, 118(4), e2022620118. https://doi.org/10.1073/pnas.2022620118

Excofffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics, 37(24), btab468-. https://doi.org/10.1093/bioinformatics/btab468

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. PLoS Genetics, 9(10), e1003905. https://doi.org/10.1371/journal.pgen.1003905

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving Inversions. Trends in Ecology & Evolution, 34(3), 239–248. https://doi.org/10.1016/j.tree.2018.12.005

Farré, M., Micheletti, D., & Ruiz-Herrera, A. (2013). Recombination Rates and Genomic Shuffling in Human and Chimpanzee—A New Twist in the Chromosomal Speciation Theory. Molecular Biology and Evolution, 30(4), 853–864. https://doi.org/10.1093/molbev/mss272

Ferretti, L., Ramos‐Onsins, S. E., & Pérez‐Enciso, M. (2013). Population genomics from pool sequencing. Molecular Ecology, 22(22), 5561–5576. https://doi.org/10.1111/mec.12522
Fifer, J. E., Yasuda, N., Yamakita, T., Bove, C. B., & Davies, S. W. (2022). Genetic divergence and range expansion in a western North Pacific coral. Science of The Total Environment, 813, 152423. https://doi.org/10.1016/j.scitotenv.2021.152423
Fifer, J., Yasuda, N., Yamakita, T., & Davies, S. (2021). Demographic Inferences and Loci Under Selection in a Recently Expanded Coral Population. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3869113
Flot, J.-F., Dahl, M., & André, C. (2013). Lophelia pertusa corals from the Ionian and Barents seas share identical nuclear ITS2 and near-identical mitochondrial genome sequences. BMC Research Notes, 6(1), 144. https://doi.org/10.1186/1756-0500-6-144
Forêt, S., Wilson, S. R., & Burden, C. J. (2009). Empirical distribution of k-word matches in biological sequences. Pattern Recognition, 42(4), 539–548. https://doi.org/10.1016/j.patcog.2008.06.026
Forsman, R. B. (2005). Life and death on the coral reef: an ecological perspective on scholarly publishing in the health sciences. Journal of the Medical Library Association : JMLA, 93(1), 7–15.
Fox, E. A., Wright, A. E., Fumagalli, M., & Vieira, F. G. (2019). ngsLD: evaluating linkage disequilibrium using genotype likelihoods. Bioinformatics, 35(19), 3855–3856. https://doi.org/10.1093/bioinformatics/btz200
Fuentes‐Pardo, A. P., & Ruzzante, D. E. (2017). Whole‐genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. Molecular Ecology, 26(20), 5369–5406. https://doi.org/10.1111/mec.14264
Fuller, Z. L., Koury, S. A., Leonard, C. J., Young, R. E., Ikegami, K., Westlake, J., Richards, S., Schaeffer, S. W., & Phadnis, N. (2020). Extensive Recombination Suppression and Epistatic Selection Causes Chromosome-Wide Differentiation of a Selfish Sex Chromosome in Drosophila pseudoobscura. Genetics, 216(1), 205–226. https://doi.org/10.1534/genetics.120.303460
Fuller, Z. L., Mocellin, V. J. L., Morris, L. A., Cantin, N., Shepherd, J., Sarre, L., Peng, J., Liao, Y., Pickrell, J., Andolfatto, P., Matz, M., Bay, L. K., & Przeworski, M. (2020). Population genetics of the coral *Acropora millepora*: Toward genomic prediction of bleaching. Science (New York, N.Y.), 369(6501). https://doi.org/10.1126/science.aba4674
Fuller, Z. L., Mocellin, V. J. L., Morris, L., Cantin, N., Shepherd, J., Sarre, L., Peng, J., Liao, Y., Pickrell, J., Andolfatto, P. L., Matz, M. V., Bay, L. K., & Przeworski, M. (2019). Population genetics of the coral *Acropora millepora*: Towards a genomic predictor of bleaching. BioRxiv, 867754. https://doi.org/10.1101/867754
Fumagalli, M. (2013). Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. PLoS ONE, 8(11), e79667. https://doi.org/10.1371/journal.pone.0079667
Fumagalli, M., Vieira, F. G., Korneliussen, T. S., Linderoth, T., Huerta-Sánchez, E., Albrechtsen, A., & Nielsen, R. (2013). Quantifying Population Genetic Differentiation from Next-Generation Sequencing Data. Genetics, 195(3), 979–992. https://doi.org/10.1534/genetics.113.154740
Fumagalli, M., Vieira, F. G., Linderoth, T., & Nielsen, R. (2014). ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics, 30(10), 1486–1487. https://doi.org/10.1093/bioinformatics/btu041
Futschik, A., & Schlötterer, C. (2010). The Next Generation of Molecular Markers From Massively Parallel Sequencing of Pooled DNA Samples. Genetics, 186(1), 207–218. https://doi.org/10.1534/genetics.110.114397
Galinsky, K. J., Bhatia, G., Loh, P.-R., Georgiev, S., Mukherjee, S., Patterson, N. J., & Price, A. L. (2016). Fast Principal-Component Analysis Reveals Convergent Evolution of

ADH1B in Europe and East Asia. The American Journal of Human Genetics, 98(3), 456–472. https://doi.org/10.1016/j.ajhg.2015.12.022
Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J., & Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular Ecology, 22(11), 3165–3178. https://doi.org/10.1111/mec.12089
Gilmour, J. P., Smith, L. D., Heyward, A. J., Baird, A. H., & Pratchett, M. S. (2013). Recovery of an Isolated Coral Reef System Following Severe Disturbance. Science, 340(6128), 69–71. https://doi.org/10.1126/science.1232310
González-Pech, R. A., Stephens, T. G., Chen, Y., Mohamed, A. R., Cheng, Y., Shah, S., Dougan, K. E., Fortuin, M. D. A., Lagorce, R., Burt, D. W., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2021). Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus Symbiodinium. BMC Biology, 19(1), 73. https://doi.org/10.1186/s12915-021-00994-6
Goodbody‐Gringley, G., Woollacott, R. M., & Giribet, G. (2012). Population structure and connectivity in the Atlantic scleractinian coral Montastraea cavernosa (Linnaeus, 1767). Marine Ecology, 33(1), 32–48. https://doi.org/10.1111/j.1439-0485.2011.00452.x
Gouin, A., Legeai, F., Nouhaud, P., Whibley, A., Simon, J.-C., & Lemaitre, C. (2014). Whole-genome re-sequencing of non-model organisms: lessons from unmapped reads. Heredity, 114(5), 494–501. https://doi.org/10.1038/hdy.2014.85
Gourlé, H., Karlsson-Lindsjö, O., Hayer, J., & Bongcam-Rudloff, E. (2019). Simulating Illumina metagenomic data with InSilicoSeq. Bioinformatics, 35(3), 521–522. https://doi.org/10.1093/bioinformatics/bty630
Greenfield, P., & Roehm, U. (2013). Answering biological questions by querying k‐mer databases. Concurrency and Computation: Practice and Experience, 25(4), 497–509. https://doi.org/10.1002/cpe.2938
Grinblat, M., Cooke, I., Shlesinger, T., Ben-Zvi, O., Loya, Y., Miller, D. J., & Cowman, P. F. (2021). Biogeography, reproductive biology and phylogenetic divergence within the Fungiidae (mushroom corals). Molecular Phylogenetics and Evolution, 164, 107265. https://doi.org/10.1016/j.ympev.2021.107265
Grossman, S. R., Shlyakhter, I., Shylakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., Lander, E. S., Schaffner, S. F., & Sabeti, P. C. (2010). A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection. Science, 327(5967), 883–886. https://doi.org/10.1126/science.1183863
Guest, J. (2021). The Great Barrier Reef: How repeated marine heat waves are reshaping an iconic marine ecosystem. Current Biology, 31(23), R1530–R1532. https://doi.org/10.1016/j.cub.2021.10.066
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), e1000695. https://doi.org/10.1371/journal.pgen.1000695
Hancock‐Hanser, B. L., Frey, A., Leslie, M. S., Dutton, P. H., Archer, F. I., & Morin, P. A. (2013). Targeted multiplex next‐generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. Molecular Ecology Resources, 13(2), 254–268. https://doi.org/10.1111/1755-0998.12059
Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., Ruegg, K., & Palstra, F. (2011). Understanding and Estimating Effective Population Size for Practical Application in Marine Species Management. Conservation Biology, 25(3), 438–449. https://doi.org/10.1111/j.1523-1739.2010.01637.x
Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. Genome Research, 16(6), 730–737. https://doi.org/10.1101/gr.5045006

Harringmeyer, O. S., & Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. Nature Ecology & Evolution, 1–15. https://doi.org/10.1038/s41559-022-01890-0
Hayashibara, T., & Shimoike, K. (2002). Cryptic species of *Acropora digitifera*. Coral Reefs, 21(2), 224–225. https://doi.org/10.1007/s00338-002-0229-6
Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. Nature, 405(6789), 907–913. https://doi.org/10.1038/35016000
Hoadley, K. D., Pettay, Daniel. T., Lewis, A., Wham, D., Grasso, C., Smith, R., Kemp, D. W., LaJeunesse, T., & Warner, M. E. (2021). Different functional traits among closely related algal symbionts dictate stress endurance for vital Indo‐Pacific reef‐building corals. Global Change Biology, 27(20), 5295–5309. https://doi.org/10.1111/gcb.15799
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), e2003355. https://doi.org/10.1371/journal.pbio.2003355
Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. Nature Reviews Genetics, 10(9), 639–650. https://doi.org/10.1038/nrg2611
Hopley, D., Smithers, S. G., & Parnell, K. (2007). The Geomorphology of the Great Barrier Reef: Development, Diversity and Change. Cambridge University Press; Cambridge Core. https://doi.org/10.1017/cbo9780511535543
Howells, E. J., Abrego, D., Meyer, E., Kirk, N. L., & Burt, J. A. (2016). Host adaptation and unexpected symbiont partners enable reef‐building corals to tolerate extreme temperatures. Global Change Biology, 22(8), 2702‐2714. https://doi.org/10.1111/gcb.13250
Howells, E. J., Abrego, D., Liew, Y. J., Burt, J. A., Meyer, E., & Aranda, M. (2021). Enhancing the heat tolerance of reef-building corals to future warming. Science Advances, 7(34), eabg6070. https://doi.org/10.1126/sciadv.abg6070
Howells, E. J., Bay, L. K., & Bay, R. A. (2022). Coral Reef Conservation and Restoration in the Omics Age. Coral Reefs of the World, 55–70. https://doi.org/10.1007/978-3-031-07055-6_4
Howells, E. J., Beltran, V. H., Larsen, N. W., Bay, L. K., Willis, B. L., & van Oppen, M. J. H. (2012). Coral thermal tolerance shaped by local adaptation of photosymbionts. Nature Climate Change, 2(2), 116–120. https://doi.org/10.1038/nclimate1330
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
Howells, E. J., van Oppen, M. J. H., & Willis, B. L. (2009). High genetic differentiation and cross-shelf patterns of genetic diversity among Great Barrier Reef populations of Symbiodinium. Coral Reefs, 28(1), 215–225. https://doi.org/10.1007/s00338-008-0450-z
Howells, E. J., Willis, B. L., Bay, L. K., & van Oppen, M. J. H. (2013). Spatial and temporal genetic structure of Symbiodinium populations within a common reef‐building coral on the Great Barrier Reef. Molecular Ecology, 22(14), 3693‐3708. https://doi.org/10.1111/mec.12342
Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. Molecular Ecology, 29(14), 2535–2549. https://doi.org/10.1111/mec.15428
Huddleston, J., Chaisson, M. J. P., Steinberg, K. M., Warren, W., Hoekzema, K., Gordon, D., Graves-Lindsay, T. A., Munson, K. M., Kronenberg, Z. N., Vives, L., Peluso, P., Boitano, M., Chin, C.-S., Korlach, J., Wilson, R. K., & Eichler, E. E. (2017). Discovery and genotyping of structural variation from long-read haploid genome sequence data. Genome Research, 27(5), 677–685. https://doi.org/10.1101/gr.214007.116
Hudson, R. R., & Kaplan, N. L. (1995). Deleterious background selection with recombination. Genetics, 141(4), 1605–1617. https://doi.org/10.1093/genetics/141.4.1605

Hughes, L., Dean, A., Steffen, W., & Rice, M. (2018). Lethal consequences: climate change impacts on the Great Barrier Reef. Climate Council of Australia.
Hughes, T. P., Anderson, K. D., Connolly, S. R., Heron, S. F., Kerry, J. T., Lough, J. M., Baird, A. H., Baum, J. K., Berumen, M. L., Bridge, T. C., Claar, D. C., Eakin, C. M., Gilmour, J. P., Graham, N. A. J., Harrison, H., Hobbs, J.-P. A., Hoey, A. S., Hoogenboom, M., Lowe, R. J., … 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
Hume, B. C. C., D'Angelo, C., Smith, E. G., Stevens, J. R., Burt, J., & Wiedenmann, J. (2015). Symbiodinium thermophilum sp. nov., a thermotolerant symbiotic alga prevalent in corals of the world's hottest sea, the Persian/Arabian Gulf. Scientific Reports, 5(1), 8562. https://doi.org/10.1038/srep08562
Hume, B. C. C., Smith, E. G., Ziegler, M., Warrington, H. J. M., Burt, J. A., LaJeunesse, T. C., Wiedenmann, J., & Voolstra, C. R. (2019). SymPortal: A novel analytical framework and platform for coral algal symbiont next‐generation sequencing ITS2 profiling. Molecular Ecology Resources, 19(4), 1063–1080. https://doi.org/10.1111/1755-0998.13004
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
Jaspers, C., Fraune, S., Arnold, A. E., Miller, D. J., Bosch, T. C. G., Voolstra, C. R., & Participants, C. of A. A. of S. B. R. C. (2019). Resolving structure and function of metaorganisms through a holistic framework combining reductionist and integrative approaches. Zoology, 133, 81–87. https://doi.org/10.1016/j.zool.2019.02.007
Jiang, Y., Jiang, Y., Wang, S., Zhang, Q., & Ding, X. (2019). Optimal sequencing depth design for whole genome re-sequencing in pigs. BMC Bioinformatics, 20(1), 556. https://doi.org/10.1186/s12859-019-3164-z
Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. Molecular Ecology, 25(1), 185–202. https://doi.org/10.1111/mec.13304
Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017). Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. Genetics, 206(3), 1549–1567. https://doi.org/10.1534/genetics.117.200493
Kapun, M., Fabian, D. K., Goudet, J., & Flatt, T. (2016). Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. Molecular Biology and Evolution, 33(5), 1317–1336. https://doi.org/10.1093/molbev/msw016
Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010
Kirkpatrick, M., & Barton, N. (2006). Chromosome Inversions, Local Adaptation and Speciation. Genetics, 173(1), 419–434. https://doi.org/10.1534/genetics.105.047985
Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., Rosa, M. C. D., Righino, B., Johansen, T., Otterå, H., Sonesson, A., Lien, S., & Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. Molecular Ecology, 25(10), 2130–2143. https://doi.org/10.1111/mec.13592
Klassmann, A., & Gautier, M. (2022). Detecting selection using extended haplotype homozygosity (EHH)-based statistics in unphased or unpolarized data. PLoS ONE, 17(1), e0262024. https://doi.org/10.1371/journal.pone.0262024
Kofler, R., & Schlötterer, C. (2014). A Guide for the Design of Evolve and Resequencing Studies. Molecular Biology and Evolution, 31(2), 474–483. https://doi.org/10.1093/molbev/mst221
Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Korneliussen, T. S., & Moltke, I. (2015). NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics, 31(24), 4009–4011. https://doi.org/10.1093/bioinformatics/btv509
Korunes, K. L., & Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Molecular Ecology Resources, 21(4), 1359–1368. https://doi.org/10.1111/1755-0998.13326
Knowlton, N. (2001). The future of coral reefs. Proceedings of the National Academy of Sciences, 98(10), 5419–5425. https://doi.org/10.1073/pnas.091092998
Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P., & Kronforst, M. R. (2014). doublesex is a mimicry supergene. Nature, 507(7491), 229–232. https://doi.org/10.1038/nature13112
Küpper, C., Stocks, M., Risse, J. E., Remedios, N. dos, Farrell, L. L., McRae, S. B., Morgan, T. C., Karlionova, N., Pinchuk, P., Verkuil, Y. I., Kitaysky, A. S., Wingfield, J. C., Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D. B., & Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. Nature Genetics, 48(1), 79–83. https://doi.org/10.1038/ng.3443
Kurland, S., Wheat, C. W., Mancera, M. P. C., Kutschera, V. E., Hill, J., Andersson, A., Rubin, C., Andersson, L., Ryman, N., & Laikre, L. (2019). Exploring a Pool‐seq‐only approach for gaining population genomic insights in nonmodel species. Ecology and Evolution, 9(19), 11448–11463. https://doi.org/10.1002/ece3.5646
Ladner, J. T., & Palumbi, S. R. (2012). Extensive sympatry, cryptic diversity and introgression throughout the geographic distribution of two coral species complexes. Molecular Ecology, 21(9), 2224–2238. https://doi.org/10.1111/j.1365-294x.2012.05528.x
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, 28(16), 2570-2580.e6. https://doi.org/10.1016/j.cub.2018.07.008
Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics, 11(10), 733–739. https://doi.org/10.1038/nrg2825
Lefouili, M., & Nam, K. (2022). The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. Scientific Reports, 12(1), 11331. https://doi.org/10.1038/s41598-022-15563-2
Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature, 475(7357), 493–496. https://doi.org/10.1038/nature10231
Li, H., & Ralph, P. (2018). 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, Y., Sidore, C., Kang, H. M., Boehnke, M., & Abecasis, G. R. (2011). Low-coverage sequencing: Implications for design of complex trait association studies. Genome Research, 21(6), 940–951. https://doi.org/10.1101/gr.117259.110
Lien, Y.-T., Nakano, Y., Plathong, S., Fukami, H., Wang, J.-T., & Chen, C. A. (2007). Occurrence of the putatively heat-tolerant Symbiodinium phylotype D in high-latitudinal outlying coral communities. Coral Reefs, 26(1), 35–44. https://doi.org/10.1007/s00338-006-0185-7
Lindsay, C., Mick, O., Alexandra, S., G., B., M., K., & T., S. (2015). Geomorphic patterns, internal architecture and reef growth in a macrotidal, high-turbidity setting of coral reefs from the Kimberley bioregion. Australian Journal of Maritime & Ocean Affairs, 7. https://doi.org/10.1080/18366503.2015.1021411
Lippert, R. A., Huang, H., & Waterman, M. S. (2002). Distributional regimes for the number of k-word matches between two random sequences. Proceedings of the National Academy of Sciences, 99(22), 13980–13989. https://doi.org/10.1073/pnas.202468099

Liu, H., Stephens, T. G., González-Pech, R. A., Beltran, V. H., Lapeyre, B., Bongaerts, P., Cooke, I., Aranda, M., Bourne, D. G., Forêt, S., Miller, D. J., van Oppen, M. J. H., Voolstra, C. R., Ragan, M. A., & Chan, C. X. (2018). Symbiodinium genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. Communications Biology, 1(1), 95. https://doi.org/10.1038/s42003-018-0098-3

Lo, R., Dougan, K. E., Chen, Y., Shah, S., Bhattacharya, D., & Chan, C. X. (2022). Alignment-Free Analysis of Whole-Genome Sequences From Symbiodiniaceae Reveals Different Phylogenetic Signals in Distinct Regions. Frontiers in Plant Science, 13, 815714. https://doi.org/10.3389/fpls.2022.815714

Lotterhos, K. E. (2019). The Effect of Neutral Recombination Variation on Genome Scans for Selection. G3: Genes|Genomes|Genetics, 9(6), 1851–1867. https://doi.org/10.1534/g3.119.400088

Lotterhos, K. E., Card, D. C., Schaal, S. M., Wang, L., Collins, C., & Verity, B. (2017). Composite measures of selection can improve the signal‐to‐noise ratio in genome scans. Methods in Ecology and Evolution, 8(6), 717–727. https://doi.org/10.1111/2041-210x.12774

Lou, R. N., Jacobs, A., Wilder, A., & Therkildsen, N. O. (2021). A beginner's guide to low‐coverage whole genome sequencing for population genomics. Molecular Ecology. https://doi.org/10.1111/mec.16077

Lou, R. N., & Therkildsen, N. O. (2022). Batch effects in population genomic studies with low‐coverage whole genome sequencing data: Causes, detection and mitigation. Molecular Ecology Resources, 22(5), 1678–1692. https://doi.org/10.1111/1755-0998.13559

Luczak, B. B., James, B. T., & Girgis, H. Z. (2017). A survey and evaluations of histogram-based statistics in alignment-free sequence comparison. Briefings in Bioinformatics, 20(4), 1222–1237. https://doi.org/10.1093/bib/bbx161

Ludt, W. B., & Rocha, L. A. (2015). Shifting seas: the impacts of Pleistocene sea‐level fluctuations on the evolution of tropical marine taxa. Journal of Biogeography, 42(1), 25–38. https://doi.org/10.1111/jbi.12416

Luikart, G., Kardos, M., Hand, B. K., Rajora, O. P., Aitken, S. N., & Hohenlohe, P. A. (2018). Population Genomics, Concepts, Approaches and Applications. Population Genomics, 3–79. https://doi.org/10.1007/13836_2018_60

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

Luo, X., Xu, L., Wang, Y., Dong, J., Chen, Y., Tang, M., Fan, L., Zhu, Y., & Liu, L. (2020). An ultra‐high‐density genetic map provides insights into genome synteny, recombination landscape and taproot skin colour in radish (Raphanus sativus L.). Plant Biotechnology Journal, 18(1), 274–286. https://doi.org/10.1111/pbi.13195

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., & Simianer, H. (2015). Properties of different selection signature statistics and a new strategy for combining them. Heredity, 115(5), 426–436. https://doi.org/10.1038/hdy.2015.42

MacNeil, M. A., Mellin, C., Matthews, S., Wolff, N. H., McClanahan, T. R., Devlin, M., Drovandi, C., Mengersen, K., & Graham, N. A. J. (2019). Water quality mediates resilience on the Great Barrier Reef. Nature Ecology & Evolution, 3(4), 620–627. https://doi.org/10.1038/s41559-019-0832-3

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science, 350(6267), 1493–1498. https://doi.org/10.1126/science.aac9927

Manthey, J. D., Campillo, L. C., Burns, K. J., & Moyle, R. G. (2016). Comparison of Target-Capture and Restriction-Site Associated DNA Sequencing for Phylogenomics: A Test in Cardinalid Tanagers (Aves, Genus: Piranga). Systematic Biology, 65(4), 640–650. https://doi.org/10.1093/sysbio/syw005

Manzello, D. P., Matz, M. V., Enochs, I. C., Valentino, L., Carlton, R. D., Kolodziej, G., Serrano, X., Towle, E. K., & Jankulak, M. (2018). Role of host genetics and heat tolerant algal symbionts in sustaining populations of the endangered coral Orbicella faveolata in the Florida Keys with ocean warming. Global Change Biology, 25(3), 1016–1031. https://doi.org/10.1111/gcb.14545

Mao, Y., Economo, E. P., & Satoh, N. (2018). The Roles of Introgression and Climate Change in the Rise to Dominance of Acropora Corals. Current Biology, 28(21), 3373-3382.e5. https://doi.org/10.1016/j.cub.2018.08.061

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics, 27(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Marr, T., Torney, D. C., Burks, C., Davison, D., & Sirotkin, K. M. (2018). Computers and DNA. 109–125. https://doi.org/10.4324/9780429501463-11

Matias, A. M. A., Popovic, I., Thia, J. A., Cooke, I. R., Torda, G., Lukoschek, V., Bay, L. K., Kim, S. W., & Riginos, C. (2022). Cryptic diversity and spatial genetic variation in the coral *Acropora tenuis* and its endosymbionts across the Great Barrier Reef. Evolutionary Applications. https://doi.org/10.1111/eva.13435

Matsuda, S. B., Chakravarti, L. J., Cunning, R., Huffmyer, A. S., Nelson, C. E., Gates, R. D., & van Oppen, M. J. H. (2022). Temperature‐mediated acquisition of rare heterologous symbionts promotes survival of coral larvae under ocean warming. Global Change Biology. https://doi.org/10.1111/gcb.16057

Mattle-Greminger, M. P., Sonay, T. B., Nater, A., Pybus, M., Desai, T., Valles, G. de, Casals, F., Scally, A., Bertranpetit, J., Marques-Bonet, T., Schaik, C. P. van, Anisimova, M., & Krützen, M. (2018). Genomes reveal marked differences in the adaptive evolution between orangutan species. Genome Biology, 19(1), 193. https://doi.org/10.1186/s13059-018-1562-6

Matute, D. R. (2013). The role of founder effects on the evolution of reproductive isolation. Journal of Evolutionary Biology, 26(11), 2299–2311. https://doi.org/10.1111/jeb.12246

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. A., & Haller, B. C. (2020). Estimating the potential for coral adaptation to global warming across the Indo‐West Pacific. Global Change Biology, 26(6), 3473‐3481. https://doi.org/10.1111/gcb.15060

McCartney‐Melstad, E., Mount, G. G., & Shaffer, H. B. (2016). Exon capture optimization in amphibians with large genomes. Molecular Ecology Resources, 16(5), 1084‐1094. https://doi.org/10.1111/1755-0998.12538

McCormick, R. F., Truong, S. K., & Mullet, J. E. (2015). RIG: Recalibration and Interrelation of Genomic Sequence Data with the GATK. G3: Genes|Genomes|Genetics, 5(4), 655–665. https://doi.org/10.1534/g3.115.017012

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & 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

McKinney, G. J., Seeb, L. W., Larson, W. A., Gomez‐Uchida, D., Limborg, M. T., Brieuc, M. S. O., Everett, M. V., Naish, K. A., Waples, R. K., & Seeb, J. E. (2016). An integrated linkage map reveals candidate genes underlying adaptive variation in Chinook salmon (Oncorhynchus tshawytscha). Molecular Ecology Resources, 16(3), 769–783. https://doi.org/10.1111/1755-0998.12479

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. Molecular Ecology Resources, 11(1), 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x
Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics, 210(2), 719–731. https://doi.org/10.1534/genetics.118.301336
Meisner, J., Albrechtsen, A., & Hanghøj, K. (2021). Detecting selection in low-coverage high-throughput sequencing data using principal component analysis. BMC Bioinformatics, 22(1), 470. https://doi.org/10.1186/s12859-021-04375-2
Melo, A. T. O., & Hale, I. (2018). Expanded functionality, increased accuracy, and enhanced speed in the de novo genotyping-by-sequencing pipeline GBS-SNP-CROP. Bioinformatics, 35(10), 1783–1785. https://doi.org/10.1093/bioinformatics/bty873
Mérot, C. (2020). Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. Molecular Ecology, 29(14), 2513–2516. https://doi.org/10.1111/mec.15500
Mérot, C., Berdan, E., Cayuela, H., Djambazian, H., Ferchaud, A.-L., Laporte, M., Normandeau, E., Ragoussis, J., Wellenreuther, M., & Bernatchez, L. (2021). Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. Molecular Biology and Evolution, 38(9), msab143-. https://doi.org/10.1093/molbev/msab143
Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. Trends in Ecology & Evolution, 35(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002
Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. Bioinformatics, 24(16), 1757–1764. https://doi.org/10.1093/bioinformatics/btn322
Morikawa, M. K., & Palumbi, S. R. (2019). Using naturally occurring climate resilient corals to construct bleaching-resistant nurseries. Proceedings of the National Academy of Sciences, 116(21), 201721415. https://doi.org/10.1073/pnas.1721415116
Mote, S., Gupta, V., De, K., Hussain, A., More, K., Nanajkar, M., & Ingole, B. (2021). Differential Symbiodiniaceae Association With Coral and Coral-Eroding Sponge in a Bleaching Impacted Marginal Coral Reef Environment. Frontiers in Marine Science, 8, 666825. https://doi.org/10.3389/fmars.2021.666825
Moya, A., Galiana, A., & Ayala, F. J. (1995). Founder-effect speciation theory: failure of experimental corroboration. Proceedings of the National Academy of Sciences, 92(9), 3983–3986. https://doi.org/10.1073/pnas.92.9.3983
Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., ffrench-Constant, R. H., Blaxter, M. L., Mallet, J., & Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1587), 343–353. https://doi.org/10.1098/rstb.2011.0198
Nakajima, Y., Wepfer, P. H., Suzuki, S., Zayasu, Y., Shinzato, C., Satoh, N., & Mitarai, S. (2017). Microsatellite markers for multiple Pocillopora genetic lineages offer new insights about coral populations. Scientific Reports, 7(1), 6729. https://doi.org/10.1038/s41598-017-06776-x
Nand, A., Zhan, Y., Salazar, O. R., Aranda, M., Voolstra, C. R., & Dekker, J. (2021). Genetic and spatial organization of the unusual chromosomes of the dinoflagellate *Symbiodinium microadriaticum*. Nature Genetics, 53(5), 618–629. https://doi.org/10.1038/s41588-021-00841-y
Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von, & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution, 32(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics, 12(6), 443–451. https://doi.org/10.1038/nrg2986

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., … Phillippy, A. M. (2022). The complete sequence of a human genome. Science, 376(6588), 44–53. https://doi.org/10.1126/science.abj6987

Nyström, M., Folke, C., Moberg, F., Nyström, M., Folke, C., & Moberg, F. (2000). Coral reef disturbance and resilience in a human-dominated environment. Trends in Ecology & Evolution, 15(10), 413–417. https://doi.org/10.1016/s0169-5347(00)01948-0

Oliver, T. A., & Palumbi, S. R. (2011). Many corals host thermally resistant symbionts in high-temperature habitat. Coral Reefs, 30(1), 241–250. https://doi.org/10.1007/s00338-010-0696-0

van Oppen, M. J. H., Bongaerts, P., Frade, P., Peplow, L. M., Boyd, S. E., Nim, H. T., & Bay, L. K. (2018). Adaptation to reef habitats through selection on the coral animal and its associated microbiome. Molecular Ecology, 27(14), 2956–2971. https://doi.org/10.1111/mec.14763

van Oppen, M. J. H., & Blackall, L. L. (2019). Coral microbiome dynamics, functions and design in a changing world. Nature Reviews Microbiology, 17(9), 557–567. https://doi.org/10.1038/s41579-019-0223-4

van Oppen, M. J. H., McDonald, B. J., Willis, B., & Miller, D. J. (2001). The Evolutionary History of the Coral Genus Acropora (Scleractinia, Cnidaria) Based on a Mitochondrial and a Nuclear Marker: Reticulation, Incomplete Lineage Sorting, or Morphological Convergence? Molecular Biology and Evolution, 18(7), 1315–1329. https://doi.org/10.1093/oxfordjournals.molbev.a003916

van Oppen, M. J. H., Palstra, F. P., Piquet, A. M.-T., & Miller, D. J. (2001). Patterns of coraldinoflagellate associations in Acropora: significance of local availability and physiology of Symbiodinium strains and hostsymbiont selectivity. Proceedings of the Royal Society of London. Series B: Biological Sciences, 268(1478), 1759–1767. https://doi.org/10.1098/rspb.2001.1733

Ortiz, J.-C., Wolff, N. H., Anthony, K. R. N., Devlin, M., Lewis, S., & Mumby, P. J. (2018). Impaired recovery of the Great Barrier Reef under cumulative stress. Science Advances, 4(7), eaar6127. https://doi.org/10.1126/sciadv.aar6127

Osman, E. O., Suggett, D. J., Voolstra, C. R., Pettay, D. T., Clark, D. R., Pogoreutz, C., Sampayo, E. M., Warner, M. E., & Smith, D. J. (2020). Coral microbiome composition along the northern Red Sea suggests high plasticity of bacterial and specificity of endosymbiotic dinoflagellate communities. Microbiome, 8(1), 8. https://doi.org/10.1186/s40168-019-0776-5

Ottenburghs, J., Honka, J., Heikkinen, M. E., Madsen, J., Müskens, G. J. D. M., & Ellegren, H. (2023). Highly differentiated loci resolve phylogenetic relationships in the Bean Goose complex. BMC Ecology and Evolution, 23(1), 2. https://doi.org/10.1186/s12862-023-02103-3

Palumbi, S. R. (2003). POPULATION GENETICS, DEMOGRAPHIC CONNECTIVITY, AND THE DESIGN OF MARINE RESERVES. Ecological Applications, 13(sp1), 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), 895–898. https://doi.org/10.1126/science.1251336

Palumbi, S. R., Walker, N. S., Hanson, E., Armstrong, K., Lippert, M., Cornwell, B., Nestor, V., & Golbuu, Y. (2023). Small‐scale genetic structure of coral populations in Palau based on whole mitochondrial genomes: Implications for future coral resilience. Evolutionary Applications. https://doi.org/10.1111/eva.13509

Pandolfi, J., & Kelley, R. (2011). The Great Barrier Reef in Time and Space: Geology and Palaeobiology.

Paradis, E., & Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633
Pellicer, J., Fay, M. F., & Leitch, I. J. (2010). The largest eukaryotic genome of them all? Botanical Journal of the Linnean Society, 164(1), 10–15. https://doi.org/10.1111/j.1095-8339.2010.01072.x
Peng, Y., Yan, H., Guo, L., Deng, C., Wang, C., Wang, Y., Kang, L., Zhou, P., Yu, K., Dong, X., Liu, X., Sun, Z., Peng, Y., Zhao, J., Deng, D., Xu, Y., Li, Y., Jiang, Q., Li, Y., … Ren, C. (2022). Reference genome assemblies reveal the origin and evolution of allohexaploid oat. Nature Genetics, 54(8), 1248–1258. https://doi.org/10.1038/s41588-022-01127-7
Pennisi, E. (2014). Disputed islands. Science, 345(6197), 611–613. https://doi.org/10.1126/science.345.6197.611
Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE, 7(5), e37135. https://doi.org/10.1371/journal.pone.0037135
Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics, 48(1), 94–100. https://doi.org/10.1038/ng.3464
Pinsky, M. L., Clark, R. D., & Bos, J. T. (2023). Coral Reef Population Genomics in an Age of Global Change. *Annual Review of Genetics*, *57*(1). https://doi.org/10.1146/annurev-genet-022123-102748
Pockrandt, C., Alzamel, M., Iliopoulos, C. S., & Reinert, K. (2020). GenMap: Ultra-fast Computation of Genome Mappability. Bioinformatics, 36(12), 3687–3692. https://doi.org/10.1093/bioinformatics/btaa222
Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M. G., & Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science, 344(6190), 1410–1414. https://doi.org/10.1126/science.1253226
Pokrovac, I., & Pezer, Ž. (2022). Recent advances and current challenges in population genomics of structural variation in animals and plants. Frontiers in Genetics, 13, 1060898. https://doi.org/10.3389/fgene.2022.1060898
Popovic, I., Matias, A. M. A., Bierne, N., & Riginos, C. (2020). Twin introductions by independent invader mussel lineages are both associated with recent admixture with a native congener in Australia. Evolutionary Applications, 13(3), 515–532. https://doi.org/10.1111/eva.12857
Poquita-Du, R. C., Huang, D., Chou, L. M., Mrinalini, & Todd, P. A. (2019). Short Term Exposure to Heat and Sediment Triggers Changes in Coral Gene Expression and Photo-Physiological Performance. Frontiers in Marine Science, 06, 121. https://doi.org/10.3389/fmars.2019.00121
Prada, C., Hanna, B., Budd, A. F., Woodley, C. M., Schmutz, J., Grimwood, J., Iglesias-Prieto, R., Pandolfi, J. M., Levitan, D., Johnson, K. G., Knowlton, N., Kitano, H., DeGiorgio, M., & Medina, M. (2016). Empty Niches after Extinctions Increase Population Sizes of Modern Corals. Current Biology, 26(23), 3190–3194. https://doi.org/10.1016/j.cub.2016.09.039
Prada, C., & Hellberg, M. E. (2021). Speciation ‐ by ‐ depth on coral reefs: Sympatric divergence with gene flow or cryptic transient isolation? Journal of Evolutionary Biology, 34(1), 128 ‐ 137. https://doi.org/10.1111/jeb.13731
Purcell, S. (2002). Intertidal reefs under extreme tidal flux in Buccaneer Archipelago, Western Australia. Coral Reefs, 21(2), 191–192. https://doi.org/10.1007/s00338-002-0223-z

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. Molecular Ecology, 23(24), 5937–5942. https://doi.org/10.1111/mec.12965

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D. M., Auvil, L., Capitanu, B., Ma, J., Lewin, H. A., Qian, X., Lang, Y., Zhou, R., Wang, L., Wang, K., … Liu, J. (2012). The yak genome and adaptation to life at high altitude. Nature Genetics, 44(8), 946–949. https://doi.org/10.1038/ng.2343

Quattrini, A. M., Rodríguez, E., Faircloth, B. C., Cowman, P. F., Brugler, M. R., Farfan, G. A., Hellberg, M. E., Kitahara, M. V., Morrison, C. L., Paz-García, D. A., Reimer, J. D., & McFadden, C. S. (2020). Palaeoclimate ocean conditions shaped the evolution of corals and their skeletons through deep time. Nature Ecology & Evolution, 4(11), 1531–1538. https://doi.org/10.1038/s41559-020-01291-1

Quigley, K. M., Bay, L. K., & van Oppen, M. J. H. (2019). The active spread of adaptive variation for reef resilience. Ecology and Evolution, 9(19), 11122–11135. https://doi.org/10.1002/ece3.5616

Quigley, K. M., Bay, L. K., & van Oppen, M. J. H. (2020). Genome‐wide SNP analysis reveals an increase in adaptive genetic variation through selective breeding of coral. Molecular Ecology, 29(12), 2176‐2188. https://doi.org/10.1111/mec.15482

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rädecker, N., Pogoreutz, C., Gegner, H. M., Cárdenas, A., Roth, F., Bougoure, J., Guagliardo, P., Wild, C., Pernice, M., Raina, J.-B., Meibom, A., & Voolstra, C. R. (2021). Heat stress destabilizes symbiotic nutrient cycling in corals. Proceedings of the National Academy of Sciences, 118(5), e2022653118. https://doi.org/10.1073/pnas.2022653118

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. Nature, 461(7263), 489–494. https://doi.org/10.1038/nature08365

Reinert, G., Chew, D., Sun, F., & Waterman, M. S. (2009). Alignment-Free Sequence Comparison (I): Statistics and Power. Journal of Computational Biology, 16(12), 1615–1634. https://doi.org/10.1089/cmb.2009.0198

Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q., & Shank, T. M. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. Molecular Ecology, 22(11), 2953–2970. https://doi.org/10.1111/mec.12228

Renema, W., Pandolfi, J. M., Kiessling, W., Bosellini, F. R., Klaus, J. S., Korpanty, C., Rosen, B. R., Santodomingo, N., Wallace, C. C., Webster, J. M., & Johnson, K. G. (2016). Are coral reefs victims of their own past success? Science Advances, 2(4), e1500850. https://doi.org/10.1126/sciadv.1500850

Richards, Z. T., Garcia, R. A., Wallace, C. C., Rosser, N. L., & Muir, P. R. (2015). A Diverse Assemblage of Reef Corals Thriving in a Dynamic Intertidal Reef Setting (Bonaparte Archipelago, Kimberley, Australia). PLOS ONE, 10(2), e0117791. https://doi.org/10.1371/journal.pone.0117791

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(6), 581–601. https://doi.org/10.1093/cz/zow067

Riginos, C., Hock, K., Matias, A. M., Mumby, P. J., van Oppen, M. J. H., & Lukoschek, V. (2019). Asymmetric dispersal is a critical element of concordance between biophysical dispersal models and spatial genetic structure in Great Barrier Reef corals. Diversity and Distributions, 25(11), 1684–1696. https://doi.org/10.1111/ddi.12969

Rinkevich, B. (2008). Management of coral reefs: We have gone wrong when neglecting active reef restoration. Marine Pollution Bulletin, 56(11), 1821–1824. https://doi.org/10.1016/j.marpolbul.2008.08.014

Rippe, J. P., Dixon, G., Fuller, Z. L., Liao, Y., & Matz, M. (2021). Environmental specialization and cryptic genetic divergence in two massive coral species from the Florida Keys Reef Tract. Molecular Ecology, 30(14), 3468–3484. https://doi.org/10.1111/mec.15931
Rippe, J. P., Matz, M. V., Green, E. A., Medina, M., Khawaja, N. Z., Pongwarin, T., C., J. H. P., Castillo, K. D., & Davies, S. W. (2017). Population structure and connectivity of the mountainous star coral, Orbicella faveolata, throughout the wider Caribbean region. Ecology and Evolution, 7(22), 9234–9246. https://doi.org/10.1002/ece3.3448
Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols, 12(12), 2640–2659. https://doi.org/10.1038/nprot.2017.123
Rocker, M. M., Francis, D. S., Fabricius, K. E., Willis, B. L., & Bay, L. K. (2017). Variation in the health and biochemical condition of the coral *Acropora tenuis* along two water quality gradients on the Great Barrier Reef, Australia. Marine Pollution Bulletin, 119(2), 106–119. https://doi.org/10.1016/j.marpolbul.2017.03.066
Rosenberg, E., & Zilber-Rosenberg, I. (2018). The hologenome concept of evolution after 10 years. Microbiome, 6(1), 78. https://doi.org/10.1186/s40168-018-0457-9
Rosser, N. L. (2015). Asynchronous spawning in sympatric populations of a hard coral reveals cryptic species and ancient genetic lineages. Molecular Ecology, 24(19), 5006–5019. https://doi.org/10.1111/mec.13372
Rosser, N. L., & Veron, J. E. N. (2011). Australian corals thriving out of water in an extreme environment. Coral Reefs, 30(1), 21–21. https://doi.org/10.1007/s00338-010-0689-z
Salojärvi, J., Smolander, O.-P., Nieminen, K., Rajaraman, S., Safronov, O., Safdari, P., Lamminmäki, A., Immanen, J., Lan, T., Tanskanen, J., Rastas, P., Amiryousefi, A., Jayaprakash, B., Kammonen, J. I., Hagqvist, R., Eswaran, G., Ahonen, V. H., Serra, J. A., Asiegbu, F. O., … Kangasjärvi, J. (2017). Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. Nature Genetics, 49(6), 904–912. https://doi.org/10.1038/ng.3862
Sampayo, E. M., Ridgway, T., Bongaerts, P., & Hoegh-Guldberg, O. (2008). Bleaching susceptibility and mortality of corals are determined by fine-scale differences in symbiont type. Proceedings of the National Academy of Sciences, 105(30), 10444–10449. https://doi.org/10.1073/pnas.0708049105
Santodomingo, N., Wallace, C. C., & Johnson, K. G. (2015). Fossils reveal a high diversity of the staghorn coral genera Acropora and Isopora (Scleractinia: Acroporidae) in the Neogene of Indonesia. Zoological Journal of the Linnean Society, 175(4), 677–763. https://doi.org/10.1111/zoj.12295
Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nature Genetics, 46(8), 919–925. https://doi.org/10.1038/ng.3015
Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nature Reviews Genetics, 15(11), 749–763. https://doi.org/10.1038/nrg3803
Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K. M., Auger, K., Chow, W., Collins, J., … Church, D. M. (2017). Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Research, 27(5), 849–864. https://doi.org/10.1101/gr.213611.116
Schoepf, V., Carrion, S. A., Pfeifer, S. M., Naugle, M., Dugal, L., Bruyn, J., & McCulloch, M. T. (2019). Stress-resistant corals may not acclimatize to ocean warming but maintain heat tolerance under cooler temperatures. Nature Communications, 10(1), 4031. https://doi.org/10.1038/s41467-019-12065-0
Schoepf, V., Stat, M., Falter, J. L., & McCulloch, M. T. (2015). Limits to the thermal tolerance of corals adapted to a highly fluctuating, naturally extreme temperature environment. Scientific Reports, 5(1), 17639. https://doi.org/10.1038/srep17639

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. https://doi.org/10.1111/eva.12944

Shepp, L. (1964). Normal Functions of Normal Random Variables. SIAM Review, 6(4), 459–460. https://doi.org/10.1137/1006100

Shinzato, C., Khalturin, K., Inoue, J., Zayasu, Y., Kanda, M., Kawamitsu, M., Yoshioka, Y., Yamashita, H., Suzuki, G., & Satoh, N. (2020). Eighteen Coral Genomes Reveal the Evolutionary Origin of Acropora Strategies to Accommodate Environmental Changes. Molecular Biology and Evolution, 38(1), msaa216. https://doi.org/10.1093/molbev/msaa216

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., Fujie, M., Fujiwara, M., Koyanagi, R., Ikuta, T., Fujiyama, A., Miller, D. J., & 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., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H., Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., … 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

Silverstein, R. N., Cunning, R., & Baker, A. C. (2015). Change in algal symbiont communities after bleaching, not prior heat exposure, increases heat tolerance of reef corals. Global Change Biology, 21(1), 236–249. https://doi.org/10.1111/gcb.12706

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., Bai, Z., Lorenzo, F. R., Xing, J., Jorde, L. B., Prchal, J. T., & Ge, R. (2010). Genetic Evidence for High-Altitude Adaptation in Tibet. Science, 329(5987), 72–75. https://doi.org/10.1126/science.1189406

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., Strand, A. I., Li, Q., Raney, B., Balakrishnan, C. N., Griffith, S. C., McVean, G., & Przeworski, M. (2015). Stable recombination hotspots in birds. Science, 350(6263), 928–932. https://doi.org/10.1126/science.aad0843

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating Individual Admixture Proportions from Next Generation Sequencing Data. Genetics, 195(3), 693–702. https://doi.org/10.1534/genetics.113.154138

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. Systematic Biology, 63(1), 83–95. https://doi.org/10.1093/sysbio/syt061

Smith, E. G., Hazzouri, K. M., Choi, J. Y., Delaney, P., Al-Kharafi, M., Howells, E. J., Aranda, M., & Burt, J. A. (2022). Signatures of selection underpinning rapid coral adaptation to the world's warmest reefs. Science Advances, 8(2), eabl7287. https://doi.org/10.1126/sciadv.abl7287

Smith-Keune, C., & van Oppen, M. (2006). Genetic structure of a reef-building coral from thermally distinct environments on the Great Barrier Reef. Coral Reefs, 25(3), 493–502. https://doi.org/10.1007/s00338-006-0129-2

Solihuddin, T., Collins, Lindsay. B., Blakeway, D., & Leary, M. J. O. (2015). Holocene coral reef growth and sea level in a macrotidal, high turbidity setting: Cockatoo Island, Kimberley Bioregion, northwest Australia. Marine Geology, 359, 50–60. https://doi.org/10.1016/j.margeo.2014.11.011

Solihuddin, T., O'Leary, M. J., Blakeway, D., Parnum, I., Kordi, M., & Collins, L. B. (2016). Holocene reef evolution in a macrotidal setting: Buccaneer Archipelago, Kimberley Bioregion, Northwest Australia. Coral Reefs, 35(3), 783–794. https://doi.org/10.1007/s00338-016-1424-1

Song, K., Ren, J., Reinert, G., Deng, M., Waterman, M. S., & Sun, F. (2014). New developments of alignment-free sequence comparison: measures, statistics and next-generation sequencing. Briefings in Bioinformatics, 15(3), 343–353. https://doi.org/10.1093/bib/bbt067

Souza, J. N. de, Nunes, F. L. D., Zilberberg, C., Sanchez, J. A., Migotto, A. E., Hoeksema, B. W., Serrano, X. M., Baker, A. C., & Lindner, A. (2017). Contrasting patterns of connectivity among endemic and widespread fire coral species (Millepora spp.) in the tropical Southwestern Atlantic. Coral Reefs, 36(3), 701–716. https://doi.org/10.1007/s00338-017-1562-0

Spalding, M. D., & Brown, B. E. (2015). Warm-water coral reefs and climate change. Science, 350(6262), 769–771. https://doi.org/10.1126/science.aad0349

Stat, M., Bird, C. E., Pochon, X., Chasqui, L., Chauka, L. J., Concepcion, G. T., Logan, D., Takabayashi, M., Toonen, R. J., & Gates, R. D. (2011). Variation in Symbiodinium ITS2 Sequence Assemblages among Coral Colonies. PLoS ONE, 6(1), e15854. https://doi.org/10.1371/journal.pone.0015854

Stat, M., Morris, E., & Gates, R. D. (2008). Functional diversity in coral–dinoflagellate symbiosis. Proceedings of the National Academy of Sciences, 105(27), 9256–9261. https://doi.org/10.1073/pnas.0801328105

Stolarski, J., Kitahara, M. V., Miller, D. J., Cairns, S. D., Mazur, M., & Meibom, A. (2011). The ancient evolutionary origins of Scleractinia revealed by azooxanthellate corals. BMC Evolutionary Biology, 11(1), 316. https://doi.org/10.1186/1471-2148-11-316

Sturm, A. B., Eckert, R. J., Méndez, J. G., González-Díaz, P., & Voss, J. D. (2020). Population genetic structure of the great star coral, Montastraea cavernosa, across the Cuban archipelago with comparisons between microsatellite and SNP markers. Scientific Reports, 10(1), 15432. https://doi.org/10.1038/s41598-020-72112-5

Szarmach, S. J., Brelsford, A., Witt, C. C., & Toews, D. P. L. (2021). Comparing divergence landscapes from reduced‐representation and whole genome resequencing in the yellow‐rumped warbler (Setophaga coronata) species complex. Molecular Ecology, 30(23), 5994–6005. https://doi.org/10.1111/mec.15940

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics, 31(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Templeton, A. R. (2008). The reality and importance of founder speciation in evolution. BioEssays, 30(5), 470–479. https://doi.org/10.1002/bies.20745

Terhorst, J., Kamm, J. A., & Song, Y. S. (2016). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nature Genetics, 49(2), 303–309. https://doi.org/10.1038/ng.3748

Teske, P., Sandoval-Castillo, J., Sebille, E. van, Waters, J., & Beheregaray, L. (2015). On-shelf larval retention limits population connectivity in a coastal broadcast spawner. Marine Ecology Progress Series, 532, 1–12. https://doi.org/10.3354/meps11362

Therkildsen, N. O., & Palumbi, S. R. (2017). Practical low‐coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Molecular Ecology Resources, 17(2), 194‐208. https://doi.org/10.1111/1755-0998.12593

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., Kennington, W. J., Stat, M., Wilkinson, S. P., Kool, J. T., & Kendrick, G. A. (2015). Isolation by resistance across a complex coral reef seascape. Proceedings of the Royal Society B: Biological Sciences, 282(1812), 20151217. https://doi.org/10.1098/rspb.2015.1217

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

Thomas, L., Underwood, J. N., Adam, A. A. S., Richards, Z. T., Dugal, L., Miller, K. J., & Gilmour, J. P. (2020). Contrasting patterns of genetic connectivity in brooding and spawning corals across a remote atoll system in northwest Australia. Coral Reefs, 39(1), 55–60. https://doi.org/10.1007/s00338-019-01884-8

Thomas, L., Underwood, J. N., Rose, N. H., Fuller, Z. L., Richards, Z. T., Dugal, L., Grimaldi, C. M., Cooke, I. R., Palumbi, S. R., & Gilmour, J. P. (2022). Spatially varying selection between habitats drives physiological shifts and local adaptation in a broadcast spawning coral on a remote atoll in Western Australia. Science Advances, 8(17), eabl9185. https://doi.org/10.1126/sciadv.abl9185

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., … Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nature Communications, 5(1), 5770. https://doi.org/10.1038/ncomms6770

Titus, B. M., & Daly, M. (2022). Population genomics for symbiotic anthozoans: can reduced representation approaches be used for taxa without reference genomes? Heredity, 128(5), 338–351. https://doi.org/10.1038/s41437-022-00531-3

Todd, E. V., Black, M. A., & Gemmell, N. J. (2016). The power and promise of RNA‐seq in ecology and evolution. Molecular Ecology, 25(6), 1224‐1241. https://doi.org/10.1111/mec.13526

Tonk, L., Sampayo, E. M., Weeks, S., Magno-Canto, M., & Hoegh-Guldberg, O. (2013). Host-Specific Interactions with Environmental Factors Shape the Distribution of Symbiodinium across the Great Barrier Reef. PLoS ONE, 8(7), e68533. https://doi.org/10.1371/journal.pone.0068533

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ, 1, e203. https://doi.org/10.7717/peerj.203

Torda, G., Donelson, J. M., Aranda, M., Barshis, D. J., Bay, L., Berumen, M. L., Bourne, D. G., Cantin, N., Foret, S., Matz, M., Miller, D. J., Moya, A., Putnam, H. M., Ravasi, T., van Oppen, M. J. H., Thurber, R. V., Vidal-Dupiol, J., Voolstra, C. R., Watson, S.-A., … Munday, P. L. (2017). Rapid adaptive responses to climate change in corals. Nature Climate Change, 7(9), 627–636. https://doi.org/10.1038/nclimate3374

Torquato, F., Bouwmeester, J., Range, P., Marshell, A., Priest, M. A., Burt, J. A., Møller, P. R., & Ben-Hamadou, R. (2022). Population genetic structure of a major reef-building coral species *Acropora downingi* in northeastern Arabian Peninsula. Coral Reefs, 41(3), 743–752. https://doi.org/10.1007/s00338-021-02158-y

Torres, A. F., Valino, D. A. M., & Ravago-Gotanco, R. (2021). Zooxanthellae Diversity and Coral-Symbiont Associations in the Philippine Archipelago: Specificity and Adaptability Across Thermal Gradients. Frontiers in Marine Science, 8, 731023. https://doi.org/10.3389/fmars.2021.731023

Totikov, A., Tomarovsky, A., Prokopov, D., Yakupova, A., Bulyonkova, T., Derezanin, L., Rasskazov, D., Wolfsberger, W. W., Koepfli, K.-P., Oleksyk, T. K., & Kliver, S. (2021). Chromosome-Level Genome Assemblies Expand Capabilities of Genomics for Conservation Biology. Genes, 12(9), 1336. https://doi.org/10.3390/genes12091336

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics, 13(1), 36–46. https://doi.org/10.1038/nrg3117

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic Islands of Speciation in Anopheles gambiae. PLoS Biology, 3(9), e285. https://doi.org/10.1371/journal.pbio.0030285

Underwood, J. N., Richards, Z., Berry, O., Oades, D., Howard, A., & Gilmour, J. P. (2020). Extreme seascape drives local recruitment and genetic divergence in brooding and spawning corals in remote north ‐ west Australia. Evolutionary Applications, 13(9), 2404–2421. https://doi.org/10.1111/eva.13033

Underwood, J. N., Richards, Z. T., Miller, K. J., Puotinen, M. L., & Gilmour, J. P. (2018). Genetic signatures through space, time and multiple disturbances in a ubiquitous brooding coral. Molecular Ecology, 27(7), 1586–1602. https://doi.org/10.1111/mec.14559

Underwood, J. N., SMITH, L. D., van Oppen, M. J. H., & Gilmour, J. P. (2006). Multiple scales of genetic connectivity in a brooding coral on isolated reefs following catastrophic bleaching: GENETIC CONNECTIVITY IN A BROODING CORAL. Molecular Ecology, 16(4), 771–784. https://doi.org/10.1111/j.1365-294x.2006.03187.x

Underwood, J. N., Wilson, S. K., Ludgerus, L., & Evans, R. D. (2013). Integrating connectivity science and spatial conservation management of coral reefs in north-west Australia. Journal for Nature Conservation, 21(3), 163–172. https://doi.org/10.1016/j.jnc.2012.12.001

van Oppen, M. J. H., Peplow, L. M., Kininmonth, S., & Berkelmans, R. (2011). Historical and contemporary factors shape the population genetic structure of the broadcast spawning coral, *Acropora millepora*, on the Great Barrier Reef. Molecular Ecology, 20(23), 4899–4914. https://doi.org/10.1111/j.1365-294x.2011.05328.x

Voolstra, C. R., Quigley, K. M., Davies, S. W., Parkinson, J. E., Peixoto, R. S., Aranda, M., Baker, A. C., Barno, A. R., Barshis, D. J., Benzoni, F., Bonito, V., Bourne, D. G., Buitrago-López, C., Bridge, T. C. L., Chan, C. X., Combosch, D. J., Craggs, J., Frommlet, J. C., Herrera, S., … Sweet, M. (2021). Consensus Guidelines for Advancing Coral Holobiont Genome and Specimen Voucher Deposition. Frontiers in Marine Science, 8, 701784. https://doi.org/10.3389/fmars.2021.701784

Voolstra, C. R., Suggett, D. J., Peixoto, R. S., Parkinson, J. E., Quigley, K. M., Silveira, C. B., Sweet, M., Muller, E. M., Barshis, D. J., Bourne, D. G., & Aranda, M. (2021). Extending the natural adaptive capacity of coral holobionts. Nature Reviews Earth & Environment, 2(11), 747–762. https://doi.org/10.1038/s43017-021-00214-3

Wan, L., Reinert, G., Sun, F., & Waterman, M. S. (2010). Alignment-Free Sequence Comparison (II): Theoretical Power of Comparison Statistics. Journal of Computational Biology, 17(11), 1467–1490. https://doi.org/10.1089/cmb.2010.0056

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods, 9(8), 808–810. https://doi.org/10.1038/nmeth.2023

Wang, X., Pedersen, C.-E. T., Athanasiadis, G., Garcia-Erill, G., Hanghøj, K., Bertola, L. D., Rasmussen, M. S., Schubert, M., Liu, X., Li, Z., Lin, L., Jørsboe, E., Nursyifa, C., Liu, S., Muwanika, V., Masembe, C., Chen, L., Wang, W., Moltke, I., … Heller, R. (2022). Persistent gene flow suggests an absence of reproductive isolation in an African antelope speciation model. https://doi.org/10.1101/2022.12.08.519574

Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency ‐ free inference of close familial relationships from genotypes or low ‐ depth sequencing data. Molecular Ecology, 28(1), 35 ‐ 48. https://doi.org/10.1111/mec.14954

Warner, P. A., van Oppen, M. J. H., & Willis, B. L. (2015). Unexpected cryptic species diversity in the widespread coral Seriatopora hystrix masks spatial ‐ genetic patterns of connectivity. Molecular Ecology, 24(12), 2993 ‐ 3008. https://doi.org/10.1111/mec.13225

Webster, J. M., & Davies, P. J. (2003). Coral variation in two deep drill cores: significance for the Pleistocene development of the Great Barrier Reef. Sedimentary Geology, 159(1–2), 61–80. https://doi.org/10.1016/s0037-0738(03)00095-2
Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., Peona, V., Pophaly, S. D., Sedlazeck, F. J., Suh, A., Warmuth, V. M., & Wolf, J. B. W. (2020). Discovery and population genomics of structural variation in a songbird genus. Nature Communications, 11(1), 3403. https://doi.org/10.1038/s41467-020-17195-4
Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. Trends in Ecology & Evolution, 33(6), 427–440. https://doi.org/10.1016/j.tree.2018.04.002
Wham, D. C., Ning, G., & LaJeunesse, T. C. (2017). Symbiodinium glynnii sp. nov., a species of stress-tolerant symbiotic dinoflagellates from pocilloporid and montiporid corals in the Pacific Ocean. Phycologia, 56(4), 396–409. https://doi.org/10.2216/16-86.1
Whitaker, K. (2006). Genetic evidence for mixed modes of reproduction in the coral Pocillopora damicornis and its effect on population structure. Marine Ecology Progress Series, 306, 115–124. https://doi.org/10.3354/meps306115
Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology, 15(3), R46. https://doi.org/10.1186/gb-2014-15-3-r46
Xu, S., Song, N., Zhao, L., Cai, S., Han, Z., & Gao, T. (2017). Genomic evidence for local adaptation in the ovoviviparous marine fish Sebastiscus marmoratus with a background of population homogeneity. Scientific Reports, 7(1), 1562. https://doi.org/10.1038/s41598-017-01742-z
Ying, H., Cooke, I., Sprungala, S., Wang, W., Hayward, D. C., Tang, Y., Huttley, G., Ball, E. E., Forêt, S., & Miller, D. J. (2018). Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. Genome Biology, 19(1), 175. https://doi.org/10.1186/s13059-018-1552-8
Ying, H., Hayward, D. C., Cooke, I., Wang, W., Moya, A., Siemering, K. R., Sprungala, S., Ball, E. E., Forêt, S., & Miller, D. J. (2019). The Whole-Genome Sequence of the Coral *Acropora millepora*. Genome Biology and Evolution, 11(5), 1374–1379. https://doi.org/10.1093/gbe/evz077
Yorifuji, M., Harii, S., Nakamura, R., & Fudo, M. (2017). Shift of symbiont communities in *Acropora tenuis* juveniles under heat stress. PeerJ, 5, e4055. https://doi.org/10.7717/peerj.4055
Young, C., Schopmeyer, S., & 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
Yu, G., Lam, T. T.-Y., Zhu, H., & Guan, Y. (2018). Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. Molecular Biology and Evolution, 35(12), 3041–3043. https://doi.org/10.1093/molbev/msy194
Zayasu, Y., Takeuchi, T., Nagata, T., Kanai, M., Fujie, M., Kawamitsu, M., Chinen, W., Shinzato, C., & Satoh, N. (2021). Genome‐wide SNP genotyping reveals hidden population structure of an acroporid species at a subtropical coral island: Implications for coral restoration. Aquatic Conservation: Marine and Freshwater Ecosystems, 31(9), 2429–2439. https://doi.org/10.1002/aqc.3626
Zhang, J., Richards, Z. T., Adam, A. A. S., Chan, C. X., Shinzato, C., Gilmour, J., Thomas, L., Strugnell, J. M., Miller, D. J., & Cooke, I. (2022). Evolutionary Responses of a Reef-building Coral to Climate Change at the End of the Last Glacial Maximum. Molecular Biology and Evolution, 39(10). https://doi.org/10.1093/molbev/msac201
Zhang, S., Wang, G.-D., Ma, P., Zhang, L., Yin, T.-T., Liu, Y., Otecko, N. O., Wang, M., Ma, Y., Wang, L., Mao, B., Savolainen, P., & Zhang, Y. (2020). Genomic regions under selection in the feralization of the dingoes. Nature Communications, 11(1), 671. https://doi.org/10.1038/s41467-020-14515-6

Zhao, L., Qu, F., Song, N., Han, Z., Gao, T., & Zhang, Z. (2021). Population genomics provides insights into the population structure and temperature-driven adaptation of Collichthys lucidus. BMC Genomics, 22(1), 729. https://doi.org/10.1186/s12864-021-08045-8

Zhao, T., Duan, Z., Genchev, G. Z., & Lu, H. (2020). Closing Human Reference Genome Gaps: Identifying and Characterizing Gap-Closing Sequences. G3: Genes|Genomes|Genetics, 10(8), 2801–2809. https://doi.org/10.1534/g3.120.401280

Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., Pan, X., Gong, M., Zhang, Y., Guo, Y., Wang, Y., Liu, J., Cai, Y., Chen, Q., Okpeku, M., Colli, L., Cai, D., Wang, K., Huang, S., … Jiang, Y. (2020). The origin of domestication genes in goats. Science Advances, 6(21), eaaz5216. https://doi.org/10.1126/sciadv.aaz5216

Zhu, Y., Bergland, A. O., González, J., & Petrov, D. A. (2012). Empirical Validation of Pooled Whole Genome Population Re-Sequencing in Drosophila melanogaster. PLoS ONE, 7(7), e41901. https://doi.org/10.1371/journal.pone.0041901

Ziegler, M., Eguíluz, V. M., Duarte, C. M., & Voolstra, C. R. (2018). Rare symbionts may contribute to the resilience of coral–algal assemblages. The ISME Journal, 12(1), 161–172. https://doi.org/10.1038/ismej.2017.151

Zielezinski, A., Girgis, H. Z., Bernard, G., Leimeister, C.-A., Tang, K., Dencker, T., Lau, A. K., Röhling, S., Choi, J. J., Waterman, M. S., Comin, M., Kim, S.-H., Vinga, S., Almeida, J. S., Chan, C. X., James, B. T., Sun, F., Morgenstern, B., & Karlowski, W. M. (2019). Benchmarking of alignment-free sequence comparison methods. Genome Biology, 20(1), 144. https://doi.org/10.1186/s13059-019-1755-7

Zielezinski, A., Vinga, S., Almeida, J., & Karlowski, W. M. (2017). Alignment-free sequence comparison: benefits, applications, and tools. Genome Biology, 18(1), 186. https://doi.org/10.1186/s13059-017-1319-7

Zilber‐Rosenberg, I., & Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiology Reviews, 32(5), 723–735. https://doi.org/10.1111/j.1574-6976.2008.00123.x

Supplementary Material

[blank page]

Supplementary Table 2.1 Sample information of *Acropora tenuis* samples (n = 228) sequenced with low-coverage whole-genome sequencing

ID	Pop	Reef	Location	# of Reads	Total bases	Sequencing depth	% Mapped	Genome coverage (% of the base in assembly covered by at least one read	Mapping depth (average depth across the genome)	Net mapping depth (average depth in covered genome)	Passed QC?	Reason to exclude
ARL_11	ARL	Arlington Reef	Offshore	22439072	2256641454	4.64	95.37%	86.30%	4.13057	4.78607	Yes	-
ARL_13	ARL	Arlington Reef	Offshore	15951578	1603343021	3.29	86.95%	56.45%	2.29613	4.06732	No	not A.tenius
ARL_14	ARL	Arlington Reef	Offshore	18781942	1887629025	3.88	87.97%	58.23%	2.74255	4.70976	No	not A.tenius
ARL_15	ARL	Arlington Reef	Offshore	19277960	1936835512	3.98	89.19%	82.52%	3.29972	3.99877	Yes	Admixed
ARL_16	ARL	Arlington Reef	Offshore	19115022	1920737036	3.95	93.96%	84.23%	3.43189	4.07453	Yes	-
ARL_17	ARL	Arlington Reef	Offshore	16161432	1620575796	3.33	89.71%	80.11%	2.79136	3.48432	Yes	-
ARL_19	ARL	Arlington Reef	Offshore	20879550	2098191483	4.31	95.89%	84.85%	3.8524	4.54019	Yes	-
ARL_1	ARL	Arlington Reef	Offshore	16284322	1635920404	3.36	87.80%	55.56%	2.37513	4.27473	No	not A.tenius
ARL_20	ARL	Arlington Reef	Offshore	13617076	1365941083	2.81	88.15%	52.20%	1.99985	3.8309	No	not A.tenius
ARL_21	ARL	Arlington Reef	Offshore	16223398	1629426122	3.35	84.83%	54.58%	2.19529	4.02236	No	not A.tenius
ARL_22	ARL	Arlington Reef	Offshore	16361816	1642620864	3.37	88.66%	55.84%	2.41891	4.3317	No	not A.tenius
ARL_23	ARL	Arlington Reef	Offshore	17395882	1744913659	3.58	87.23%	55.83%	2.49819	4.4744	No	not A.tenius

ARL_27	ARL	Arlington Reef	Offshore	15627468	1571082312	3.23	94.45%	80.86%	2.85863	3.53526	Yes	-
ARL_28	ARL	Arlington Reef	Offshore	14077088	1415596776	2.91	92.74%	78.05%	2.53543	3.2486	Yes	-
ARL_29	ARL	Arlington Reef	Offshore	20461054	2057052288	4.23	94.94%	85.14%	3.75706	4.41256	Yes	-
ARL_2	ARL	Arlington Reef	Offshore	14112956	1418196011	2.91	85.30%	76.14%	2.32203	3.04953	Yes	-
ARL_30	ARL	Arlington Reef	Offshore	14615568	1469577518	3.02	90.69%	79.21%	2.56942	3.24397	Yes	-
ARL_3	ARL	Arlington Reef	Offshore	15365674	1542132135	3.17	87.32%	54.71%	2.23413	4.08348	No	not A.tenius
ARL_4	ARL	Arlington Reef	Offshore	19038616	1910392774	3.92	94.40%	83.60%	3.45569	4.1337	Yes	-
ARL_5	ARL	Arlington Reef	Offshore	22963566	2307312709	4.74	91.35%	79.75%	3.6666	4.5979	No	not A.tenius
DI-1-10	DI	Dunk Island	Inshore	11407876	1152195476	2.37	95.75%	71.71%	2.12171	2.95863	Yes	-
DI-1-12	DI	Dunk Island	Inshore	12489526	1261442126	2.59	96.55%	75.04%	2.3499	3.13133	Yes	-
DI-1-15	DI	Dunk Island	Inshore	13517170	1365234170	2.80	95.98%	76.39%	2.53586	3.31959	Yes	-
DI-1-17	DI	Dunk Island	Inshore	13062902	1319353102	2.71	92.38%	75.25%	2.34167	3.11204	Yes	-
DI-1-1	DI	Dunk Island	Inshore	15271400	1542411400	3.17	95.76%	76.95%	2.80644	3.64723	Yes	-
DI-1-20	DI	Dunk Island	Inshore	12681536	1280835136	2.63	95.45%	73.73%	2.35115	3.18905	Yes	-
DI-1-23	DI	Dunk Island	Inshore	9697308	979428108	2.01	94.23%	66.02%	1.76506	2.67352	Yes	-
DI-1-25	DI	Dunk Island	Inshore	13881412	1402022612	2.88	92.22%	75.36%	2.48364	3.29585	Yes	-
DI-1-27	DI	Dunk Island	Inshore	11045228	1115568028	2.29	92.97%	70.23%	1.99683	2.84339	Yes	-
DI-1-2	DI	Dunk Island	Inshore	12826598	1295486398	2.66	95.73%	74.89%	2.39466	3.19765	Yes	-
DI-1-32	DI	Dunk Island	Inshore	12625190	1275144190	2.62	95.57%	74.78%	2.35141	3.14436	Yes	-
DI-1-3	DI	Dunk Island	Inshore	20372002	2057572202	4.23	96.77%	83.38%	3.82885	4.59206	Yes	-
DI-1-4	DI	Dunk Island	Inshore	15201706	1535372306	3.15	94.19%	77.01%	2.77186	3.59946	Yes	-
DI-1-6	DI	Dunk Island	Inshore	17828078	1800635878	3.70	94.03%	81.02%	3.25035	4.01175	Yes	-
DI-1-8	DI	Dunk Island	Inshore	11639418	1175581218	2.41	92.92%	71.09%	2.10233	2.95709	Yes	-
DI-2-10	DI	Dunk Island	Inshore	12636456	1276282056	2.62	92.52%	73.36%	2.26102	3.08199	Yes	-
DI-2-13	DI	Dunk Island	Inshore	16630506	1679681106	3.45	91.73%	79.22%	2.95793	3.73373	Yes	-

DI-2-16	DI	Dunk Island	Inshore	15916554	1607571954	3.30	87.88%	77.77%	2.71993	3.49722	Yes	-
DI-2-17	DI	Dunk Island	Inshore	15508600	1566368600	3.22	91.10%	77.38%	2.73454	3.53378	Yes	-
DI-2-18	DI	Dunk Island	Inshore	12453888	1257842688	2.58	90.76%	71.67%	2.18316	3.04623	Yes	-
DI-2-19	DI	Dunk Island	Inshore	17228130	1740041130	3.57	92.77%	79.13%	3.10036	3.91807	Yes	-
DI-2-1	DI	Dunk Island	Inshore	13652208	1378873008	2.83	95.24%	74.71%	2.5166	3.36847	Yes	-
DI-2-22	DI	Dunk Island	Inshore	12003122	1212315322	2.49	91.01%	69.39%	2.06982	2.98295	Yes	-
DI-2-24	DI	Dunk Island	Inshore	17022078	1719229878	3.53	93.25%	78.78%	3.07165	3.89899	Yes	-
DI-2-26	DI	Dunk Island	Inshore	13962284	1410190684	2.90	92.35%	77.18%	2.51398	3.25741	Yes	-
DI-2-28	DI	Dunk Island	Inshore	13069438	1320013238	2.71	93.68%	73.61%	2.36527	3.21312	Yes	-
DI-2-29	DI	Dunk Island	Inshore	10877392	1098616592	2.26	94.81%	71.62%	2.01827	2.81793	Yes	-
DI-2-30	DI	Dunk Island	Inshore	13986806	1412667406	2.90	88.56%	75.35%	2.40493	3.19153	Yes	-
DI-2-4	DI	Dunk Island	Inshore	7955274	803482674	1.65	94.80%	61.67%	1.46431	2.37425	Yes	Admixed
DI-2-7	DI	Dunk Island	Inshore	14093664	1423460064	2.92	86.81%	74.77%	2.35091	3.14419	Yes	-
FI-1-10	FI	Fitzroy Island	Inshore	8925458	901471258	1.85	93.53%	62.90%	1.60572	2.55282	Yes	-
FI-1-12	FI	Fitzroy Island	Inshore	13819860	1395805860	2.87	91.84%	76.74%	2.47628	3.22687	Yes	-
FI-1-13	FI	Fitzroy Island	Inshore	16280352	1644315552	3.38	94.04%	79.99%	2.9897	3.73768	Yes	-
FI-1-14	FI	Fitzroy Island	Inshore	10762290	1086991290	2.23	92.84%	68.82%	1.9366	2.81416	Yes	-
FI-1-16	FI	Fitzroy Island	Inshore	8892028	898094828	1.84	93.08%	64.04%	1.61054	2.5148	Yes	-
FI-1-18	FI	Fitzroy Island	Inshore	13404820	1353886820	2.78	89.24%	74.64%	2.32842	3.11953	Yes	-
FI-1-1	FI	Fitzroy Island	Inshore	17277302	1745007502	3.58	94.53%	79.74%	3.09411	3.88023	Yes	-
FI-1-21	FI	Fitzroy Island	Inshore	16348904	1651239304	3.39	93.64%	79.40%	2.97801	3.75078	Yes	-
FI-1-24	FI	Fitzroy Island	Inshore	15676598	1583336398	3.25	91.60%	77.21%	2.76766	3.58481	Yes	-
FI-1-26	FI	Fitzroy Island	Inshore	16794120	1696206120	3.48	94.60%	79.59%	3.07675	3.86567	Yes	-
FI-1-27	FI	Fitzroy Island	Inshore	14862538	1501116338	3.08	93.42%	77.38%	2.7021	3.49212	Yes	-
FI-1-30	FI	Fitzroy Island	Inshore	14052434	1419295834	2.92	93.66%	75.04%	2.53598	3.37929	Yes	-
FI-1-3	FI	Fitzroy Island	Inshore	114928038	11607731838	23.84	96.01%	92.76%	21.1214	22.7707	Yes	-

FI-1-4	FI	Fitzroy Island	Inshore	9903788	1000282588	2.05	94.02%	68.38%	1.81689	2.65694	Yes	-
FI-1-5	FI	Fitzroy Island	Inshore	16339884	1650328284	3.39	91.74%	79.59%	2.90047	3.64411	Yes	-
FI-2-10	FI	Fitzroy Island	Inshore	16827330	1699560330	3.49	94.36%	79.41%	3.07383	3.87062	Yes	-
FI-2-13	FI	Fitzroy Island	Inshore	18230676	1841298276	3.78	95.40%	82.05%	3.3799	4.11952	Yes	-
FI-2-14	FI	Fitzroy Island	Inshore	22721790	2294900790	4.71	94.94%	83.90%	4.173	4.97363	Yes	-
FI-2-19	FI	Fitzroy Island	Inshore	13022542	1315276742	2.70	93.53%	74.76%	2.36814	3.16753	Yes	-
FI-2-21	FI	Fitzroy Island	Inshore	12242798	1236522598	2.54	93.00%	74.01%	2.21729	2.99581	Yes	-
FI-2-23	FI	Fitzroy Island	Inshore	10979406	1108920006	2.28	93.69%	71.29%	2.00062	2.80632	Yes	-
FI-2-25	FI	Fitzroy Island	Inshore	15642020	1579844020	3.25	96.31%	78.34%	2.93212	3.74286	Yes	-
FI-2-27	FI	Fitzroy Island	Inshore	15981944	1614176344	3.32	93.05%	78.43%	2.87403	3.66438	Yes	-
FI-2-29	FI	Fitzroy Island	Inshore	15048162	1519864362	3.12	94.12%	77.96%	2.74491	3.52099	Yes	-
FI-2-2	FI	Fitzroy Island	Inshore	16050608	1621111408	3.33	88.00%	76.58%	2.72715	3.5613	Yes	-
FI-2-30	FI	Fitzroy Island	Inshore	4478452	452323652	0.93	97.14%	43.50%	0.832416	1.91365	Yes	-
FI-2-5	FI	Fitzroy Island	Inshore	14478178	1462295978	3.00	95.27%	78.43%	2.69223	3.43258	Yes	-
FI-2-6	FI	Fitzroy Island	Inshore	15866210	1602487210	3.29	95.09%	79.08%	2.9325	3.70848	Yes	-
FI-2-8	FI	Fitzroy Island	Inshore	14277810	1442058810	2.96	93.16%	77.18%	2.58356	3.34756	Yes	-
FI-2-9	FI	Fitzroy Island	Inshore	14748348	1489583148	3.06	93.15%	77.25%	2.65216	3.43307	Yes	-
JB_10	JB	John Brewer Reef	Offshore	13560962	1359810920	2.79	92.43%	76.99%	2.42552	3.15026	Yes	-
JB_11	JB	John Brewer Reef	Offshore	16163558	1624517344	3.34	91.35%	78.79%	2.83933	3.60351	Yes	-
JB_12	JB	John Brewer Reef	Offshore	15319042	1539388772	3.16	93.88%	79.61%	2.78801	3.50215	Yes	-
JB_13	JB	John Brewer Reef	Offshore	14808056	1488345910	3.06	92.97%	79.34%	2.66539	3.3594	Yes	-
JB_15	JB	John Brewer Reef	Offshore	15018672	1508783154	3.10	93.59%	79.90%	2.71816	3.40215	Yes	-
JB_16	JB	John Brewer Reef	Offshore	17290734	1737616965	3.57	87.13%	80.68%	2.90966	3.60659	Yes	-

JB_17	JB	John Brewer Reef	Offshore	12305534	1235238116	2.54	93.72%	75.95%	2.23465	2.94215	Yes	-
JB_18	JB	John Brewer Reef	Offshore	13582072	1362731508	2.80	92.47%	74.74%	2.41854	3.23573	Yes	-
JB_19	JB	John Brewer Reef	Offshore	15607330	1567513846	3.22	92.41%	80.20%	2.79088	3.47979	Yes	-
JB_1	JB	John Brewer Reef	Offshore	16476170	1656634431	3.40	96.26%	82.17%	3.07821	3.74613	Yes	-
JB_20	JB	John Brewer Reef	Offshore	15982366	1606273446	3.30	89.58%	79.98%	2.76921	3.46242	Yes	-
JB_21	JB	John Brewer Reef	Offshore	16587914	1667816298	3.43	93.19%	81.74%	2.99287	3.66132	Yes	-
JB_22	JB	John Brewer Reef	Offshore	14748266	1482054225	3.04	95.36%	79.75%	2.72194	3.41315	Yes	-
JB_23	JB	John Brewer Reef	Offshore	14037912	1410808515	2.90	88.69%	77.72%	2.40725	3.09724	Yes	-
JB_2	JB	John Brewer Reef	Offshore	20446292	2055780608	4.22	93.44%	84.62%	3.70185	4.37476	Yes	-
JB_3	JB	John Brewer Reef	Offshore	19147964	1923683384	3.95	91.90%	83.90%	3.40697	4.0607	Yes	-
JB_4	JB	John Brewer Reef	Offshore	16345110	1641947591	3.37	92.39%	80.96%	2.92068	3.6077	Yes	-
JB_5	JB	John Brewer Reef	Offshore	16001642	1607094701	3.30	87.56%	55.19%	2.31752	4.19912	No	not A.tenius
JB_8	JB	John Brewer Reef	Offshore	19310406	1941161568	3.99	94.02%	83.34%	3.50814	4.20943	Yes	-
JB_9	JB	John Brewer Reef	Offshore	14708740	1476337675	3.03	94.14%	77.25%	2.6778	3.46644	Yes	-
MI-1-10	MI	Magnetic Island	Inshore	6821966	689018566	1.42	92.21%	54.02%	1.20298	2.22671	Yes	-
MI-1-12	MI	Magnetic Island	Inshore	4862210	491083210	1.01	93.38%	46.50%	0.875726	1.8832	No	Relatedness

MI-1-13	MI	Magnetic Island	Inshore	5135290	518664290	1.07	91.82%	45.42%	0.885531	1.94976	Yes	-
MI-1-16	MI	Magnetic Island	Inshore	13265622	1339827822	2.75	88.85%	70.66%	2.24959	3.18346	No	Relatedness
MI-1-19	MI	Magnetic Island	Inshore	9730388	982769188	2.02	91.85%	64.71%	1.7091	2.64123	Yes	-
MI-1-1	MI	Magnetic Island	Inshore	15348364	1550184764	3.18	91.97%	77.11%	2.73205	3.5431	Yes	Admixed
MI-1-22	MI	Magnetic Island	Inshore	12320986	1244419586	2.56	93.94%	71.10%	2.24037	3.15116	Yes	-
MI-1-2	MI	Magnetic Island	Inshore	13084170	1321501170	2.71	94.41%	74.57%	2.41471	3.23796	Yes	-
MI-1-3	MI	Magnetic Island	Inshore	4191526	423344126	0.87	93.32%	42.20%	0.749027	1.77497	Yes	-
MI-1-4	MI	Magnetic Island	Inshore	124838162	12608654362	25.90	90.54%	91.29%	21.0872	23.1003	Yes	-
MI-1-5	MI	Magnetic Island	Inshore	9095742	918669942	1.89	93.65%	64.72%	1.66202	2.5681	No	Relatedness
MI-1-6	MI	Magnetic Island	Inshore	14617666	1476384266	3.03	95.17%	77.06%	2.71152	3.51888	Yes	-
MI-1-8	MI	Magnetic Island	Inshore	11103610	1121464610	2.30	94.86%	70.95%	2.05008	2.88949	Yes	-
MI-2-12	MI	Magnetic Island	Inshore	9471226	956593826	1.97	92.13%	64.23%	1.66774	2.59635	Yes	-
MI-2-14	MI	Magnetic Island	Inshore	6228386	629066986	1.29	90.55%	52.77%	1.092	2.06922	Yes	-
MI-2-16	MI	Magnetic Island	Inshore	4478058	452283858	0.93	88.24%	38.82%	0.739871	1.9057	No	Relatedness
MI-2-18	MI	Magnetic Island	Inshore	7661034	773764434	1.59	86.81%	55.97%	1.27349	2.27537	Yes	-
MI-2-21	MI	Magnetic Island	Inshore	6699250	676624250	1.39	89.53%	54.34%	1.15518	2.12574	No	Relatedness

MI-2-23	MI	Magnetic Island	Inshore	8963460	905309460	1.86	94.20%	62.73%	1.61804	2.57919	Yes	-
MI-2-24	MI	Magnetic Island	Inshore	10490758	1059566558	2.18	89.10%	64.85%	1.78651	2.75483	Yes	-
MI-2-27	MI	Magnetic Island	Inshore	11192884	1130481284	2.32	89.20%	68.97%	1.92619	2.79278	Yes	-
MI-2-28	MI	Magnetic Island	Inshore	9828210	992649210	2.04	89.33%	63.99%	1.68178	2.62823	Yes	-
MI-2-29	MI	Magnetic Island	Inshore	6713958	678109758	1.39	93.37%	53.02%	1.18724	2.23934	Yes	-
MI-2-3	MI	Magnetic Island	Inshore	10484406	1058925006	2.18	93.39%	66.67%	1.87358	2.81011	No	Relatedness
MI-2-4	MI	Magnetic Island	Inshore	9162154	925377554	1.90	90.26%	62.66%	1.58515	2.52963	Yes	-
MI-2-6	MI	Magnetic Island	Inshore	4682482	472930682	0.97	90.95%	44.12%	0.811325	1.83899	Yes	-
MI-2-8	MI	Magnetic Island	Inshore	11597204	1171317604	2.41	90.55%	67.17%	1.99746	2.9736	Yes	-
MI-2-9	MI	Magnetic Island	Inshore	16330178	1649347978	3.39	90.92%	75.95%	2.83473	3.73252	Yes	-
PI-1-10	PI	Pelorus Island	Inshore	12482644	1260747044	2.59	94.60%	73.57%	2.30012	3.12664	Yes	-
PI-1-14	PI	Pelorus Island	Inshore	15615770	1577192770	3.24	95.95%	79.52%	2.92451	3.67759	Yes	-
PI-1-16	PI	Pelorus Island	Inshore	12436992	1256136192	2.58	95.17%	71.47%	2.28373	3.19516	Yes	Admixed
PI-1-17	PI	Pelorus Island	Inshore	15078162	1522894362	3.13	93.78%	78.12%	2.74873	3.5185	Yes	-
PI-1-1	PI	Pelorus Island	Inshore	15524040	1567928040	3.22	93.49%	78.90%	2.82297	3.57769	Yes	-
PI-1-20	PI	Pelorus Island	Inshore	16026216	1618647816	3.32	96.39%	80.23%	3.00958	3.75099	Yes	-
PI-1-22	PI	Pelorus Island	Inshore	19501798	1969681598	4.05	96.06%	83.20%	3.63591	4.37033	Yes	-
PI-1-24	PI	Pelorus Island	Inshore	15829730	1598802730	3.28	95.45%	79.68%	2.92643	3.67252	Yes	-
PI-1-25	PI	Pelorus Island	Inshore	12626500	1275276500	2.62	95.19%	75.17%	2.34959	3.12591	Yes	-
PI-1-27	PI	Pelorus Island	Inshore	12772878	1290060678	2.65	94.62%	73.54%	2.3502	3.19602	Yes	-

PI-1-29	PI	Pelorus Island	Inshore	15376684	1553045084	3.19	96.59%	79.19%	2.87652	3.63227	Yes	-
PI-1-2	PI	Pelorus Island	Inshore	13047156	1317762756	2.71	95.97%	76.40%	2.43175	3.18272	Yes	-
PI-1-3	PI	Pelorus Island	Inshore	8947662	903713862	1.86	94.46%	63.83%	1.61202	2.52559	Yes	-
PI-1-4	PI	Pelorus Island	Inshore	13887110	1402598110	2.88	94.59%	75.11%	2.52568	3.3627	Yes	-
PI-1-5	PI	Pelorus Island	Inshore	13782246	1392006846	2.86	95.77%	75.85%	2.56044	3.37584	Yes	-
PI-2-12	PI	Pelorus Island	Inshore	9421596	951581196	1.95	96.71%	68.01%	1.78708	2.62773	Yes	-
PI-2-14	PI	Pelorus Island	Inshore	17097688	1726866488	3.55	97.19%	81.73%	3.24055	3.96499	Yes	-
PI-2-15	PI	Pelorus Island	Inshore	11495764	1161072164	2.39	97.59%	73.01%	2.19092	3.00081	Yes	-
PI-2-16	PI	Pelorus Island	Inshore	12425408	1254966208	2.58	97.41%	73.39%	2.3552	3.20901	Yes	-
PI-2-17	PI	Pelorus Island	Inshore	13361018	1349462818	2.77	97.07%	75.89%	2.52834	3.33162	Yes	-
PI-2-20	PI	Pelorus Island	Inshore	9990100	1009000100	2.07	96.75%	66.39%	1.85739	2.79762	Yes	-
PI-2-22	PI	Pelorus Island	Inshore	17213702	1738583902	3.57	95.98%	79.41%	3.20872	4.04063	Yes	-
PI-2-23	PI	Pelorus Island	Inshore	16141866	1630328466	3.35	97.26%	79.13%	3.02485	3.82258	Yes	-
PI-2-24	PI	Pelorus Island	Inshore	12643584	1277001984	2.62	96.37%	73.25%	2.35387	3.21354	Yes	-
PI-2-25	PI	Pelorus Island	Inshore	6680778	674758578	1.39	96.60%	56.34%	1.25474	2.22693	Yes	-
PI-2-26	PI	Pelorus Island	Inshore	9652192	974871392	2.00	97.03%	68.67%	1.82117	2.65195	Yes	-
PI-2-29	PI	Pelorus Island	Inshore	12665476	1279213076	2.63	97.42%	71.94%	2.36089	3.28176	Yes	-
PI-2-2	PI	Pelorus Island	Inshore	19344024	1953746424	4.01	96.69%	82.26%	3.63144	4.41436	Yes	-
PI-2-4	PI	Pelorus Island	Inshore	13875590	1401434590	2.88	96.11%	76.01%	2.58776	3.40445	Yes	-
PI-2-7	PI	Pelorus Island	Inshore	14060328	1420093128	2.92	97.86%	77.07%	2.68766	3.48718	Yes	-
PR-1-10	PR	Pandora Reef	Inshore	8347058	843052858	1.73	94.74%	61.55%	1.52768	2.48214	Yes	-
PR-1-11	PR	Pandora Reef	Inshore	17109176	1728026776	3.55	95.17%	81.28%	3.16965	3.89945	Yes	-
PR-1-12	PR	Pandora Reef	Inshore	17830434	1800873834	3.70	93.58%	79.97%	3.2338	4.04367	Yes	-
PR-1-13	PR	Pandora Reef	Inshore	16170964	1633267364	3.36	94.75%	78.59%	2.93419	3.73337	Yes	-
PR-1-14	PR	Pandora Reef	Inshore	13751184	1388869584	2.85	96.87%	76.37%	2.58748	3.38791	Yes	-
PR-1-18	PR	Pandora Reef	Inshore	19400240	1959424240	4.03	94.10%	80.78%	3.47257	4.29871	Yes	-
PR-1-19	PR	Pandora Reef	Inshore	15246568	1539903368	3.16	95.14%	79.04%	2.80951	3.5544	Yes	-

PR-1-1	PR	Pandora Reef	Inshore	16679146	1684593746	3.46	94.51%	80.05%	3.0618	3.82479	Yes	-
PR-1-20	PR	Pandora Reef	Inshore	10338964	1044235364	2.15	93.50%	68.12%	1.87506	2.75273	Yes	-
PR-1-21	PR	Pandora Reef	Inshore	9520252	961545452	1.98	95.85%	66.84%	1.76986	2.64784	Yes	-
PR-1-22	PR	Pandora Reef	Inshore	14207944	1435002344	2.95	96.26%	78.11%	2.67034	3.41851	Yes	-
PR-1-24	PR	Pandora Reef	Inshore	13887314	1402618714	2.88	91.15%	75.35%	2.44835	3.24949	Yes	-
PR-1-26	PR	Pandora Reef	Inshore	16105346	1626639946	3.34	93.37%	78.95%	2.8961	3.66838	Yes	-
PR-1-29	PR	Pandora Reef	Inshore	24679148	2492593948	5.12	94.51%	85.11%	4.51615	5.30627	Yes	-
PR-1-2	PR	Pandora Reef	Inshore	15824198	1598243998	3.28	92.22%	78.39%	2.83134	3.61192	Yes	-
PR-1-30	PR	Pandora Reef	Inshore	8878322	896710522	1.84	84.82%	59.63%	1.45185	2.43483	Yes	-
PR-1-3	PR	Pandora Reef	Inshore	11722356	1183957956	2.43	94.46%	72.83%	2.16536	2.97317	Yes	-
PR-1-4	PR	Pandora Reef	Inshore	12440788	1256519588	2.58	95.80%	74.41%	2.31494	3.11093	Yes	-
PR-1-7	PR	Pandora Reef	Inshore	15096740	1524770740	3.13	93.50%	77.90%	2.73908	3.51593	Yes	-
PR-1-8	PR	Pandora Reef	Inshore	15611164	1576727564	3.24	95.16%	78.01%	2.88186	3.69427	Yes	-
PR-1-9	PR	Pandora Reef	Inshore	11867194	1198586594	2.46	93.14%	72.86%	2.15946	2.9639	Yes	-
PR-2-17	PR	Pandora Reef	Inshore	16174860	1633660860	3.36	92.20%	77.25%	2.85924	3.70109	Yes	-
PR-2-20	PR	Pandora Reef	Inshore	14158428	1430001228	2.94	96.46%	77.18%	2.65644	3.44198	Yes	-
PR-2-26	PR	Pandora Reef	Inshore	12037446	1215782046	2.50	95.11%	73.76%	2.22854	3.02142	Yes	-
PR-2-27	PR	Pandora Reef	Inshore	9679706	977650306	2.01	96.67%	66.63%	1.8197	2.73115	Yes	-
PR-2-30	PR	Pandora Reef	Inshore	12444364	1256880764	2.58	96.44%	73.24%	2.31523	3.16094	Yes	-
PR-2-4	PR	Pandora Reef	Inshore	13901514	1404052914	2.88	90.69%	76.34%	2.45075	3.21016	Yes	-
PR-2-6	PR	Pandora Reef	Inshore	10893136	1100206736	2.26	93.78%	72.47%	2.00399	2.76517	Yes	-
PR-2-7	PR	Pandora Reef	Inshore	11875466	1199422066	2.46	94.44%	72.03%	2.18014	3.0268	Yes	-
PR-2-8	PR	Pandora Reef	Inshore	14858716	1500730316	3.08	91.35%	76.68%	2.63555	3.43686	Yes	-
RIB_10	RIB	Rib Reef	Offshore	14836712	1490146284	3.06	92.59%	79.82%	2.66764	3.3419	Yes	-
RIB_11	RIB	Rib Reef	Offshore	15533788	1560727687	3.21	92.58%	79.97%	2.78674	3.48496	Yes	-
RIB_12	RIB	Rib Reef	Offshore	14772468	1483211196	3.05	95.50%	80.27%	2.7427	3.41674	Yes	-
RIB_13	RIB	Rib Reef	Offshore	16915626	1700094056	3.49	90.17%	80.92%	2.94076	3.63421	Yes	-

RIB_14	RIB	Rib Reef	Offshore	17344612	1743431340	3.58	94.49%	82.40%	3.16789	3.84451	Yes	-
RIB_15	RIB	Rib Reef	Offshore	15084010	1516459165	3.12	91.40%	79.68%	2.6697	3.35072	Yes	-
RIB_16	RIB	Rib Reef	Offshore	12991912	1305394164	2.68	93.78%	76.78%	2.35839	3.07178	Yes	-
RIB_17	RIB	Rib Reef	Offshore	16844310	1692810977	3.48	94.35%	82.38%	3.09013	3.75093	Yes	-
RIB_18	RIB	Rib Reef	Offshore	15390604	1546236390	3.18	84.47%	78.51%	2.51629	3.20489	Yes	-
RIB_19	RIB	Rib Reef	Offshore	18470508	1855470444	3.81	93.42%	83.13%	3.34585	4.02485	Yes	-
RIB_20	RIB	Rib Reef	Offshore	15298680	1535810830	3.15	91.98%	79.67%	2.70854	3.39978	Yes	-
RIB_21	RIB	Rib Reef	Offshore	19652608	1974365892	4.06	93.19%	83.58%	3.51405	4.20424	Yes	-
RIB_22	RIB	Rib Reef	Offshore	19867802	1997707986	4.10	95.27%	84.37%	3.68811	4.37135	Yes	-
RIB_23	RIB	Rib Reef	Offshore	16874474	1696553516	3.49	91.12%	81.60%	2.97522	3.64626	Yes	-
RIB_24	RIB	Rib Reef	Offshore	15713514	1578458684	3.24	87.11%	78.97%	2.63004	3.33039	Yes	-
RIB_4	RIB	Rib Reef	Offshore	20269900	2037610026	4.19	91.12%	83.90%	3.55526	4.23754	Yes	-
RIB_5	RIB	Rib Reef	Offshore	17429610	1752387248	3.60	90.07%	82.18%	3.0396	3.69879	Yes	-
RIB_7	RIB	Rib Reef	Offshore	15054008	1513125066	3.11	86.61%	78.65%	2.51873	3.20252	Yes	-
RIB_8	RIB	Rib Reef	Offshore	14537088	1460836044	3.00	90.52%	78.61%	2.54876	3.24245	Yes	-
RIB_9	RIB	Rib Reef	Offshore	15354996	1542489427	3.17	92.70%	79.78%	2.75867	3.45769	Yes	-
TAY_10	TAY	Taylor Reef	Offshore	15434164	1551034762	3.19	94.28%	80.28%	2.82272	3.51591	Yes	-
TAY_11	TAY	Taylor Reef	Offshore	13722818	1378452669	2.83	92.46%	77.87%	2.45638	3.15436	Yes	-
TAY_12	TAY	Taylor Reef	Offshore	11079950	1111759518	2.28	91.42%	72.98%	1.96062	2.68663	Yes	-
TAY_14	TAY	Taylor Reef	Offshore	17162974	1724738809	3.54	97.75%	83.14%	3.25787	3.91837	Yes	-
TAY_15	TAY	Taylor Reef	Offshore	17512212	1759879097	3.62	96.51%	82.49%	3.28475	3.98205	Yes	-
TAY_17	TAY	Taylor Reef	Offshore	18679052	1877941435	3.86	91.88%	83.13%	3.31896	3.99244	Yes	-
TAY_18	TAY	Taylor Reef	Offshore	15147064	1522141275	3.13	89.89%	79.58%	2.64053	3.31814	Yes	-
TAY_19	TAY	Taylor Reef	Offshore	14940028	1500699505	3.08	92.68%	79.58%	2.67297	3.35892	Yes	-
TAY_1	TAY	Taylor Reef	Offshore	12912994	1296227158	2.66	92.51%	76.37%	2.31138	3.02675	Yes	-
TAY_20	TAY	Taylor Reef	Offshore	20014894	2012019453	4.13	91.74%	84.34%	3.5578	4.2186	Yes	-
TAY_22	TAY	Taylor Reef	Offshore	16837432	1691514448	3.47	96.33%	82.17%	3.1361	3.81645	Yes	-

TAY_23	TAY	Taylor Reef	Offshore	14733898	1479065118	3.04	96.99%	80.21%	2.77009	3.45363	Yes	-
TAY_24	TAY	Taylor Reef	Offshore	16042098	1611539647	3.31	90.13%	80.32%	2.78557	3.46825	Yes	-
TAY_2	TAY	Taylor Reef	Offshore	14931028	1499306151	3.08	96.63%	80.60%	2.79857	3.47201	Yes	-
TAY_3	TAY	Taylor Reef	Offshore	17336844	1741410471	3.58	91.20%	82.02%	3.06197	3.7333	Yes	-
TAY_4	TAY	Taylor Reef	Offshore	16841184	1691916570	3.48	94.34%	82.30%	3.0837	3.7467	Yes	-
TAY_6	TAY	Taylor Reef	Offshore	18986608	1908812923	3.92	94.66%	84.36%	3.47668	4.12121	Yes	-
TAY_7	TAY	Taylor Reef	Offshore	19580696	1968338941	4.04	94.87%	84.14%	3.59658	4.27464	Yes	-
TAY_8	TAY	Taylor Reef	Offshore	13434020	1349581510	2.77	93.60%	77.67%	2.43161	3.13059	Yes	-
TAY_9	TAY	Taylor Reef	Offshore	13781960	1385079660	2.85	93.79%	78.29%	2.50219	3.196	Yes	-

* Sample MI-1-4 and FI-1-3 were downsampled after mapping to ~3X

Supplementary Table 2.2 Number and proportion of nucleotide bases retained after filtering steps applied on genome assembly of *A. tenuis*. SNPs were called within the filtered regions

	Total length	Percentage of assembly
Total	486,812,518	100.00%
Mappability (GenMap)	266,991,395	54.84%
Short simple repeats (mdust)	263,812,390	54.19%
Scaffold > 1Mb	263,710,454	54.17%
Depth	256,230,147	52.63%

Supplementary Table 2.3 The top blast hit of mitochondrial genomes of excluded samples

Sample ID	Top Blast Hit ID	% of Identity	Mismatch	Gap	Evalue	Species Name
ARL_13	gi|1469212651|dbj|LC201841.1|	99.782	5	10	0	*Acropora echinata*
ARL_14	gi|1469212651|dbj|LC201841.1|	99.793	3	10	0	*Acropora echinata*
ARL_1	gi|1469212651|dbj|LC201841.1|	99.793	3	10	0	*Acropora echinata*
ARL_20	gi|1469212651|dbj|LC201841.1|	99.782	5	10	0	*Acropora echinata*
ARL_21	gi|1469212637|dbj|LC201827.1|	99.679	26	15	0	*Acropora florida*
ARL_22	gi|1469212651|dbj|LC201841.1|	99.782	5	10	0	*Acropora echinata*
ARL_23	gi|1469212651|dbj|LC201841.1|	99.777	6	10	0	*Acropora echinata*
ARL_3	gi|1469212651|dbj|LC201841.1|	99.777	6	10	0	*Acropora echinata*
ARL_5	gi|1469212651|dbj|LC201841.1|	99.782	5	10	0	*Acropora echinata*
JB_5	gi|1469212651|dbj|LC201841.1|	99.777	6	10	0	*Acropora echinata*

Supplementary Table 2.4 Results of genetic statistic tests in nine populations

Population	Theta Watterson	Nucleotide diversity	Tajima's D
ARL	0.007367767	0.006103151	-0.755641
DI	0.0100687	0.006300541	-1.3683892
FI	0.009926469	0.006279366	-1.3405708
JB	0.007968827	0.006122773	-0.9104806

MI	0.00812494	0.00595776	-1.0324621
PI	0.010035599	0.006301289	-1.3626927
PR	0.009472735	0.006193823	-1.2652326
RIB	0.008071582	0.006165232	-0.921151
TAY	0.008030296	0.006132442	-0.9235433

Supplementary Table 2.5 List of genes located in the outlier region of F_{ST} Scans between inshore and offshore reefs

Gene id	Pfam annotation	Putative gene id	Gene name	GO annotation
aten_0.1.m1.28054	PF00574.22; PF00106.24; PF13561.5; PF08659.9; PF01370.20	Clpp	ATP-dependent Clp protease proteolytic subunit, mitochondrial (EC 3.4.21.92) (Endopeptidase Clp)	GO:0004176; GO:0004252; GO:0005739; GO:0005759; GO:0006515; GO:0009368; GO:0042802; GO:0051117; GO:0051260; GO:0051603
aten_0.1.m1.28056	PF15743.4; PF04508.11; PF14916.5	Gpsm2 Lgn Pins	G-protein-signaling modulator 2 (Pins homolog)	GO:0000132; GO:0000166; GO:0001965; GO:0005092; GO:0005737; GO:0005813; GO:0005829; GO:0005938; GO:0007052; GO:0008022; GO:0016328; GO:0019904; GO:0031291; GO:0032991; GO:0042802; GO:0043621; GO:0051301; GO:0051661; GO:0060236; GO:0070840; GO:0097431; GO:0097575; GO:0099738; GO:1904778; GO:1905832
aten_0.1.m1.28057	PF08146.11; PF12348.7; PF13646.5	HEATR1 BAP28 QnpA-17571	HEAT repeat-containing protein 1 (Protein BAP28) (Fragment)	GO:0005730; GO:0006364
aten_0.1.m1.28058	PF12397.7; PF13646.5; PF02985.21	HEATR1 BAP28 UTP10	HEAT repeat-containing protein 1 (Protein BAP28) (U3 small nucleolar RNA-associated protein 10 homolog) [Cleaved into: HEAT repeat-containing protein 1, N-terminally processed]	GO:0000462; GO:0001650; GO:0003723; GO:0005654; GO:0005730; GO:0005739; GO:0006364; GO:0016020; GO:0030686; GO:0032040; GO:0034455; GO:0045943; GO:2000234
aten_0.1.m1.28059	PF01522.20	NA	NA	NA

Supplementary Table 2.6 Genes on locus with putative inversion. Locus intervals are defined as the interval containing SNPs which the adjusted P-value for evidence of selection lower than 0.05. Lead SNP is the SNP with the best P-value

Scaffold	Start	End	Lead SNP	Chr*	P value	SNP outliers	Overlapped genes	Locus size (Mb)
Sc0000013	1234678	3069818	2738982	Chr1	1.31E-17	3740	TECPR1; thap1; Tmem186; KLF5 BTEB2 CKLF IKLF; slc32a1 viaat; THAP12 DAP4 P52RIPK PRKRIR THAP0; recS ypbC BSU23020; TRPM6 CHAK2; LUC7L3 CROP; LUC7L3 CROP; lsm11; Atp5pd Atp5h; gckr; Rint1; Tmf1 Ara160 Gm153; mrps24-a; LSH6 OBO6 At1g07090 F10K1.20; Rpl3; Harbi1; USP7 HAUSP; TRAF3 CAP1 CRAF1; LSH7 OBO7 At1g78815 F9K20.14; ZMYM2 ZNF198; Cbx7 D15Ertd417e	1.84
Sc0000097*	831393	1452742	831393	Chr2	9.52E-10	7	Trap1 Hsp75	0.621
Sc0000135	526260	880608	876696	Chr11	1.01E-15	73	Kcnb1; Sh mns CG12348	0.354
Sc0000185*	41650	310310	259719	Chr2	1.22E-09	5	Cbx7 D15Ertd417e	0.269
Sc0000214	345	241228	191862	Sc0000151	8.62E-16	583	K02A2.6	0.241
Sc0000066	1069464	1069464	1069464	Chr7	8.56E-09	1	ATF2 CREB2 CREBP1	0
Sc0000075	1184457	1184457	1184457	Sc0000050	7.37E-09	1	AN ORF133	0

*Sc0000097 and Sc0000185 formed the concatenated signals in Chr2 in Fig 2.3. The Chr field is the matched pseudo-chromosome id of each locus

Supplementary Table 2.7 Hardy-Weinberg equilibrium test for top SNPs of pcangsd scan

Scaffold	Top SNP	Major	Minor	F*	P value
Sc0000097	831393	T	C	0.089612	0.45
Sc0000066	1069464	C	T	0.561258	8.53E-04
Sc0000135	876696	C	T	0.044878	0.68
Sc0000013	2738982	C	G	-0.073186	0.45
Sc0000214	191862	C	T	0.168938	0.14
Sc0000075	1184457	T	C	0.019569	0.87
Sc0000185	259719	G	T	0.011748	0.92

*F: Inbreeding coefficient

Supplementary Table 3.1 Sample information of *Acropora digitifera* (n=75) sequenced with whole-genome sequencing

Sample ID	Location ID	Location name	Region of origin	Habitat	Batch 1 yield bases (Mb)	Batch 2 yield bases (Mb)	Total yield bases (Mb)
AI_1_001	AI	Adele Island	Inshore	Intertidal	11,435	0	11,435
AI_1_008	AI	Adele Island	Inshore	Intertidal	8,600	2,093	10,693
AI_1_021	AI	Adele Island	Inshore	Intertidal	22,350	0	22,350
AI_1_022	AI	Adele Island	Inshore	Intertidal	27,936	0	27,936
AI_1_023	AI	Adele Island	Inshore	Intertidal	11,545	0	11,545
AI_1_025	AI	Adele Island	Inshore	Intertidal	9,781	0	9,781
AI_2_036	AI	Adele Island	Inshore	Intertidal	12,936	0	12,936
AI_2_041	AI	Adele Island	Inshore	Intertidal	12,362	0	12,362
AI_2_043	AI	Adele Island	Inshore	Intertidal	14,622	0	14,622
AI_2_136	AI	Adele Island	Inshore	Intertidal	7,215	3,740	10,955
AI_2_151	AI	Adele Island	Inshore	Intertidal	14,137	0	14,137
AI_3_047	AI	Adele Island	Inshore	Intertidal	13,867	0	13,867
AI_3_060	AI	Adele Island	Inshore	Intertidal	17,243	0	17,243
AI_3_063	AI	Adele Island	Inshore	Intertidal	18,534	0	18,534
AI_3_071	AI	Adele Island	Inshore	Intertidal	10,905	0	10,905
AR_125_374	AR	Ashmore Reef	Offshore North	Subtidal	7,823	3,590	11,413

AR_125_377	AR	Ashmore Reef	Offshore North	Subtidal	14,274	0	14,274
AR_125_385	AR	Ashmore Reef	Offshore North	Subtidal	9,015	0	9,015
AR_125_388	AR	Ashmore Reef	Offshore North	Subtidal	15,577	0	15,577
AR_125_392	AR	Ashmore Reef	Offshore North	Subtidal	7,929	3,094	11,023
AR_128_316	AR	Ashmore Reef	Offshore North	Subtidal	4,042	2,234	6,276
AR_128_318	AR	Ashmore Reef	Offshore North	Subtidal	10,527	0	10,527
AR_128_326	AR	Ashmore Reef	Offshore North	Subtidal	8,745	2,471	11,216
AR_128_328	AR	Ashmore Reef	Offshore North	Subtidal	3,442	4,965	8,407
AR_128_336	AR	Ashmore Reef	Offshore North	Subtidal	5,308	5,401	10,709
AR_132_154	AR	Ashmore Reef	Offshore North	Subtidal	7,456	4,039	11,495
AR_132_162	AR	Ashmore Reef	Offshore North	Subtidal	10,512	0	10,512
AR_132_170	AR	Ashmore Reef	Offshore North	Subtidal	5,299	5,123	10,422
AR_132_173	AR	Ashmore Reef	Offshore North	Subtidal	10,334	0	10,334
AR_132_178	AR	Ashmore Reef	Offshore North	Subtidal	5,507	5,424	10,931
AR_133_341	AR	Ashmore Reef	Offshore North	Subtidal	11,756	0	11,756
AR_133_343	AR	Ashmore Reef	Offshore North	Subtidal	7,936	3,068	11,004
AR_133_346	AR	Ashmore Reef	Offshore North	Subtidal	8,145	2,717	10,862
AR_133_354	AR	Ashmore Reef	Offshore North	Subtidal	3,040	3,792	6,832

AR_133_357	AR	Ashmore Reef	Offshore North	Subtidal	12,206	0	12,206
BR_4_077	BR	Beagle Reef	Inshore	Intertidal	5,635	1,911	7,546
BR_4_078	BR	Beagle Reef	Inshore	Intertidal	9,204	0	9,204
BR_4_081	BR	Beagle Reef	Inshore	Intertidal	9,561	0	9,561
BR_4_082	BR	Beagle Reef	Inshore	Intertidal	10,763	0	10,763
BR_4_087	BR	Beagle Reef	Inshore	Intertidal	12,028	0	12,028
BR_4_088	BR	Beagle Reef	Inshore	Intertidal	5,085	2,000	7,085
BR_4_091	BR	Beagle Reef	Inshore	Intertidal	32,515	0	32,515
BR_4_100	BR	Beagle Reef	Inshore	Intertidal	13,814	0	13,814
BR_5_112	BR	Beagle Reef	Inshore	Intertidal	13,715	0	13,715
BR_5_114	BR	Beagle Reef	Inshore	Intertidal	19,638	0	19,638
BR_5_121*	BR	Beagle Reef	Inshore	Intertidal	9,755	0	9,755
BR_5_123	BR	Beagle Reef	Inshore	Intertidal	9,379	0	9,379
BR_5_124	BR	Beagle Reef	Inshore	Intertidal	9,032	0	9,032
BR_5_129	BR	Beagle Reef	Inshore	Intertidal	10,153	0	10,153
BR_5_133	BR	Beagle Reef	Inshore	Intertidal	18,762	0	18,762
RS1_2_417	RS1	Rowley Shoals	Offshore South	Subtidal	8,488	2,676	11,164
RS1_2_422	RS1	Rowley Shoals	Offshore South	Subtidal	4,692	2,201	6,893

RS1_M11_820	RS1	Rowley Shoals	Offshore South	Subtidal	6,528	4,624	11,152
RS1_M11_840	RS1	Rowley Shoals	Offshore South	Subtidal	4,284	2,006	6,290
RS1_M12_808	RS1	Rowley Shoals	Offshore South	Subtidal	8,127	3,242	11,369
RS1_M12_817	RS1	Rowley Shoals	Offshore South	Subtidal	9,921	0	9,921
RS1_S_314	RS1	Rowley Shoals	Offshore South	Subtidal	13,162	0	13,162
RS1_S_321	RS1	Rowley Shoals	Offshore South	Subtidal	4,263	4,512	8,775
RS2_2_256	RS2	Rowley Shoals	Offshore South	Subtidal	10,402	0	10,402
RS2_C11_769	RS2	Rowley Shoals	Offshore South	Subtidal	8,000	3,737	11,737
RS2_C11_784	RS2	Rowley Shoals	Offshore South	Subtidal	22,064	0	22,064
RS2_C13_704	RS2	Rowley Shoals	Offshore South	Subtidal	8,693	2,854	11,547
RS2_C13_706	RS2	Rowley Shoals	Offshore South	Subtidal	17,710	0	17,710
RS2_C13_721	RS2	Rowley Shoals	Offshore South	Subtidal	19,524	0	19,524
RS2_C20_283	RS2	Rowley Shoals	Offshore South	Subtidal	4,004	4,679	8,683
RS2_S_734	RS2	Rowley Shoals	Offshore South	Subtidal	12,496	0	12,496
RS2_S_737	RS2	Rowley Shoals	Offshore South	Subtidal	5,669	5,584	11,253
RS3_1_184	RS3	Rowley Shoals	Offshore South	Subtidal	8,032	5,406	13,438
RS3_1_185	RS3	Rowley Shoals	Offshore South	Subtidal	9,434	0	9,434
RS3_1_191	RS3	Rowley Shoals	Offshore South	Subtidal	7,351	4,262	11,613

RS3_1_207	RS3	Rowley Shoals	Offshore South	Subtidal	6,020	5,452	11,472
RS3_S_215	RS3	Rowley Shoals	Offshore South	Subtidal	4,199	5,556	9,755
RS3_S_232	RS3	Rowley Shoals	Offshore South	Subtidal	6,849	4,430	11,279
RS3_S_246	RS3	Rowley Shoals	Offshore South	Subtidal	4,018	4,944	8,962
RS3_S_250	RS3	Rowley Shoals	Offshore South	Subtidal	0	4,840	4,840

*Mislabelled Sample likely from Rowley Shoals. See Supplementary Information for details

Supplementary Table 3.2 Statistics of alignment results of 75 *A.digitifera* samples

Sample ID	Mapping rate (%)	Mean mapping depth (X)	Mean genome coverage (%)
AI_1_001	97.667	18.767037	79.90865
AI_1_008	97.5663	18.005929	80.63667
AI_1_021	96.8774	37.40879	81.76728
AI_1_022	97.1598	46.137165	82.77188
AI_1_023	96.9446	19.128358	80.04476
AI_1_025	95.2865	15.668055	79.01263
AI_2_036	97.1875	21.288786	80.74803
AI_2_041	97.392	20.74144	78.887
AI_2_043	97.37	24.126531	80.15771
AI_2_136	97.2602	18.113422	78.6608
AI_2_151	97.292	23.868113	80.06716
AI_3_047	97.1889	22.945807	80.73957
AI_3_060	97.2106	27.933526	81.02576
AI_3_063	96.9327	30.816544	81.58648
AI_3_071	96.5805	18.063423	80.21551
AR_125_374	96.2671	19.114326	80.62103
AR_125_377	96.0844	18.483311	79.6248
AR_125_385	96.5558	15.236707	80.06581
AR_125_388	95.6461	25.334181	80.79124
AR_125_392	96.7675	18.255773	80.25345
AR_128_316	97.2345	10.609246	77.99757
AR_128_318	97.451	17.165764	81.12785
AR_128_326	97.0968	18.33821	80.68386

AR_128_328	97.0342	14.094782	78.77346
AR_128_336	97.5705	18.14965	80.61123
AR_132_154	97.7245	18.60395	80.81611
AR_132_162	96.7145	17.248979	80.72761
AR_132_170	96.2782	17.384088	80.9814
AR_132_173	94.5964	16.708212	79.97423
AR_132_178	97.2215	17.912628	79.69592
AR_133_341	97.4625	19.138851	80.30004
AR_133_343	96.7202	18.10389	79.68526
AR_133_346	97.1972	17.834644	80.60629
AR_133_354	94.5335	10.808067	77.72408
AR_133_357	96.6271	19.709395	80.96677
BR_4_077	96.5332	12.637533	78.84721
BR_4_078	96.8139	14.83753	79.06365
BR_4_081	97.0307	14.944312	78.88322
BR_4_082	96.3477	16.997906	79.23306
BR_4_087	96.5735	19.245551	79.53166
BR_4_088	96.4051	11.264454	77.78714
BR_4_091	96.0758	50.772276	82.4786
BR_4_100	96.6591	16.847131	78.93905
BR_5_112	95.7586	21.113459	79.59467
BR_5_114	96.3534	31.152646	81.46212
BR_5_121	97.8185	15.956472	79.47596
BR_5_123	96.059	15.107586	78.30701
BR_5_124	96.9857	14.422176	78.33935

BR_5_129	97.0632	15.844296	78.37352
BR_5_133	96.6321	29.721228	81.54055
RS1_2_417	97.6006	18.011018	80.39919
RS1_2_422	94.5774	10.702747	77.46733
RS1_M11_820	95.8864	17.723679	79.39771
RS1_M11_840	97.8422	10.271225	77.54022
RS1_M12_808	97.3633	17.885207	79.35851
RS1_M12_817	97.6437	15.657466	78.60756
RS1_S_314	95.5671	20.778862	80.47854
RS1_S_321	97.5365	14.537167	78.73255
RS2_2_256	97.6359	17.009786	79.97507
RS2_C11_769	97.4031	19.080128	80.95296
RS2_C11_784	97.5892	34.729218	81.95152
RS2_C13_704	94.6454	18.113346	80.08855
RS2_C13_706	97.5623	28.235592	81.32329
RS2_C13_721	97.5496	30.457152	81.38908
RS2_C20_283	97.6171	13.909271	77.71487
RS2_S_734	97.0893	20.06806	79.84422
RS2_S_737	96.7284	18.13485	80.26239
RS3_1_184	97.1275	17.435551	80.32015
RS3_1_185	97.2257	15.329198	79.54294
RS3_1_191	97.6529	18.938273	79.88599
RS3_1_207	96.4095	18.944451	80.55368
RS3_S_215	96.9705	15.813856	79.36472
RS3_S_232	95.7579	17.828818	80.1138

RS3_S_246	92.4953	14.257731	78.43161
RS3_S_250	96.9054	8.038645	75.3388
Average	96.75	19.52	79.86

Supplementary Table 3.3 Pairwise F_{ST} between locations of *A.digitifera* and genetics statistics of three populations in northwestern Australia

	Sample size	Inshore (Fst)	North Offshore (Fst)	South Offshore (Fst)	Nucleotide diversity (Pi)	sd. Pi	Tajima's D	sd. Tajima's D	Unique SNP count
Inshore	29	-	0.0559	0.0583	0.0031	0.0024	-0.6625	0.7732	1,235,530
North Offshore	20	0.0224	-	0.0357	0.0032	0.0025	-1.0121	0.6935	1,425,395
South Offshore	25	0.0237	0.0069	-	0.0031	0.0024	-0.9802	0.7042	1,511,531
	74	-	-	-	0.0031	0.0242	-0.8990	0.7350	9,656,554

Mean Fst values are presented in the bottom left and standard deviations are shown in the top right of the matrix.

Supplementary Table 3.4 Genomic regions corresponding to putative selective sweeps inferred by extended haplotype homozygosity statistics

Sacffold	Start	End	Population	Fraction with z-score > 2	Significant in	Max iHS pos	Max iHS	Max xpehh pos	Max xpehh	Max xpnsl pos	Max xpnsl
BLFC01000007	450000	500001	North Offshore	0.255102	xpehh;xpnsl	485984	2.9222	486419	6.45251	485791	5.13275
BLFC01000008	1150000	1200001	Inshore	0.376238	ihs	1151860	2.96154	1150520	3.38807	1150520	3.41195
BLFC01000008	2050000	2100001	Inshore	0.387699	xpehh;xpnsl	2062023	3.47442	2065705	5.67936	2058926	5.87126
BLFC01000016	1050000	1100001	Inshore	0.212598	xpehh	1097707	1.42546	1099364	4.30497	1097707	3.22342
BLFC01000016	1250000	1350001	North Offshore	0.271111333	ihs;xpnsl	1289824	3.74958	1347301	4.86855	1347301	4.75458
BLFC01000016	3200000	3250001	North Offshore	0.377358	ihs	3217995	4.5374	3244203	2.47276	3245051	2.43118

BLFC01000039	100000	150001	Inshore	0.712259	xpehh;xpnsl	149931	2.68655	143662	7.71652	141029	5.68318
BLFC01000047	750000	800001	South Offshore	0.211045	xpehh;xpnsl	799313	1.47046	787775	4.20586	789212	3.96488
BLFC01000047	950000	1000001	North Offshore	0.365759	xpehh	992059	1.8728	991844	3.53235	990754	3.43012
BLFC01000047	1150000	1300001	North Offshore	0.750877143	xpnsl;ihs;xpehh	1235776	4.82621	1215552	5.59139	1206578	5.5565
BLFC01000051	400000	450001	South Offshore	0.176904	xpehh	406817	1.55161	414813	5.05569	416704	3.38618
BLFC01000051	600000	650001	Inshore	0.378151	ihs	606419	4.8272	620726	3.01434	620976	3.15953
BLFC01000051	2800000	2850001	North Offshore	0.541667	xpehh	2807752	0.835428	2823848	3.76386	2807752	2.87626
BLFC01000051	3050000	3100001	North Offshore	0.447368	xpehh;xpnsl	3088178	1.77114	3088034	4.75509	3091277	4.20267
BLFC01000051	3400000	3500001	North Offshore	0.263889	xpnsl;xpehh	3467077	2.02659	3463665	4.93851	3463665	4.9018
BLFC01000051	3550000	3600001	Inshore	0.117371	xpehh	3556299	1.22063	3555441	4.47151	3558277	3.44029
BLFC01000055	150000	250001	South Offshore	0.3259185	ihs;xpehh	152730	3.88979	221955	3.50319	241336	3.61245
BLFC01000056	950000	1000001	North Offshore	0.541855	xpehh;xpnsl	988793	1.80121	986554	3.92168	982428	4.4323
BLFC01000056	1050000	1100001	South Offshore	0.2782515	xpehh;xpnsl	1062999	1.53132	1079244	4.29211	1071582	3.97007
BLFC01000057	650000	700001	North Offshore	0.260901	xpehh;xpnsl	675906	2.28721	694903	3.81337	694903	4.30612
BLFC01000057	650000	700001	South Offshore	0.300929	xpnsl	672986	2.6359	687484	3.81246	687634	4.72187
BLFC01000074	100000	250001	Inshore	0.433003	ihs;xpehh;xp	228082	3.45896	134031	5.74859	130412	4.93609

					nsl						
BLFC01000074	1200000	1250001	South Offshore	0.258	xpehh	1204568	2.4704	1231872	4.4708	1231775	3.11235
BLFC01000082	550000	600001	Inshore	0.637795	ihs	586389	1.06654	573504	-0.144667	573504	-0.0622397
BLFC01000089	100000	200001	North Offshore	0.283555	xpehh;xpnsl	132329	2.32552	182867	6.0017	182867	6.12764
BLFC01000089	100000	200001	South Offshore	0.181481	xpnsl	133800	3.72487	160740	5.33381	159890	4.46437
BLFC01000100	800000	850001	North Offshore	0.664419	xpehh;xpnsl	815192	3.05109	816730	4.65824	815371	5.07494
BLFC01000100	1550000	1700001	North Offshore	0.438408	ihs	1692875	3.60842	1688459	3.575	1687772	3.08233
BLFC01000100	2200000	2250001	North Offshore	0.374359	xpehh	2218945	1.5817	2218899	4.66895	2217153	3.59818
BLFC01000100	3250000	3350001	South Offshore	0.39983275	ihs;xpehh	3256583	5.37928	3255914	5.41373	3315158	3.57557
BLFC01000106	450000	500001	South Offshore	0.138816	xpnsl	490683	3.07135	461759	3.8686	462428	4.58558
BLFC01000123	850000	900001	Inshore	0.457392	xpehh;xpnsl	865610	2.79269	865926	5.52992	873566	4.6371
BLFC01000123	1650000	1750001	South Offshore	0.361209	ihs	1687211	3.83822	1703307	3.72396	1717376	4.25608
BLFC01000124	650000	700001	South Offshore	0.456722	ihs	688072	4.95745	668198	2.60297	681715	3.00201
BLFC01000125	1600000	1650001	North Offshore	0.419821	ihs;xpehh;xpnsl	1606552	3.95282	1608481	3.96101	1604574	4.22175
BLFC01000125	2250000	2300001	North Offshore	0.243499	xpnsl	2280625	0.665246	2258881	3.85412	2258881	4.18321

BLFC01000137	350000	400001	North Offshore	0.428571	ihs	398739	-0.694641	384921	-0.326828	365527	0.141938
BLFC01000137	650000	850001	South Offshore	0.568234	ihs	839339	1.07665	819010	2.30764	848290	1.84379
BLFC01000148	600000	800001	South Offshore	0.305265857	xpehh;xpnsl;ihs	672565	3.32549	649622	5.34751	649622	5.30524
BLFC01000152	50000	100001	North Offshore	0.333333	ihs	62674	3.66933	61837	3.239	61979	3.53374
BLFC01000152	2400000	2450001	South Offshore	0.429652	ihs	2445286	4.14004	2420695	2.39323	2421047	2.59693
BLFC01000152	2450000	2500001	North Offshore	0.27027	xpnsl	2452017	2.34819	2451789	3.2204	2452017	3.89036
BLFC01000154	0	50001	South Offshore	0.271394	ihs	40458	3.42856	15062	3.63807	15062	3.97767
BLFC01000154	200000	250001	South Offshore	0.166018	xpnsl	218934	2.79022	210172	3.69583	210172	4.14854
BLFC01000154	250000	300001	Inshore	0.2135055	xpehh;xpnsl	275363	3.05018	288245	8.52351	287600	7.56992
BLFC01000158	700000	750001	Inshore	0.804511	ihs	735929	3.26064	736084	2.45478	720890	0.711187
BLFC01000161	700000	750001	Inshore	0.1579545	xpehh;xpnsl	735197	2.23937	733922	4.78477	733224	4.30954
BLFC01000166	1100000	1150001	Inshore	0.427083	xpehh	1146780	1.17354	1149745	4.76495	1138874	3.30022
BLFC01000172	1250000	1300001	Inshore	0.195065	xpnsl	1281540	3.13632	1285870	5.31027	1289973	5.47847
BLFC01000174	50000	100001	North Offshore	0.221925	xpehh	NA	NA	97865	3.21697	97865	3.01622
BLFC01000184	1050000	1100001	South Offshore	0.141218	xpnsl	1055414	1.86688	1078691	6.75342	1078694	6.87178

BLFC01000184	1600000	1650001	South Offshore	0.343348	ihs	1612581	4.17738	1612188	2.77975	1612364	3.39998
BLFC01000185	100000	300001	South Offshore	0.7074474	ihs;xpehh;xpnsl	195245	6.07138	244363	7.2475	234147	6.36994
BLFC01000185	900000	1050001	Inshore	0.407388333	xpehh;xpnsl	921726	4.67558	970783	7.54117	976133	7.16109
BLFC01000185	1150000	1200001	North Offshore	0.184375	xpnsl	1196569	1.95592	1183079	4.35389	1184206	3.99023
BLFC01000185	1550000	1600001	South Offshore	0.191781	xpehh	1599844	1.73014	1579603	3.99305	1579939	3.26355
BLFC01000201	150000	200001	North Offshore	0.273756	ihs	165771	3.3562	157733	1.81893	157733	1.98348
BLFC01000201	800000	850001	North Offshore	0.275488	ihs	822711	4.24267	822860	4.11203	822983	3.8522
BLFC01000201	1300000	1350001	South Offshore	0.30343	ihs	1300679	4.07306	1306341	2.83224	1306700	3.09001
BLFC01000201	2350000	2400001	South Offshore	0.235808	xpehh	2356090	1.38741	2357086	5.18438	2357899	2.77929
BLFC01000201	2450000	2500001	Inshore	0.5625	ihs	2496207	-0.218496	2465725	-1.33292	2458890	-1.09102
BLFC01000208	350000	400001	South Offshore	0.212993	ihs;xpehh	368318	3.10485	368907	4.48945	368907	4.38297
BLFC01000211	400000	450001	North Offshore	0.483516	ihs	440191	3.81843	437344	1.34023	437344	1.42985
BLFC01000211	400000	450001	South Offshore	0.279461	xpehh;xpnsl	439894	2.87341	440260	4.07964	440640	4.0841
BLFC01000211	650000	700001	North Offshore	0.344086	xpehh	699906	1.81461	698751	3.75343	698762	2.77235
BLFC01000235	3050000	3100001	North Offshore	0.306173	ihs	3092512	4.43787	3092607	3.77124	3092607	4.05552

BLFC01000235	3250000	3300001	Inshore	0.250296	ihs;xpnsl	3292500	3.44087	3267449	4.23103	3269793	4.28589
BLFC01000243	650000	700001	South Offshore	0.272727	xpehh;xpnsl	NA	NA	694851	3.81108	694851	4.00147
BLFC01000255	50000	300001	North Offshore	0.2945238	xpehh;xpnsl;ihs	290628	3.31835	50521	4.54118	126805	4.40546
BLFC01000255	450000	500001	South Offshore	0.497006	ihs	454321	0.39473	499867	2.80723	482643	2.8241
BLFC01000255	450000	600001	North Offshore	0.3780862	xpehh;xpnsl	531479	2.60591	460165	4.94422	527783	4.47521
BLFC01000256	400000	450001	North Offshore	0.197839	xpehh;xpnsl	429849	3.7501	408496	4.48318	414789	4.43736
BLFC01000256	1300000	1400001	South Offshore	0.54021	xpehh;xpnsl	NA	NA	1340234	4.26542	1338357	4.94323
BLFC01000265	200000	250001	Inshore	0.316338	xpehh;xpnsl	243988	2.76047	243955	6.85913	243305	5.64723
BLFC01000274	600000	650001	Inshore	0.199074	xpehh	606898	1.50525	610892	4.68778	611034	3.63221
BLFC01000277	150000	250001	Inshore	0.42405575	ihs;xpehh	162685	4.18741	182749	5.26146	233228	4.09552
BLFC01000277	1200000	1250001	Inshore	0.115049	xpnsl	1224134	3.22165	1224225	4.65349	1224225	4.93185
BLFC01000286	1950000	2000001	South Offshore	0.6537425	ihs;xpehh	1983407	2.1411	1980416	4.90958	1983265	4.43014
BLFC01000289	450000	500001	Inshore	0.502683	ihs	459833	3.38314	495066	4.19997	457677	3.61231
BLFC01000289	650000	700001	South Offshore	0.137755	xpnsl	656672	3.27542	656438	5.05119	656222	5.55124
BLFC01000298	200000	350001	Inshore	0.2556995	xpehh;xpnsl;ihs	323174	3.64882	276007	6.28043	276012	5.50613

BLFC01000298	550000	600001	North Offshore	0.269872	xpehh	563562	3.58033	583068	3.36411	579100	3.70553
BLFC01000298	1250000	1300001	North Offshore	0.4646895	ihs;xpehh	1259733	3.94498	1259813	4.32137	1259466	3.62625
BLFC01000298	1300000	1350001	South Offshore	0.484087	xpehh;xpnsl	1339669	4.19213	1340907	4.61273	1339491	5.0155
BLFC01000299	1950000	2000001	North Offshore	0.177283	xpnsl	1984765	3.04118	1964277	4.41676	1964277	5.16142
BLFC01000303	500000	550001	South Offshore	0.151644	xpnsl	500318	2.88282	537904	4.98364	537897	5.17207
BLFC01000309	400000	450001	North Offshore	0.581699	ihs	405596	1.90702	416660	3.20605	409195	3.15931
BLFC01000309	1800000	1850001	South Offshore	0.157895	xpnsl	1849931	2.04899	1849675	4.60565	1849683	5.39742
BLFC01000309	2150000	2200001	South Offshore	0.366959	xpehh;xpnsl	2171887	2.90382	2156503	4.43253	2198792	5.16289
BLFC01000309	2200000	2300001	Inshore	0.485879333	xpehh;xpnsl	2232049	3.14423	2232121	5.58234	2223799	5.53468
BLFC01000309	2950000	3100001	Inshore	0.359859667	xpehh;xpnsl;ihs	3082293	3.68533	2958900	6.18554	2979591	4.96952
BLFC01000310	200000	250001	North Offshore	0.692308	ihs	249797	0.379109	217391	1.6042	217391	1.80807
BLFC01000310	2100000	2150001	North Offshore	0.610577	ihs	2108853	2.14722	2116656	1.10515	2111612	1.44163
BLFC01000317	1000000	1050001	South Offshore	0.664179	xpehh;xpnsl	1023081	3.71682	1030537	7.80811	1020457	8.07608
BLFC01000324	350000	400001	North Offshore	0.360305	ihs	365621	3.51909	385187	3.77968	361654	3.48772
BLFC01000324	1200000	1400001	North Offshore	0.348308	xpehh;xpnsl	1286918	2.93002	1288100	5.8801	1286780	5.26683
BLFC01000324	1600000	1800001	Inshore	0.726401	ihs	1780814	3.49861	1708264	3.41471	1708264	2.73406

BLFC01000326	1550000	2150001	Inshore	0.607887571	ihs;xpehh;xpnsl	1582542	4.77814	1576014	10.0019	2062083	6.20647
BLFC01000341	950000	1000001	North Offshore	0.616188	xpehh;xpnsl	994021	4.18992	993849	5.82197	996061	4.62675
BLFC01000341	1500000	1550001	North Offshore	0.6898735	xpehh;xpnsl	1527448	4.20109	1530904	5.67713	1530907	5.47135
BLFC01000341	1750000	1800001	North Offshore	0.4	ihs	1778378	3.19529	1774559	2.62968	1774559	2.79603
BLFC01000348	1350000	1400001	North Offshore	0.23435	xpehh	1366037	3.4491	1353915	3.76859	1354246	4.13578
BLFC01000348	2050000	2150001	South Offshore	0.1921055	xpehh	2100431	2.7034	2100967	3.99516	2096247	2.92689
BLFC01000348	2400000	2450001	Inshore	0.528777	ihs	2438917	3.02893	2439187	2.69656	2447469	3.70918
BLFC01000368	800000	850001	Inshore	0.171182	xpnsl	836679	3.33585	828502	5.66153	829154	4.73148
BLFC01000375	100000	450001	South Offshore	0.716148235	xpehh;xpnsl;ihs	322331	6.4166	345929	7.02986	348028	6.79577
BLFC01000375	1150000	1250001	Inshore	0.231304	ihs;xpnsl	1175296	3.81487	1240943	4.92929	1207822	4.5262
BLFC01000393	1400000	1450001	Inshore	0.270959	xpehh;xpnsl	1434493	3.20752	1429927	6.41692	1429938	5.61381
BLFC01000393	1650000	1700001	North Offshore	0.3440085	xpehh;xpnsl	1669514	2.45685	1661048	5.11696	1656339	4.10532
BLFC01000393	1950000	2000001	South Offshore	0.269982	xpnsl	1988908	2.61434	1988869	4.39732	1988869	4.63423
BLFC01000404	1000000	1050001	Inshore	0.4215555	xpehh;xpnsl	1021451	3.29176	1006863	7.53204	1018567	5.66851
BLFC01000404	1250000	1300001	South Offshore	0.4583335	xpehh;xpnsl	1251980	3.9054	1265972	5.37037	1252410	5.1683

BLFC01000407	1800000	2050001	Inshore	0.29483075	xpehh;xpnsl	1982851	2.45675	2022077	4.76862	1920125	4.4626
BLFC01000407	1950000	2000001	South Offshore	0.3119895	ihs;xpnsl	1959096	3.36117	1974816	3.24776	1972809	4.45478
BLFC01000410	1050000	1100001	North Offshore	0.370062	ihs	1074411	4.02913	1088656	4.21978	1088657	4.73257
BLFC01000413	150000	200001	Inshore	0.332061	ihs	191337	3.18934	153512	4.39477	182520	5.13402
BLFC01000427	850000	900001	North Offshore	0.341615	ihs	871309	4.88865	890623	1.64889	885083	1.88585
BLFC01000439	1750000	1800001	Inshore	0.190418	xpehh;xpnsl	1782101	3.09458	1778121	6.6544	1781879	5.47662
BLFC01000439	2650000	2700001	North Offshore	0.3842195	ihs	2651208	3.82403	2681484	3.03687	2683656	3.06446
BLFC01000440	0	50001	North Offshore	0.316109	xpnsl	48594	0.885972	12811	2.14292	12448	3.54545
BLFC01000450	750000	800001	Inshore	0.356902	ihs	780571	4.00246	780537	3.97284	780537	4.36385
BLFC01000450	1550000	1600001	North Offshore	0.210708	xpehh	1555456	3.17822	1559262	3.83659	1559878	3.37236
BLFC01000451	750000	800001	North Offshore	0.384762	ihs	799728	3.65083	799763	4.32125	783028	3.61419
BLFC01000454	300000	400001	South Offshore	0.44689	ihs;xpehh;xpnsl	335383	4.05609	347677	5.07427	347677	4.96914
BLFC01000454	400000	450001	North Offshore	0.258427	xpehh;xpnsl	447741	3.80033	448423	5.45898	447493	4.25351
BLFC01000468	300000	400001	South Offshore	0.4326684	ihs;xpehh;xpnsl	348972	4.35188	352619	4.45588	304020	4.57003
BLFC01000480	800000	900001	South Offshore	0.241073333	xpehh;xpnsl	865666	2.5504	865666	5.44176	829452	5.23233

BLFC01000511	800000	850001	North Offshore	0.40528	ihs	842704	4.02222	817531	3.4843	817531	3.62734
BLFC01000511	1750000	1800001	North Offshore	0.2661	xpnsl	1781400	2.55979	1792138	5.94308	1792264	5.70655
BLFC01000522	350000	400001	Inshore	0.152866	xpnsl	398790	2.86414	360611	4.19177	357699	4.84042
BLFC01000522	1250000	1300001	Inshore	0.146452	xpnsl	1264724	2.57584	1270847	4.23247	1270692	5.02645
BLFC01000522	1650000	1750001	North Offshore	0.3431335	ihs;xpehh;xpnsl	1687504	4.89095	1706724	5.28105	1703531	4.83047
BLFC01000524	1050000	1250001	South Offshore	0.327208333	ihs;xpnsl	1227242	4.3606	1090735	3.89076	1074719	3.74598
BLFC01000536	200000	250001	North Offshore	0.2074585	xpehh;xpnsl	208127	2.59076	247924	5.19173	237565	5.49663
BLFC01000542	200000	300001	South Offshore	0.245811667	xpehh;xpnsl	205646	2.11922	260419	4.6446	207167	4.58087
BLFC01000557	300000	350001	North Offshore	0.276119	ihs	331688	4.02651	348333	2.53673	324277	2.59004
BLFC01000557	550000	600001	South Offshore	0.225452	xpehh	551751	2.00114	589317	4.90754	591820	4.12636
BLFC01000565	100000	150001	Inshore	0.253134	xpehh;xpnsl	107730	2.07586	148097	6.14604	148125	4.27051
BLFC01000565	300000	350001	South Offshore	0.151515	xpnsl	300176	1.19143	300030	2.01249	300328	3.37558
BLFC01000573	150000	200001	North Offshore	0.531469	ihs	191939	3.65615	174514	2.44623	174514	2.48401
BLFC01000583	50000	100001	South Offshore	0.532258	ihs	86256	4.1762	85157	4.06314	85157	4.26142
BLFC01000591	150000	200001	Inshore	0.625	ihs	152165	-0.364502	152165	1.08421	152749	0.768111

BLFC01000593	400000	450001	Inshore	0.417021	ihs	402815	4.07372	432730	3.53433	432753	2.71711
BLFC01000593	900000	950001	North Offshore	0.230769	xpnsl	918032	2.28491	919671	3.89028	937186	4.00939
BLFC01000596	550000	600001	South Offshore	0.4290955	xpehh;xpnsl	587983	3.90135	588532	6.7476	588104	7.08919
BLFC01000596	1400000	1450001	South Offshore	0.292264	xpehh;xpnsl	1448514	3.24296	1448546	4.68104	1448546	5.18493
BLFC01000596	3450000	3500001	South Offshore	0.4375	ihs	3490848	-0.476861	3491113	0.825963	3491113	0.599846
BLFC01000599	750000	800001	South Offshore	0.237716	ihs	754285	4.05232	753253	3.43917	753253	3.8605
BLFC01000600	400000	450001	South Offshore	0.577400333	ihs;xpehh;xpnsl	426075	4.48381	416162	5.49617	408076	4.71321
BLFC01000600	2250000	2300001	Inshore	0.313609	xpehh	2296020	2.87062	2268473	5.46134	2268866	4.08089
BLFC01000600	2900000	2950001	Inshore	0.161359	xpehh;xpnsl	2908924	2.58154	2933378	5.38865	2910961	4.92485
BLFC01000600	3650000	3700001	North Offshore	0.3987825	xpehh;xpnsl	3681477	3.82942	3689455	4.94101	3690450	5.49045
BLFC01000600	3650000	3700001	South Offshore	0.19102	xpnsl	3668892	2.41694	3689505	4.38796	3681585	5.02246
BLFC01000600	3700000	3800001	Inshore	0.63069525	xpehh;xpnsl	3745266	4.32864	3746482	7.58099	3746368	8.04479
BLFC01000610	900000	1100001	Inshore	0.46805	ihs	930967	3.64815	932661	2.95286	969143	3.23842
BLFC01000632	550000	600001	South Offshore	0.324192333	ihs;xpehh;xpnsl	585867	3.96543	594850	5.08634	594850	5.45417
BLFC01000632	1650000	1700001	North Offshore	0.441632	ihs;xpehh;xpnsl	1690891	4.20392	1692851	4.07223	1693193	4.69256

BLFC01000632	2550000	3150001	North Offshore	0.630428222	ihs;xpehh;xpnsl	2666594	4.53171	2766132	6.5529	2745812	6.25649
BLFC01000632	2950000	3000001	South Offshore	0.487619	xpehh;xpnsl	2953899	2.90961	2979390	5.95263	2979157	5.30259
BLFC01000639	1300000	1550001	Inshore	0.57258575	xpehh;xpnsl;ihs	1314102	4.69142	1313413	9.09364	1313410	6.80179
BLFC01000639	2100000	2150001	North Offshore	0.393939	ihs	2109437	4.24518	2104496	2.08834	2104496	2.40122
BLFC01000645	400000	450001	South Offshore	0.503442	xpehh;xpnsl	438105	3.3883	433651	6.74455	416385	6.46731
BLFC01000647	200000	250001	North Offshore	0.203591	xpehh;xpnsl	222994	2.62356	216809	5.15908	225705	4.80798
BLFC01000647	1450000	1500001	South Offshore	0.309729	xpehh;xpnsl	1487040	2.85304	1488151	6.43082	1484884	6.3634
BLFC01000647	2800000	2850001	North Offshore	0.315294	ihs	2801528	3.94515	2845494	2.85509	2845327	2.81782
BLFC01000647	3000000	3050001	South Offshore	0.358333	ihs	3019098	3.86544	3034506	2.83336	3009327	2.34418
BLFC01000653	2050000	2150001	Inshore	0.609653	ihs;xpehh;xpnsl	2139950	3.32769	2122914	7.79232	2134707	6.41791
BLFC01000655	200000	250001	Inshore	0.616541	ihs	227203	2.41192	206072	4.1239	206072	3.36382
BLFC01000660	700000	750001	South Offshore	0.478405	ihs	716842	4.04429	722705	2.85123	722705	2.88726
BLFC01000690	2600000	2850001	Inshore	0.393651667	ihs;xpehh;xpnsl	2804187	4.9162	2797440	6.75395	2792852	5.62148
BLFC01000690	3200000	3250001	Inshore	0.375	ihs	3229127	3.86506	3226927	5.1136	3229236	4.30668
BLFC01000692	0	50001	South Offshore	0.457143	ihs	37142	2.542	19756	3.33336	39646	2.69383

BLFC01000706	250000	500001	Inshore	0.428532125	xpehh;xpnsl;ihs	319734	1.30669	456899	5.6807	367564	4.47299
BLFC01000706	750000	1000001	Inshore	0.447594125	xpehh;xpnsl;ihs	906519	3.13826	905676	6.18	951652	5.29937
BLFC01000715	0	50001	Inshore	0.128755	xpehh	NA	NA	31253	4.12845	41730	3.71548
BLFC01000715	150000	200001	North Offshore	0.439446	ihs	195396	3.34687	186618	2.3984	151960	2.49257
BLFC01000718	1150000	1200001	North Offshore	0.44186	ihs	1191247	4.52965	1191484	2.32567	1194415	2.00251
BLFC01000718	1250000	1400001	Inshore	0.612255333	ihs	1354627	5.14527	1272275	4.60231	1262432	4.1672
BLFC01000726	800000	850001	North Offshore	0.517153	xpehh;xpnsl	830578	2.27447	823443	5.79018	830395	5.51176
BLFC01000729	0	150001	North Offshore	0.7487974	xpehh;xpnsl	NA	NA	63515	5.36156	63564	5.32173
BLFC01000729	150000	200001	Inshore	0.597765667	ihs;xpehh;xpnsl	195585	3.98814	174735	7.28515	190650	5.37354
BLFC01000730	900000	1000001	North Offshore	0.39099175	xpehh;xpnsl	998578	0.60738	950712	6.01023	954769	4.40188
BLFC01000732	1950000	2000001	Inshore	0.116501	xpehh;xpnsl	1974413	3.12729	1975613	5.41867	1974126	4.47239
BLFC01000734	350000	400001	Inshore	0.393651	ihs	387760	3.32911	354842	1.95482	363834	2.70633
BLFC01000734	800000	850001	South Offshore	0.307081667	ihs;xpehh;xpnsl	818792	3.71819	816971	5.43089	818617	4.33688
BLFC01000745	300000	400001	Inshore	0.437223	ihs	301932	3.43471	376570	3.80586	380360	3.58522
BLFC01000745	1250000	1300001	North Offshore	0.2774225	xpehh;xpnsl	1292579	3.5625	1293810	4.15836	1296204	4.41272

BLFC01000745	1350000	1450001	South Offshore	0.242210333	xpehh;xpnsl	1410532	4.65199	1397710	6.03626	1397710	6.66027
BLFC01000745	1600000	1650001	North Offshore	0.219701	xpehh;xpnsl	1602482	2.65348	1611042	5.33661	1611695	5.17966
BLFC01000756	600000	650001	South Offshore	0.422414	ihs	605716	3.40673	613473	2.12088	613365	2.48479
BLFC01000763	350000	400001	South Offshore	0.5	ihs	393319	2.52172	393198	2.65204	393588	1.18242
BLFC01000765	100000	300001	South Offshore	0.68207725	xpehh;xpnsl;ihs	260843	4.56762	245927	4.80572	280146	4.76781
BLFC01000766	900000	950001	South Offshore	0.160895	xpnsl	911678	1.59938	922591	3.28198	922592	4.58846
BLFC01000770	650000	700001	Inshore	0.384615	xpehh;xpnsl	667419	2.60077	669890	6.99494	669138	5.0953
BLFC01000770	850000	900001	Inshore	0.370463	ihs	874844	3.75985	873371	2.64995	885654	3.12568
BLFC01000770	1200000	1250001	Inshore	0.535471333	ihs;xpehh;xpnsl	1226445	3.13562	1224244	7.1799	1220200	7.69895
BLFC01000770	2800000	2900001	Inshore	0.3759082	ihs;xpehh;xpnsl	2889016	4.00873	2889755	6.45083	2815433	5.72164
BLFC01000770	3100000	3150001	Inshore	0.365079	ihs	3106316	4.14539	3102571	1.34513	3102571	1.80103
BLFC01000773	200000	250001	Inshore	0.115385	xpnsl	239839	2.35306	210324	4.40627	210355	4.44733
BLFC01000773	500000	550001	Inshore	0.2638625	xpehh;xpnsl	533499	2.20094	504862	5.21464	504862	4.13906
BLFC01000773	1700000	1750001	North Offshore	0.351923	ihs	1726920	3.83509	1718249	2.77975	1722866	3.42698
BLFC01000773	2100000	2250001	Inshore	0.260515167	xpehh;xpnsl	2129760	2.83286	2182234	6.40305	2130558	5.28582

BLFC01000774	1650000	1700001	South Offshore	0.250549	ihs	1663236	3.20923	1659497	3.277	1659497	3.92118
BLFC01000778	700000	800001	South Offshore	0.253742	xpehh;xpnsl	734553	2.72001	753711	4.14284	735231	3.73357
BLFC01000778	750000	800001	North Offshore	0.2469135	xpehh;xpnsl	750471	1.36658	764228	3.84589	765995	3.72396
BLFC01000816	400000	450001	Inshore	0.157343	xpehh;xpnsl	402132	3.45902	403474	5.44239	402285	6.01676
BLFC01000818	2100000	2150001	South Offshore	0.379377	xpehh	2134532	1.43798	2147329	4.31912	2147556	4.14656
BLFC01000820	1050000	1100001	Inshore	0.507538	ihs	1086673	3.59923	1057876	2.00743	1051012	1.99627
BLFC01000827	1250000	1300001	South Offshore	0.206161	xpehh	1258935	1.99368	1263260	4.06275	1261773	3.39039
BLFC01000829	450000	500001	South Offshore	0.144041	xpnsl	NA	NA	496452	4.32431	498648	5.74433
BLFC01000834	1150000	1200001	South Offshore	0.183502	xpnsl	1160239	2.20555	1176394	4.13243	1176066	4.89078
BLFC01000834	1200000	1250001	North Offshore	0.196829	xpehh;xpnsl	1233970	3.73843	1206477	4.09158	1209588	4.26273
BLFC01000834	3350000	3400001	Inshore	0.380719	ihs	3350192	3.50333	3351106	4.11796	3360212	3.22617
BLFC01000834	3550000	3600001	South Offshore	0.251001	ihs	3598939	4.23803	3574761	2.9998	3574761	3.56807
BLFC01000838	550000	600001	Inshore	0.617647	ihs	552502	-0.140772	550234	1.72985	587308	1.54485
BLFC01000838	1750000	1950001	North Offshore	0.433622429	xpnsl;xpehh	1909343	2.22538	1834551	3.97819	1772970	4.37986
BLFC01000846	6200000	6250001	Inshore	0.11169	xpnsl	6212824	2.93805	6211714	4.02555	6210244	4.06198
BLFC01000847	1200000	1350001	North Offshore	0.44201175	xpehh;xpnsl	1212897	4.21653	1346965	4.74441	1211187	4.79447

BLFC01000850	1200000	1350001	South Offshore	0.2869092	xpehh;xpnsl	1346448	1.80067	1246900	4.41474	1238088	4.37861
BLFC01000857	0	100001	North Offshore	0.61031575	xpehh;xpnsl	40776	3.82486	42058	5.10738	15088	5.28861
BLFC01000857	0	150001	South Offshore	0.534269	ihs;xpehh;xpnsl	116600	4.80004	100598	4.44045	124352	4.82474
BLFC01000857	650000	700001	South Offshore	0.409722	ihs	694331	3.44061	692428	3.57331	688785	3.62432
BLFC01000877	650000	700001	South Offshore	0.16833	xpnsl	691299	3.66229	684308	3.51196	685205	4.22496
BLFC01000889	250000	300001	North Offshore	0.333333	xpehh	251439	0.79986	252189	4.0166	252809	3.22085
BLFC01000909	50000	100001	North Offshore	0.315789	ihs	99061	2.20358	97342	3.28092	97342	3.90422
BLFC01000927	850000	900001	North Offshore	0.2371795	xpehh;xpnsl	894119	0.491962	868086	3.42451	862424	3.41805
BLFC01000928	550000	600001	Inshore	0.6355935	xpehh;xpnsl	NA	NA	560179	8.10162	566045	5.05025
BLFC01000929	100000	150001	South Offshore	0.166667	xpnsl	104355	0.89123	104514	3.48166	104892	3.30918
BLFC01000930	500000	550001	North Offshore	0.17801	xpnsl	505081	0.950698	505179	3.79061	505179	3.71702
BLFC01000954	750000	800001	South Offshore	0.2166665	xpehh;xpnsl	770155	2.08234	755113	4.75934	754264	4.43396
BLFC01000954	1100000	1150001	South Offshore	0.477612	ihs	1121823	4.48059	1117444	3.65213	1117769	3.52993

Supplementary Table 3.5 List of genes overlapping with selective sweeps identified by extended haplotype homozygosity statistics

Scaffold id	Start	End	Population	Gene ID	Uniprot id
BLFC01000008	1150000	1200001	inshore	adig_s0073.g90	TTN1_CAEEL
BLFC01000008	2050000	2100001	inshore	adig_s0073.g153	CPP1_ACRMI
BLFC01000008	2050000	2100001	inshore	adig_s0073.g154	BARH1_HUMAN
BLFC01000008	2050000	2100001	inshore	adig_s0073.g155	DLX4A_DANRE
BLFC01000008	2050000	2100001	inshore	adig_s0073.g156	MSHA_STRAW
BLFC01000008	2050000	2100001	inshore	adig_s0073.g158	GOGA4_MOUSE
BLFC01000016	1050000	1100001	inshore	adig_s0011.g61	SLN13_HUMAN
BLFC01000016	1050000	1100001	inshore	adig_s0011.g62	SLN14_RABIT
BLFC01000016	1050000	1100001	inshore	adig_s0011.g63	MLC2_DROME
BLFC01000039	100000	150001	inshore	adig_s0146.g9	SEPT8_XENTR
BLFC01000039	100000	150001	inshore	adig_s0146.g10	LRBA_MOUSE
BLFC01000051	600000	650001	inshore	adig_s0009.g24	S35F5_MOUSE
BLFC01000051	600000	650001	inshore	adig_s0009.g25	ENPP5_RAT
BLFC01000051	600000	650001	inshore	adig_s0009.g26	Y1946_NEIG1
BLFC01000051	3550000	3600001	inshore	adig_s0009.g174	VIT_BOMMO
BLFC01000074	100000	250001	inshore	adig_s0057.g5	HUTU_MOUSE
BLFC01000074	100000	250001	inshore	adig_s0057.g6	HUTH_BOVIN
BLFC01000074	100000	250001	inshore	adig_s0057.g7	HUTH_BOVIN
BLFC01000074	100000	250001	inshore	adig_s0057.g8	CATC_HUMAN
BLFC01000074	100000	250001	inshore	adig_s0057.g9	ODPB_MOUSE
BLFC01000074	100000	250001	inshore	adig_s0057.g10	ARF4_XENLA
BLFC01000074	100000	250001	inshore	adig_s0057.g11	ROST_DROME
BLFC01000074	100000	250001	inshore	adig_s0057.g12	KDPA_POLAQ

BLFC01000074	100000	250001	inshore	adig_s0057.g13	KLKB1_BOVIN
BLFC01000074	100000	250001	inshore	adig_s0057.g14	ABLM1_MOUSE
BLFC01000074	100000	250001	inshore	adig_s0057.g15	Y1623_METJA
BLFC01000074	100000	250001	inshore	adig_s0057.g16	NAGK_MOUSE
BLFC01000074	100000	250001	inshore	adig_s0057.g17	MOXD1_MOUSE
BLFC01000082	550000	600001	inshore	adig_s0168.g24	AGRA2_DANRE
BLFC01000123	850000	900001	inshore	adig_s0039.g83	ZN519_HUMAN
BLFC01000123	850000	900001	inshore	adig_s0039.g84	QCR7_FASHE
BLFC01000123	850000	900001	inshore	adig_s0039.g85	RAB38_MOUSE
BLFC01000123	850000	900001	inshore	adig_s0039.g86	GDF5_HUMAN
BLFC01000154	250000	300001	inshore	adig_s0150.g21	PXDN_XENTR
BLFC01000154	250000	300001	inshore	adig_s0150.g22	RPOC2_WHEAT
BLFC01000154	250000	300001	inshore	adig_s0150.g23	PERC_DROME
BLFC01000154	250000	300001	inshore	adig_s0150.g24	PXDN_HUMAN
BLFC01000154	250000	300001	inshore	adig_s0150.g25	PXDN_XENTR
BLFC01000158	700000	750001	inshore	adig_s0103.g26	BLM_XENLA
BLFC01000161	700000	750001	inshore	adig_s0021.g39	JMJD6_CAEBR
BLFC01000166	1100000	1150001	inshore	adig_s0026.g37	YCCS_ECOLI
BLFC01000166	1100000	1150001	inshore	adig_s0026.g40	YCCS_ECOLI
BLFC01000166	1100000	1150001	inshore	adig_s0026.g41	PI3R6_MOUSE
BLFC01000166	1100000	1150001	inshore	adig_s0026.g42	ATPB_GEOSE
BLFC01000172	1250000	1300001	inshore	adig_s0078.g98	EBP2_MOUSE
BLFC01000172	1250000	1300001	inshore	adig_s0078.g99	T38B1_CAEEL
BLFC01000172	1250000	1300001	inshore	adig_s0078.g100	FMT_SYMTH
BLFC01000172	1250000	1300001	inshore	adig_s0078.g101	MB212_DANRE

BLFC01000172	1250000	1300001	inshore	adig_s0078.g102	MB21L_DROPS
BLFC01000172	1250000	1300001	inshore	adig_s0078.g103	PURT_DESAG
BLFC01000172	1250000	1300001	inshore	adig_s0078.g104	POLG_ILHV
BLFC01000172	1250000	1300001	inshore	adig_s0078.g105	ECH1_MOUSE
BLFC01000185	900000	1050001	inshore	adig_s0062.g71	S27A4_MACFA
BLFC01000185	900000	1050001	inshore	adig_s0062.g74	DAXX_CANLF
BLFC01000185	900000	1050001	inshore	adig_s0062.g75	WDR47_MOUSE
BLFC01000185	900000	1050001	inshore	adig_s0062.g76	MAGI3_MOUSE
BLFC01000185	900000	1050001	inshore	adig_s0062.g77	STX6_MOUSE
BLFC01000185	900000	1050001	inshore	adig_s0062.g78	HMA2_ORYSJ
BLFC01000185	900000	1050001	inshore	adig_s0062.g79	Y0701_DICDI
BLFC01000201	2450000	2500001	inshore	adig_s0044.g131	RL22_CHLMU
BLFC01000235	3250000	3300001	inshore	NA	NA
BLFC01000265	200000	250001	inshore	adig_s0125.g19	IMPA1_BOVIN
BLFC01000265	200000	250001	inshore	adig_s0125.g21	Y2179_DICDI
BLFC01000265	200000	250001	inshore	adig_s0125.g23	CO6A5_HUMAN
BLFC01000265	200000	250001	inshore	adig_s0125.g24	CP087_DANRE
BLFC01000274	600000	650001	inshore	adig_s0110.g23	RL19_FRATT
BLFC01000274	600000	650001	inshore	adig_s0110.g24	TF29_SCHPO
BLFC01000277	150000	250001	inshore	adig_s0082.g15	THOC2_RHIFE
BLFC01000277	150000	250001	inshore	adig_s0082.g17	RIT1_HUMAN
BLFC01000277	150000	250001	inshore	adig_s0082.g19	SFR1_HUMAN
BLFC01000277	150000	250001	inshore	adig_s0082.g20	PCSK5_BRACL
BLFC01000277	1200000	1250001	inshore	adig_s0082.g82	NOSIP_DANRE
BLFC01000277	1200000	1250001	inshore	adig_s0082.g83	TRPA1_RAT

BLFC01000277	1200000	1250001	inshore	adig_s0082.g84	TRPA1_DROME
BLFC01000289	450000	500001	inshore	adig_s0153.g18	BMPER_MOUSE
BLFC01000289	450000	500001	inshore	adig_s0153.g19	MOODY_DROPS
BLFC01000289	450000	500001	inshore	adig_s0153.g20	EV162_ASFM2
BLFC01000298	200000	350001	inshore	adig_s0120.g13	MTNN_DESPS
BLFC01000298	200000	350001	inshore	adig_s0120.g14	HSP12_CAEEL
BLFC01000298	200000	350001	inshore	adig_s0120.g15	HSP12_CAEEL
BLFC01000298	200000	350001	inshore	adig_s0120.g16	HSP12_CAEEL
BLFC01000298	200000	350001	inshore	adig_s0120.g17	HSP12_CAEEL
BLFC01000298	200000	350001	inshore	adig_s0120.g18	GCP60_HUMAN
BLFC01000298	200000	350001	inshore	adig_s0120.g19	TREA_APIME
BLFC01000298	200000	350001	inshore	adig_s0120.g20	RFCL_HALWD
BLFC01000298	200000	350001	inshore	adig_s0120.g21	GELA_DICDI
BLFC01000298	200000	350001	inshore	adig_s0120.g22	GRLN_DICDI
BLFC01000298	200000	350001	inshore	adig_s0120.g23	AP4S_ARATH
BLFC01000298	200000	350001	inshore	adig_s0120.g24	CKB2_CAEEL
BLFC01000298	200000	350001	inshore	adig_s0120.g25	OFT30_ARATH
BLFC01000298	200000	350001	inshore	adig_s0120.g26	PRS23_MOUSE
BLFC01000298	200000	350001	inshore	adig_s0120.g28	KI16B_HUMAN
BLFC01000309	2200000	2300001	inshore	adig_s0069.g143	SAT_SYNFM
BLFC01000309	2200000	2300001	inshore	adig_s0069.g144	RLMN_FRAAA
BLFC01000309	2200000	2300001	inshore	adig_s0069.g146	YOR2_AZOVI
BLFC01000309	2950000	3100001	inshore	adig_s0069.g181	GELS_HOMAM
BLFC01000309	2950000	3100001	inshore	adig_s0069.g182	TM9S3_HUMAN
BLFC01000309	2950000	3100001	inshore	adig_s0069.g184	PAX2_MOUSE

BLFC01000309	2950000	3100001	inshore	adig_s0069.g185	DMX1B_DANRE
BLFC01000309	2950000	3100001	inshore	adig_s0069.g186	DMX1B_DANRE
BLFC01000309	2950000	3100001	inshore	adig_s0069.g187	DMX1B_DANRE
BLFC01000309	2950000	3100001	inshore	adig_s0069.g188	DMX1B_DANRE
BLFC01000309	2950000	3100001	inshore	adig_s0069.g189	DMX1B_DANRE
BLFC01000309	2950000	3100001	inshore	adig_s0069.g190	PITX2_HUMAN
BLFC01000324	1600000	1800001	inshore	adig_s0041.g84	AAGAB_HUMAN
BLFC01000324	1600000	1800001	inshore	adig_s0041.g86	PXC2_ARATH
BLFC01000324	1600000	1800001	inshore	adig_s0041.g87	B2CL1_HUMAN
BLFC01000324	1600000	1800001	inshore	adig_s0041.g88	DRB1_ORYSJ
BLFC01000324	1600000	1800001	inshore	adig_s0041.g89	ULK3_CHICK
BLFC01000326	1550000	2150001	inshore	adig_s0038.g80	ODR4_CHICK
BLFC01000326	1550000	2150001	inshore	adig_s0038.g81	NPTX2_RAT
BLFC01000326	1550000	2150001	inshore	adig_s0038.g82	USF_AQUPY
BLFC01000326	1550000	2150001	inshore	adig_s0038.g83	HELS_METJA
BLFC01000326	1550000	2150001	inshore	adig_s0038.g84	CASP8_RAT
BLFC01000326	1550000	2150001	inshore	adig_s0038.g85	DPOL_PPV01
BLFC01000326	1550000	2150001	inshore	adig_s0038.g86	FMT_UNCTG
BLFC01000326	1550000	2150001	inshore	adig_s0038.g87	CAPSD_GMDNV
BLFC01000326	1550000	2150001	inshore	adig_s0038.g88	LIPB_EHRCR
BLFC01000326	1550000	2150001	inshore	adig_s0038.g89	ABCAD_MOUSE
BLFC01000326	1550000	2150001	inshore	adig_s0038.g90	DNHD1_HUMAN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g91	LIPB_EHRCR
BLFC01000326	1550000	2150001	inshore	adig_s0038.g92	SHE10_VANPO
BLFC01000326	1550000	2150001	inshore	adig_s0038.g93	CYB_EPICE

BLFC01000326	1550000	2150001	inshore	adig_s0038.g94	PGES2_BOVIN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g95	DNJA1_PONAB
BLFC01000326	1550000	2150001	inshore	adig_s0038.g96	HM20A_MOUSE
BLFC01000326	1550000	2150001	inshore	adig_s0038.g97	ZMYM3_HUMAN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g98	NPFF2_MOUSE
BLFC01000326	1550000	2150001	inshore	adig_s0038.g99	UGTK5_MANES
BLFC01000326	1550000	2150001	inshore	adig_s0038.g100	QOR_HUMAN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g101	FRAS1_HUMAN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g102	IRAK4_HUMAN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g103	TCP11_HUMAN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g104	RPA2_EUPOC
BLFC01000326	1550000	2150001	inshore	adig_s0038.g105	OIT3_RAT
BLFC01000326	1550000	2150001	inshore	adig_s0038.g106	OIT3_BOVIN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g107	TLL2_HUMAN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g108	OIT3_BOVIN
BLFC01000326	1550000	2150001	inshore	adig_s0038.g109	TLL2_HUMAN
BLFC01000348	2400000	2450001	inshore	adig_s0001.g105	RSPRY_MACFA
BLFC01000348	2400000	2450001	inshore	adig_s0001.g106	TAF4B_MOUSE
BLFC01000348	2400000	2450001	inshore	adig_s0001.g107	SL9A8_HUMAN
BLFC01000348	2400000	2450001	inshore	adig_s0001.g108	LRP1_MOUSE
BLFC01000368	800000	850001	inshore	adig_s0025.g56	DNAK_NITMS
BLFC01000368	800000	850001	inshore	adig_s0025.g57	XRN2_MOUSE
BLFC01000368	800000	850001	inshore	adig_s0025.g58	NICN1_MOUSE
BLFC01000368	800000	850001	inshore	adig_s0025.g59	HNK2_XENLA
BLFC01000375	1150000	1250001	inshore	adig_s0052.g69	SKAP2_TAKRU

BLFC01000375	1150000	1250001	inshore	adig_s0052.g70	DXR_CORGL
BLFC01000375	1150000	1250001	inshore	adig_s0052.g71	ANO5_HUMAN
BLFC01000375	1150000	1250001	inshore	adig_s0052.g72	RL22_ASHYP
BLFC01000375	1150000	1250001	inshore	adig_s0052.g73	ANO4_MOUSE
BLFC01000375	1150000	1250001	inshore	adig_s0052.g74	ANO1_HUMAN
BLFC01000375	1150000	1250001	inshore	adig_s0052.g75	ANO1_HUMAN
BLFC01000375	1150000	1250001	inshore	adig_s0052.g76	ANO4_MOUSE
BLFC01000375	1150000	1250001	inshore	adig_s0052.g77	HLYA_PROMI
BLFC01000375	1150000	1250001	inshore	adig_s0052.g78	Y869_STAS1
BLFC01000393	1400000	1450001	inshore	adig_s0102.g96	NOTC2_RAT
BLFC01000393	1400000	1450001	inshore	adig_s0102.g97	NEST_MESAU
BLFC01000393	1400000	1450001	inshore	adig_s0102.g98	NEST_MESAU
BLFC01000393	1400000	1450001	inshore	adig_s0102.g99	Y3712_PECCP
BLFC01000393	1400000	1450001	inshore	adig_s0102.g100	BRSK2_HUMAN
BLFC01000404	1000000	1050001	inshore	adig_s0020.g95	DRS1_ASPTN
BLFC01000404	1000000	1050001	inshore	adig_s0020.g96	TPRGL_BOVIN
BLFC01000404	1000000	1050001	inshore	adig_s0020.g97	SIP11_MAIZE
BLFC01000404	1000000	1050001	inshore	adig_s0020.g98	OPN4_PODSI
BLFC01000404	1000000	1050001	inshore	adig_s0020.g99	NU188_MOUSE
BLFC01000407	1800000	2050001	inshore	adig_s0145.g113	RAD50_ARCFU
BLFC01000407	1800000	2050001	inshore	adig_s0145.g114	POL2_BBWV2
BLFC01000407	1800000	2050001	inshore	adig_s0145.g115	TTN1_CAEEL
BLFC01000407	1800000	2050001	inshore	adig_s0145.g116	P52K_HUMAN
BLFC01000407	1800000	2050001	inshore	adig_s0145.g117	BLH11_ARATH
BLFC01000407	1800000	2050001	inshore	adig_s0145.g119	RNT_SHEB8

BLFC01000407	1800000	2050001	inshore	adig_s0145.g122	ALMA7_EMIHU
BLFC01000407	1800000	2050001	inshore	adig_s0145.g123	ALMA7_EMIHU
BLFC01000413	150000	200001	inshore	adig_s0114.g4	PTPRF_MOUSE
BLFC01000413	150000	200001	inshore	adig_s0114.g6	COHA1_MESAU
BLFC01000413	150000	200001	inshore	adig_s0114.g5	COOA1_HUMAN
BLFC01000413	150000	200001	inshore	adig_s0114.g7	SUOX_HUMAN
BLFC01000439	1750000	1800001	inshore	adig_s0042.g95	R3HD4_BOVIN
BLFC01000439	1750000	1800001	inshore	adig_s0042.g96	RBCC1_MOUSE
BLFC01000439	1750000	1800001	inshore	adig_s0042.g97	METK1_MOUSE
BLFC01000439	1750000	1800001	inshore	adig_s0042.g98	NALCN_HUMAN
BLFC01000439	1750000	1800001	inshore	adig_s0042.g99	CHST3_TETCF
BLFC01000450	750000	800001	inshore	adig_s0015.g55	K0319_HUMAN
BLFC01000450	750000	800001	inshore	adig_s0015.g56	CAC1A_APIME
BLFC01000450	750000	800001	inshore	adig_s0015.g57	CAC1A_HUMAN
BLFC01000522	350000	400001	inshore	adig_s0035.g27	OIT3_RAT
BLFC01000522	350000	400001	inshore	adig_s0035.g28	POLG_HCVSA
BLFC01000522	350000	400001	inshore	adig_s0035.g29	CQ098_HUMAN
BLFC01000522	350000	400001	inshore	adig_s0035.g30	GLI1_CHICK
BLFC01000522	350000	400001	inshore	adig_s0035.g31	PGCA_BOVIN
BLFC01000522	1250000	1300001	inshore	adig_s0035.g90	DOT1_DEBHA
BLFC01000522	1250000	1300001	inshore	adig_s0035.g91	S39AE_MOUSE
BLFC01000522	1250000	1300001	inshore	adig_s0035.g92	LUZP4_HUMAN
BLFC01000565	100000	150001	inshore	adig_s0198.g9	BMPR2_HUMAN
BLFC01000565	100000	150001	inshore	adig_s0198.g10	CDK2_CARAU
BLFC01000565	100000	150001	inshore	adig_s0198.g11	ORC4_PONAB

BLFC01000565	100000	150001	inshore	adig_s0198.g12	DMAW_CLAFS
BLFC01000565	100000	150001	inshore	adig_s0198.g13	CBH_CLOPE
BLFC01000591	150000	200001	inshore	NA	NA
BLFC01000593	400000	450001	inshore	adig_s0051.g26	DYDC1_HUMAN
BLFC01000593	400000	450001	inshore	adig_s0051.g27	TYRO_CHICK
BLFC01000600	2250000	2300001	inshore	adig_s0005.g225	ATS6_HUMAN
BLFC01000600	2250000	2300001	inshore	adig_s0005.g226	ULK3_CHICK
BLFC01000600	2250000	2300001	inshore	adig_s0005.g227	Y9963_DICDI
BLFC01000600	2900000	2950001	inshore	adig_s0005.g259	F234B_RAT
BLFC01000600	2900000	2950001	inshore	adig_s0005.g260	FBX3_BOVIN
BLFC01000600	2900000	2950001	inshore	adig_s0005.g261	DDX6_DICDI
BLFC01000600	2900000	2950001	inshore	adig_s0005.g262	ERR3_PONAB
BLFC01000600	2900000	2950001	inshore	adig_s0005.g263	BCAL2_ARATH
BLFC01000600	3700000	3800001	inshore	adig_s0005.g341	FRIS_LYMST
BLFC01000600	3700000	3800001	inshore	adig_s0005.g344	FRIS_LYMST
BLFC01000600	3700000	3800001	inshore	adig_s0005.g345	PEX19_RAT
BLFC01000600	3700000	3800001	inshore	adig_s0005.g346	PHS2_HUMAN
BLFC01000600	3700000	3800001	inshore	adig_s0005.g348	Y850_VIBCH
BLFC01000600	3700000	3800001	inshore	adig_s0005.g349	GCSPA_THEON
BLFC01000600	3700000	3800001	inshore	adig_s0005.g350	LRC58_XENLA
BLFC01000600	3700000	3800001	inshore	adig_s0005.g351	GABR2_RAT
BLFC01000600	3700000	3800001	inshore	adig_s0005.g352	CAD96_DROME
BLFC01000600	3700000	3800001	inshore	adig_s0005.g353	PGBM_MOUSE
BLFC01000600	3700000	3800001	inshore	adig_s0005.g354	RET_HUMAN
BLFC01000610	900000	1100001	inshore	adig_s0050.g52	CHS7_GIBZE

BLFC01000610	900000	1100001	inshore	adig_s0050.g53	TLL1_CHICK
BLFC01000610	900000	1100001	inshore	adig_s0050.g54	K0319_HUMAN
BLFC01000610	900000	1100001	inshore	adig_s0050.g55	ADRB2_MACMU
BLFC01000610	900000	1100001	inshore	adig_s0050.g56	HIS6_CHLPD
BLFC01000610	900000	1100001	inshore	adig_s0050.g59	SOX14_DANRE
BLFC01000610	900000	1100001	inshore	adig_s0050.g60	GALR2_RAT
BLFC01000610	900000	1100001	inshore	adig_s0050.g61	DOCK9_MOUSE
BLFC01000639	1300000	1550001	inshore	adig_s0017.g54	PXDN_MOUSE
BLFC01000639	1300000	1550001	inshore	adig_s0017.g55	BECN1_RAT
BLFC01000639	1300000	1550001	inshore	adig_s0017.g56	AP4B1_HUMAN
BLFC01000639	1300000	1550001	inshore	adig_s0017.g57	SYGB_RUTMC
BLFC01000639	1300000	1550001	inshore	adig_s0017.g58	RNK_DROME
BLFC01000639	1300000	1550001	inshore	adig_s0017.g59	RPOB_AETCO
BLFC01000639	1300000	1550001	inshore	adig_s0017.g60	MED27_ANOGA
BLFC01000639	1300000	1550001	inshore	adig_s0017.g61	XYLK_HUMAN
BLFC01000639	1300000	1550001	inshore	adig_s0017.g62	TNR16_CHICK
BLFC01000639	1300000	1550001	inshore	adig_s0017.g63	TNR14_HUMAN
BLFC01000639	1300000	1550001	inshore	adig_s0017.g64	TNR14_HUMAN
BLFC01000639	1300000	1550001	inshore	adig_s0017.g65	TAF1B_HUMAN
BLFC01000639	1300000	1550001	inshore	adig_s0017.g66	AGMR_PSEAE
BLFC01000639	1300000	1550001	inshore	adig_s0017.g67	TNR16_MOUSE
BLFC01000639	1300000	1550001	inshore	adig_s0017.g68	HSP7C_CAEBR
BLFC01000639	1300000	1550001	inshore	adig_s0017.g69	COP1_MOUSE
BLFC01000639	1300000	1550001	inshore	adig_s0017.g70	PI4KB_HUMAN
BLFC01000639	1300000	1550001	inshore	adig_s0017.g71	ATG26_ASPTN

BLFC01000653	2050000	2150001	inshore	adig_s0034.g140	INO80_MOUSE
BLFC01000653	2050000	2150001	inshore	adig_s0034.g142	CB042_MOUSE
BLFC01000653	2050000	2150001	inshore	adig_s0034.g143	TISB_RAT
BLFC01000653	2050000	2150001	inshore	adig_s0034.g144	RA51B_HUMAN
BLFC01000653	2050000	2150001	inshore	adig_s0034.g145	SPSB4_HUMAN
BLFC01000653	2050000	2150001	inshore	adig_s0034.g146	MTRR_RAT
BLFC01000653	2050000	2150001	inshore	adig_s0034.g147	MTMR2_DANRE
BLFC01000653	2050000	2150001	inshore	adig_s0034.g148	MSRE_MOUSE
BLFC01000655	200000	250001	inshore	adig_s0250.g15	AMGO2_MOUSE
BLFC01000690	2600000	2850001	inshore	adig_s0030.g199	NA
BLFC01000690	2600000	2850001	inshore	adig_s0030.g200	MICA1_DANRE
BLFC01000690	2600000	2850001	inshore	adig_s0030.g201	CF157_XENTR
BLFC01000690	2600000	2850001	inshore	adig_s0030.g202	TDIF2_HUMAN
BLFC01000690	2600000	2850001	inshore	adig_s0030.g204	COQ4_NEMVE
BLFC01000690	2600000	2850001	inshore	adig_s0030.g205	TBX5A_DANRE
BLFC01000690	2600000	2850001	inshore	adig_s0030.g207	LSM2_MOUSE
BLFC01000690	2600000	2850001	inshore	adig_s0030.g208	CSK2B_DANRE
BLFC01000690	2600000	2850001	inshore	adig_s0030.g210	PIN1_HUMAN
BLFC01000690	2600000	2850001	inshore	adig_s0030.g213	FBP1_STRPU
BLFC01000690	2600000	2850001	inshore	adig_s0030.g212	SYF1_RAT
BLFC01000690	2600000	2850001	inshore	adig_s0030.g215	CRB_DROME
BLFC01000690	2600000	2850001	inshore	adig_s0030.g216	Y1388_SYNY3
BLFC01000690	2600000	2850001	inshore	adig_s0030.g217	CMC_DICDI
BLFC01000690	2600000	2850001	inshore	adig_s0030.g219	RT33_MOUSE
BLFC01000690	2600000	2850001	inshore	adig_s0030.g221	EPI1_CAEEL

BLFC01000690	2600000	2850001	inshore	adig_s0030.g222	LAMA2_HUMAN
BLFC01000690	2600000	2850001	inshore	adig_s0030.g223	GFI1_RAT
BLFC01000690	2600000	2850001	inshore	adig_s0030.g224	ACADM_BOVIN
BLFC01000690	3200000	3250001	inshore	adig_s0030.g252	RABX5_MOUSE
BLFC01000690	3200000	3250001	inshore	adig_s0030.g253	RL27_BORBP
BLFC01000690	3200000	3250001	inshore	adig_s0030.g254	HSLU_CHLTE
BLFC01000706	250000	500001	inshore	adig_s0094.g8	FAT4_HUMAN
BLFC01000706	250000	500001	inshore	adig_s0094.g9	Y1015_HAEIN
BLFC01000706	250000	500001	inshore	adig_s0094.g10	DIM_PEA
BLFC01000706	750000	1000001	inshore	adig_s0094.g20	MABP1_HUMAN
BLFC01000706	750000	1000001	inshore	adig_s0094.g21	POGK_HUMAN
BLFC01000706	750000	1000001	inshore	adig_s0094.g22	SOX5_MOUSE
BLFC01000706	750000	1000001	inshore	adig_s0094.g23	SMBP2_MOUSE
BLFC01000706	750000	1000001	inshore	adig_s0094.g24	F25A4_MOUSE
BLFC01000706	750000	1000001	inshore	adig_s0094.g25	ADEC_ROSDO
BLFC01000706	750000	1000001	inshore	adig_s0094.g26	FCN2_HUMAN
BLFC01000715	0	50001	inshore	adig_s0231.g3	CTTB2_DASNO
BLFC01000715	0	50001	inshore	adig_s0231.g2	Y1197_DICDI
BLFC01000718	1250000	1400001	inshore	adig_s0112.g34	KCMA1_XENLA
BLFC01000718	1250000	1400001	inshore	adig_s0112.g35	LAMB4_DANRE
BLFC01000718	1250000	1400001	inshore	adig_s0112.g36	LRMDA_MOUSE
BLFC01000718	1250000	1400001	inshore	adig_s0112.g37	LRMDA_HUMAN
BLFC01000718	1250000	1400001	inshore	adig_s0112.g38	F241B_MOUSE
BLFC01000718	1250000	1400001	inshore	adig_s0112.g40	ZN503_DANRE
BLFC01000718	1250000	1400001	inshore	adig_s0112.g41	CP086_RAT

BLFC01000729	150000	200001	inshore	adig_s0124.g4	HEPHL_ACRMI
BLFC01000729	150000	200001	inshore	adig_s0124.g5	NIPA2_PONAB
BLFC01000729	150000	200001	inshore	adig_s0124.g6	LRC51_BOVIN
BLFC01000729	150000	200001	inshore	adig_s0124.g7	TM216_BOVIN
BLFC01000729	150000	200001	inshore	adig_s0124.g8	MPAS_CLAUC
BLFC01000732	1950000	2000001	inshore	adig_s0032.g129	C27C1_LITCT
BLFC01000732	1950000	2000001	inshore	adig_s0032.g130	NIPLA_DANRE
BLFC01000732	1950000	2000001	inshore	adig_s0032.g131	COME_METJA
BLFC01000732	1950000	2000001	inshore	adig_s0032.g132	TM241_XENLA
BLFC01000734	350000	400001	inshore	adig_s0081.g27	TDRD1_MOUSE
BLFC01000734	350000	400001	inshore	adig_s0081.g28	CAHD1_MOUSE
BLFC01000734	350000	400001	inshore	adig_s0081.g29	LACC1_HUMAN
BLFC01000734	350000	400001	inshore	adig_s0081.g31	LITAF_CHICK
BLFC01000734	350000	400001	inshore	adig_s0081.g32	RS12_ORENI
BLFC01000734	350000	400001	inshore	adig_s0081.g33	ALK_MOUSE
BLFC01000745	300000	400001	inshore	adig_s0118.g22	ECE1_BOVIN
BLFC01000745	300000	400001	inshore	adig_s0118.g23	PRUN1_BOVIN
BLFC01000745	300000	400001	inshore	adig_s0118.g24	LHPL2_DANRE
BLFC01000745	300000	400001	inshore	adig_s0118.g25	LID2_SCHPO
BLFC01000745	300000	400001	inshore	adig_s0118.g27	CYBP_PONAB
BLFC01000745	300000	400001	inshore	adig_s0118.g28	AP3S1_MOUSE
BLFC01000745	300000	400001	inshore	adig_s0118.g29	PRUN2_XENTR
BLFC01000745	300000	400001	inshore	adig_s0118.g30	ST17A_RABIT
BLFC01000745	300000	400001	inshore	adig_s0118.g31	DEPD7_RAT
BLFC01000745	300000	400001	inshore	adig_s0118.g32	UBP47_HUMAN

BLFC01000770	650000	700001	inshore	NA	NA
BLFC01000770	850000	900001	inshore	NA	NA
BLFC01000770	1200000	1250001	inshore	NA	NA
BLFC01000770	2800000	2900001	inshore	NA	NA
BLFC01000770	3100000	3150001	inshore	NA	NA
BLFC01000773	200000	250001	inshore	adig_s0046.g20	CERT_XENTR
BLFC01000773	200000	250001	inshore	adig_s0046.g22	CP17A_ICTPU
BLFC01000773	500000	550001	inshore	adig_s0046.g42	SIR_SYNE7
BLFC01000773	500000	550001	inshore	adig_s0046.g43	UB2Q1_HUMAN
BLFC01000773	500000	550001	inshore	adig_s0046.g44	TTC28_MOUSE
BLFC01000773	500000	550001	inshore	adig_s0046.g45	MUC4_MOUSE
BLFC01000773	500000	550001	inshore	adig_s0046.g46	PCDH8_HUMAN
BLFC01000773	2100000	2250001	inshore	adig_s0046.g154	GCY_STRPU
BLFC01000773	2100000	2250001	inshore	adig_s0046.g156	ANPRA_MOUSE
BLFC01000773	2100000	2250001	inshore	adig_s0046.g157	GC76C_DROME
BLFC01000773	2100000	2250001	inshore	adig_s0046.g158	GCY_STRPU
BLFC01000816	400000	450001	inshore	adig_s0180.g37	GDAP2_NEMVE
BLFC01000816	400000	450001	inshore	adig_s0180.g39	CFA47_HUMAN
BLFC01000816	400000	450001	inshore	adig_s0180.g40	CFA47_HUMAN
BLFC01000820	1050000	1100001	inshore	adig_s0065.g82	CPT1A_RAT
BLFC01000820	1050000	1100001	inshore	adig_s0065.g83	TRPC4_HUMAN
BLFC01000820	1050000	1100001	inshore	adig_s0065.g84	SETX_MOUSE
BLFC01000834	3350000	3400001	inshore	adig_s0028.g243	GLGB_CHLCV
BLFC01000834	3350000	3400001	inshore	adig_s0028.g244	CTNS_DICDI
BLFC01000834	3350000	3400001	inshore	adig_s0028.g245	DEOC_HUMAN

BLFC01000834	3350000	3400001	inshore	adig_s0028.g246	AMN1_PONAB
BLFC01000838	550000	600001	inshore	NA	NA
BLFC01000846	6200000	6250001	inshore	adig_s0002.g563	ITSN1_MOUSE
BLFC01000846	6200000	6250001	inshore	adig_s0002.g564	ITSN2_MOUSE
BLFC01000846	6200000	6250001	inshore	adig_s0002.g565	LTN1_CHICK
BLFC01000846	6200000	6250001	inshore	adig_s0002.g566	LTN1_MOUSE
BLFC01000846	6200000	6250001	inshore	adig_s0002.g567	MGT4C_CHICK
BLFC01000846	6200000	6250001	inshore	adig_s0002.g568	LIS1_XENTR
BLFC01000928	550000	600001	inshore	adig_s0036.g20	TB182_HUMAN
BLFC01000928	550000	600001	inshore	adig_s0036.g21	SSY22_ORYSJ
BLFC01000928	550000	600001	inshore	adig_s0036.g22	WDR35_RAT
BLFC01000007	450000	500001	northoffshore	adig_s0063.g11	ANK3_HUMAN
BLFC01000007	450000	500001	northoffshore	adig_s0063.g12	Y4891_SELML
BLFC01000007	450000	500001	northoffshore	adig_s0063.g14	NOT5_YEAST
BLFC01000007	450000	500001	northoffshore	adig_s0063.g13	NOT3_YEAST
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g83	LIP1_CAEEL
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g84	LIP1_CAEEL
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g85	LIP1_CAEEL
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g86	LIP1_CAEEL
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g87	RHO1_ASHGO
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g88	S15A2_DANRE
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g89	MPC1_RAT
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g90	COAE_XANAC
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g91	DZIP3_HUMAN
BLFC01000016	1250000	1350001	northoffshore	adig_s0011.g92	PTHD3_MOUSE

BLFC01000016	3200000	3250001	northoffshore	adig_s0011.g235	BRWD1_HUMAN
BLFC01000016	3200000	3250001	northoffshore	adig_s0011.g236	DHDDS_HUMAN
BLFC01000016	3200000	3250001	northoffshore	adig_s0011.g237	LIN41_DANRE
BLFC01000016	3200000	3250001	northoffshore	adig_s0011.g238	ICEF1_HUMAN
BLFC01000047	950000	1000001	northoffshore	adig_s0061.g32	SACS_HUMAN
BLFC01000047	950000	1000001	northoffshore	adig_s0061.g33	SACS_HUMAN
BLFC01000047	950000	1000001	northoffshore	adig_s0061.g34	FHDC1_MOUSE
BLFC01000047	950000	1000001	northoffshore	adig_s0061.g35	EX7L_SHEAM
BLFC01000047	950000	1000001	northoffshore	adig_s0061.g36	SACS_MOUSE
BLFC01000047	950000	1000001	northoffshore	adig_s0061.g37	SACS_HUMAN
BLFC01000047	1150000	1300001	northoffshore	adig_s0061.g51	SYDND_GLOVI
BLFC01000047	1150000	1300001	northoffshore	adig_s0061.g52	AA2AR_HORSE
BLFC01000047	1150000	1300001	northoffshore	adig_s0061.g53	TTC28_MOUSE
BLFC01000047	1150000	1300001	northoffshore	adig_s0061.g54	IOLA_LISIN
BLFC01000047	1150000	1300001	northoffshore	adig_s0061.g55	CAPP2_MESCR
BLFC01000047	1150000	1300001	northoffshore	adig_s0061.g56	MFSD6_MOUSE
BLFC01000051	2800000	2850001	northoffshore	adig_s0009.g158	PLI_CRODU
BLFC01000051	2800000	2850001	northoffshore	adig_s0009.g159	Y205_PASMU
BLFC01000051	2800000	2850001	northoffshore	adig_s0009.g160	KPSD5_ECOLX
BLFC01000051	3050000	3100001	northoffshore	adig_s0009.g162	ENTK_MOUSE
BLFC01000051	3400000	3500001	northoffshore	adig_s0009.g171	POL_FIVT2
BLFC01000051	3400000	3500001	northoffshore	adig_s0009.g172	NA
BLFC01000051	3400000	3500001	northoffshore	adig_s0009.g173	THAP1_SALSA
BLFC01000056	950000	1000001	northoffshore	adig_s0099.g28	POL4_DROME
BLFC01000056	950000	1000001	northoffshore	adig_s0099.g29	DUT_CANGA

BLFC01000056	950000	1000001	northoffshore	adig_s0099.g30	KCNJ1_HUMAN
BLFC01000056	950000	1000001	northoffshore	adig_s0099.g31	MD2BP_HUMAN
BLFC01000056	950000	1000001	northoffshore	adig_s0099.g32	COG2_HUMAN
BLFC01000057	650000	700001	northoffshore	adig_s0068.g34	CSR2B_HUMAN
BLFC01000057	650000	700001	northoffshore	adig_s0068.g35	ZAP70_MOUSE
BLFC01000057	650000	700001	northoffshore	adig_s0068.g36	GPSM2_MOUSE
BLFC01000057	650000	700001	northoffshore	adig_s0068.g37	SPAT1_HUMAN
BLFC01000057	650000	700001	northoffshore	adig_s0068.g38	NODB_BRAEL
BLFC01000057	650000	700001	northoffshore	adig_s0068.g39	RDH12_MOUSE
BLFC01000057	650000	700001	northoffshore	adig_s0068.g40	CLPP_MOUSE
BLFC01000089	100000	200001	northoffshore	NA	NA
BLFC01000100	800000	850001	northoffshore	adig_s0008.g33	ADRB2_MESAU
BLFC01000100	800000	850001	northoffshore	adig_s0008.g34	ADRB2_MESAU
BLFC01000100	800000	850001	northoffshore	adig_s0008.g35	ARRD2_MOUSE
BLFC01000100	800000	850001	northoffshore	adig_s0008.g36	RLP24_PONAB
BLFC01000100	1550000	1700001	northoffshore	adig_s0008.g75	YRD6_CAEEL
BLFC01000100	1550000	1700001	northoffshore	adig_s0008.g76	YRD6_CAEEL
BLFC01000100	1550000	1700001	northoffshore	adig_s0008.g77	RTJK_DROME
BLFC01000100	1550000	1700001	northoffshore	adig_s0008.g78	NPS6_GIBZE
BLFC01000100	1550000	1700001	northoffshore	adig_s0008.g79	UBC6_KLULA
BLFC01000100	1550000	1700001	northoffshore	adig_s0008.g81	MCTP1_MOUSE
BLFC01000100	2200000	2250001	northoffshore	adig_s0008.g101	TPPE_ARATH
BLFC01000100	2200000	2250001	northoffshore	adig_s0008.g102	YRD6_CAEEL
BLFC01000125	1600000	1650001	northoffshore	adig_s0010.g97	NCL_ARATH
BLFC01000125	1600000	1650001	northoffshore	adig_s0010.g98	ZIPA_PSEPF

BLFC01000125	2250000	2300001	northoffshore	adig_s0010.g121	YRD6_CAEEL
BLFC01000125	2250000	2300001	northoffshore	adig_s0010.g122	PZRN3_MOUSE
BLFC01000137	350000	400001	northoffshore	NA	NA
BLFC01000152	50000	100001	northoffshore	adig_s0014.g4	NALP4_HUMAN
BLFC01000152	50000	100001	northoffshore	adig_s0014.g5	LRC31_HUMAN
BLFC01000152	50000	100001	northoffshore	adig_s0014.g6	ZPP_ACRMI
BLFC01000152	2450000	2500001	northoffshore	adig_s0014.g149	PARL_BOVIN
BLFC01000152	2450000	2500001	northoffshore	adig_s0014.g150	RPP2D_ARATH
BLFC01000152	2450000	2500001	northoffshore	adig_s0014.g151	FA5V_PSETE
BLFC01000174	50000	100001	northoffshore	NA	NA
BLFC01000185	1150000	1200001	northoffshore	adig_s0062.g96	EFG_CORA7
BLFC01000185	1150000	1200001	northoffshore	adig_s0062.g97	TOIP1_RAT
BLFC01000185	1150000	1200001	northoffshore	adig_s0062.g98	TOR1A_MACFA
BLFC01000201	150000	200001	northoffshore	adig_s0044.g12	DDAH1_BOVIN
BLFC01000201	150000	200001	northoffshore	adig_s0044.g13	DUT_EHV1B
BLFC01000201	150000	200001	northoffshore	adig_s0044.g14	ADB4C_MELGA
BLFC01000201	150000	200001	northoffshore	adig_s0044.g15	CATB_PIG
BLFC01000201	800000	850001	northoffshore	adig_s0044.g45	IN80E_RAT
BLFC01000201	800000	850001	northoffshore	adig_s0044.g46	LAMA1_MOUSE
BLFC01000201	800000	850001	northoffshore	adig_s0044.g47	SPB1_PONAB
BLFC01000211	400000	450001	northoffshore	adig_s0090.g24	ERLEC_MOUSE
BLFC01000211	650000	700001	northoffshore	adig_s0090.g35	MSHA1_CATAD
BLFC01000235	3050000	3100001	northoffshore	adig_s0007.g118	MGDG_SOYBN
BLFC01000235	3050000	3100001	northoffshore	adig_s0007.g119	TESK2_HUMAN
BLFC01000235	3050000	3100001	northoffshore	adig_s0007.g120	PRDM6_MOUSE

BLFC01000235	3050000	3100001	northoffshore	adig_s0007.g121	ASND1_BOVIN
BLFC01000235	3050000	3100001	northoffshore	adig_s0007.g122	ASNA_NEMVE
BLFC01000255	50000	300001	northoffshore	adig_s0127.g3	PUR2_NOSS1
BLFC01000255	50000	300001	northoffshore	adig_s0127.g4	HARB1_MOUSE
BLFC01000255	50000	300001	northoffshore	adig_s0127.g5	DDL_HALOH
BLFC01000255	50000	300001	northoffshore	adig_s0127.g6	NARC_HALMT
BLFC01000255	50000	300001	northoffshore	adig_s0127.g7	XYLL3_ARATH
BLFC01000255	50000	300001	northoffshore	adig_s0127.g8	ACINU_MOUSE
BLFC01000255	50000	300001	northoffshore	adig_s0127.g9	Y581_AQUAE
BLFC01000255	50000	300001	northoffshore	adig_s0127.g10	OPN4B_GADMO
BLFC01000255	450000	600001	northoffshore	adig_s0127.g15	LACXP_LACLL
BLFC01000255	450000	600001	northoffshore	adig_s0127.g16	RL36B_YEAST
BLFC01000255	450000	600001	northoffshore	adig_s0127.g17	AGALC_ASPTN
BLFC01000255	450000	600001	northoffshore	adig_s0127.g18	ZN862_HUMAN
BLFC01000256	400000	450001	northoffshore	adig_s0174.g29	DMRT3_HUMAN
BLFC01000256	400000	450001	northoffshore	adig_s0174.g30	DMTA2_ORYLA
BLFC01000298	550000	600001	northoffshore	adig_s0120.g46	FGF1_NOTVI
BLFC01000298	550000	600001	northoffshore	adig_s0120.g48	ERH_AEDAE
BLFC01000298	550000	600001	northoffshore	adig_s0120.g49	CDKN3_HUMAN
BLFC01000298	550000	600001	northoffshore	adig_s0120.g50	HSP12_CAEEL
BLFC01000298	550000	600001	northoffshore	adig_s0120.g51	REBMT_LENAE
BLFC01000298	550000	600001	northoffshore	adig_s0120.g52	CRG1_YEAST
BLFC01000298	1250000	1300001	northoffshore	adig_s0120.g101	4CL1_TOBAC
BLFC01000298	1250000	1300001	northoffshore	adig_s0120.g102	TNFA_RAT
BLFC01000299	1950000	2000001	northoffshore	adig_s0031.g136	GPR63_HUMAN

BLFC01000299	1950000	2000001	northoffshore	adig_s0031.g138	NPY1R_CANLF
BLFC01000299	1950000	2000001	northoffshore	adig_s0031.g139	RH11_ARATH
BLFC01000299	1950000	2000001	northoffshore	adig_s0031.g140	AGRV1_DANRE
BLFC01000299	1950000	2000001	northoffshore	adig_s0031.g141	HAVR2_RAT
BLFC01000309	400000	450001	northoffshore	adig_s0069.g22	PLPL9_MOUSE
BLFC01000309	400000	450001	northoffshore	adig_s0069.g23	CSNC_ASPOZ
BLFC01000309	400000	450001	northoffshore	adig_s0069.g24	TENR_HUMAN
BLFC01000309	400000	450001	northoffshore	adig_s0069.g25	TENX_HUMAN
BLFC01000309	400000	450001	northoffshore	adig_s0069.g26	CSNC_ASPOZ
BLFC01000310	200000	250001	northoffshore	adig_s0027.g7	TENA_HUMAN
BLFC01000310	2100000	2150001	northoffshore	adig_s0027.g89	PGBD4_HUMAN
BLFC01000310	2100000	2150001	northoffshore	adig_s0027.g90	G2E3_CHICK
BLFC01000310	2100000	2150001	northoffshore	adig_s0027.g91	POK19_HUMAN
BLFC01000310	2100000	2150001	northoffshore	adig_s0027.g92	URM1_MOUSE
BLFC01000310	2100000	2150001	northoffshore	adig_s0027.g93	YG31B_YEAST
BLFC01000324	350000	400001	northoffshore	adig_s0041.g24	PRDM6_BOVIN
BLFC01000324	350000	400001	northoffshore	adig_s0041.g25	PRDM9_RAT
BLFC01000324	350000	400001	northoffshore	adig_s0041.g26	PRDM6_HUMAN
BLFC01000324	350000	400001	northoffshore	adig_s0041.g27	PRDM6_HUMAN
BLFC01000324	350000	400001	northoffshore	adig_s0041.g28	CE120_HUMAN
BLFC01000324	350000	400001	northoffshore	adig_s0041.g29	LAMA2_HUMAN
BLFC01000324	350000	400001	northoffshore	adig_s0041.g30	DRC8_MOUSE
BLFC01000324	1200000	1400001	northoffshore	adig_s0041.g75	FCER2_MOUSE
BLFC01000324	1200000	1400001	northoffshore	adig_s0041.g76	NA
BLFC01000324	1200000	1400001	northoffshore	adig_s0041.g79	COQ4_BOTFB

BLFC01000324	1200000	1400001	northoffshore	adig_s0041.g81	PPIB_CHICK
BLFC01000341	950000	1000001	northoffshore	adig_s0053.g57	NLRC3_HUMAN
BLFC01000341	950000	1000001	northoffshore	adig_s0053.g58	RAE1_MOUSE
BLFC01000341	950000	1000001	northoffshore	adig_s0053.g59	S2536_HUMAN
BLFC01000341	1500000	1550001	northoffshore	adig_s0053.g81	CPP1_ACRMI
BLFC01000341	1500000	1550001	northoffshore	adig_s0053.g83	UVRC_TRIEI
BLFC01000341	1750000	1800001	northoffshore	adig_s0053.g96	GBPH_PLAFB
BLFC01000341	1750000	1800001	northoffshore	adig_s0053.g98	COT2_MOUSE
BLFC01000348	1350000	1400001	northoffshore	adig_s0001.g61	MFRN2_DANRE
BLFC01000348	1350000	1400001	northoffshore	adig_s0001.g62	SUFU_MOUSE
BLFC01000348	1350000	1400001	northoffshore	adig_s0001.g63	BOD1_RAT
BLFC01000348	1350000	1400001	northoffshore	adig_s0001.g64	TMM53_DANRE
BLFC01000393	1650000	1700001	northoffshore	adig_s0102.g118	EFG1_BORBP
BLFC01000410	1050000	1100001	northoffshore	adig_s0163.g77	ARI3A_MOUSE
BLFC01000410	1050000	1100001	northoffshore	adig_s0163.g78	ARI1B_MOUSE
BLFC01000410	1050000	1100001	northoffshore	adig_s0163.g79	PKHD1_HUMAN
BLFC01000410	1050000	1100001	northoffshore	adig_s0163.g80	FGF1_MESAU
BLFC01000410	1050000	1100001	northoffshore	adig_s0163.g81	FGF1_NOTVI
BLFC01000410	1050000	1100001	northoffshore	adig_s0163.g82	FGF1_PIG
BLFC01000410	1050000	1100001	northoffshore	adig_s0163.g83	DCMC_HUMAN
BLFC01000427	850000	900001	northoffshore	adig_s0092.g60	MTR1A_SHEEP
BLFC01000427	850000	900001	northoffshore	adig_s0092.g61	MFAP1_BOVIN
BLFC01000427	850000	900001	northoffshore	adig_s0092.g62	GABR2_RAT
BLFC01000439	2650000	2700001	northoffshore	adig_s0042.g161	VPS10_PYRTT
BLFC01000439	2650000	2700001	northoffshore	adig_s0042.g162	KAD3_BOVIN

BLFC01000439	2650000	2700001	northoffshore	adig_s0042.g163	PUM3_RAT
BLFC01000439	2650000	2700001	northoffshore	adig_s0042.g164	ROG1_YEAST
BLFC01000440	0	50001	northoffshore	adig_s0256.g1	MLP_ACRMI
BLFC01000440	0	50001	northoffshore	adig_s0256.g2	MLP_ACRMI
BLFC01000440	0	50001	northoffshore	adig_s0256.g3	MATN4_HUMAN
BLFC01000450	1550000	1600001	northoffshore	adig_s0015.g114	SC5A6_MOUSE
BLFC01000450	1550000	1600001	northoffshore	adig_s0015.g115	SC5A8_DANRE
BLFC01000450	1550000	1600001	northoffshore	adig_s0015.g116	PEPT_CLONN
BLFC01000450	1550000	1600001	northoffshore	adig_s0015.g117	GCST_CORJK
BLFC01000450	1550000	1600001	northoffshore	adig_s0015.g118	SYV_CORGL
BLFC01000451	750000	800001	northoffshore	adig_s0054.g57	FLGI_PSESM
BLFC01000451	750000	800001	northoffshore	adig_s0054.g59	UVRB_CHLPD
BLFC01000451	750000	800001	northoffshore	adig_s0054.g60	GR101_LYMST
BLFC01000451	750000	800001	northoffshore	adig_s0054.g61	PROM1_MOUSE
BLFC01000454	400000	450001	northoffshore	adig_s0096.g10	PGCB_MOUSE
BLFC01000454	400000	450001	northoffshore	adig_s0096.g11	EYS_MACFA
BLFC01000511	800000	850001	northoffshore	adig_s0075.g73	MRP4_HUMAN
BLFC01000511	800000	850001	northoffshore	adig_s0075.g74	GELA_DICDI
BLFC01000511	800000	850001	northoffshore	adig_s0075.g75	DRD1L_OREMO
BLFC01000511	800000	850001	northoffshore	adig_s0075.g76	ADRB1_XENLA
BLFC01000511	800000	850001	northoffshore	adig_s0075.g77	FIBP_MOUSE
BLFC01000511	1750000	1800001	northoffshore	adig_s0075.g146	DMTA2_TAKRU
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g117	RYAR_TRICA
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g118	NLRC5_ICTPU
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g119	Y1101_SYNY3

BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g120	YQA3_THEAQ
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g121	CATIP_DANRE
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g122	DUS3L_XENLA
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g123	Y4110_SALAI
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g124	BTBD6_XENTR
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g125	RITA1_AILME
BLFC01000522	1650000	1750001	northoffshore	adig_s0035.g126	M3K1_HUMAN
BLFC01000536	200000	250001	northoffshore	adig_s0109.g13	COLA1_HUMAN
BLFC01000536	200000	250001	northoffshore	adig_s0109.g14	UROM_HUMAN
BLFC01000557	300000	350001	northoffshore	adig_s0152.g9	DEOB_DEIRA
BLFC01000557	300000	350001	northoffshore	adig_s0152.g10	DDGS_TRIV2
BLFC01000557	300000	350001	northoffshore	adig_s0152.g11	LIN41_RAT
BLFC01000573	150000	200001	northoffshore	adig_s0248.g9	FBX30_RAT
BLFC01000573	150000	200001	northoffshore	adig_s0248.g10	APH1A_MOUSE
BLFC01000573	150000	200001	northoffshore	adig_s0248.g11	S39AE_DANRE
BLFC01000593	900000	950001	northoffshore	NA	NA
BLFC01000600	3650000	3700001	northoffshore	adig_s0005.g338	SHE2_ASHGO
BLFC01000600	3650000	3700001	northoffshore	adig_s0005.g339	RET_HUMAN
BLFC01000600	3650000	3700001	northoffshore	adig_s0005.g340	ASPM_MACMU
BLFC01000600	3650000	3700001	northoffshore	adig_s0005.g342	FRIS_LYMST
BLFC01000600	3650000	3700001	northoffshore	adig_s0005.g341	FRIS_LYMST
BLFC01000600	3650000	3700001	northoffshore	adig_s0005.g343	FRIS_LYMST
BLFC01000632	1650000	1700001	northoffshore	adig_s0004.g115	TRPM7_RAT
BLFC01000632	1650000	1700001	northoffshore	adig_s0004.g116	ALG13_RAT
BLFC01000632	1650000	1700001	northoffshore	adig_s0004.g117	LAMC1_XENTR

BLFC01000632	1650000	1700001	northoffshore	adig_s0004.g118	TBC14_BOVIN
BLFC01000632	1650000	1700001	northoffshore	adig_s0004.g119	CRYD_XENLA
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g191	RBP2A_PLAF7
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g192	NA
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g193	SETD3_CHICK
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g194	NPFF2_HUMAN
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g195	S26A6_MOUSE
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g196	DISP2_DANRE
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g197	ZC3H1_HUMAN
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g198	NPFF2_RAT
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g199	QRFPR_BRAFL
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g200	NPFF2_MOUSE
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g201	KHK_RAT
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g202	NK2R_RABIT
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g203	AHNK2_HUMAN
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g204	PFF1_ARTBC
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g205	NK2R_RABIT
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g206	PFF1_TRIVH
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g207	CALM_TETPY
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g208	ENPP_APLCA
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g209	BEGIN_HUMAN
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g210	AT11B_HUMAN
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g211	CADN_ACRMI
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g212	NK2R_RABIT
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g213	NPFF2_HUMAN

BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g214	ACHA6_HUMAN
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g215	ACH10_CHICK
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g216	ACHA7_MACMU
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g217	GSKIP_DANRE
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g218	SQOR_MOUSE
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g219	CERS1_MOUSE
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g220	FOLR2_HUMAN
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g221	CALM_TETPY
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g222	SERIC_NEMVE
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g223	USOM5_ACRMI
BLFC01000632	2550000	3150001	northoffshore	adig_s0004.g224	ZAN_MOUSE
BLFC01000639	2100000	2150001	northoffshore	adig_s0017.g112	TRPC5_RABIT
BLFC01000639	2100000	2150001	northoffshore	adig_s0017.g113	TRPC5_RABIT
BLFC01000647	200000	250001	northoffshore	adig_s0012.g15	CTL2_DICDI
BLFC01000647	200000	250001	northoffshore	adig_s0012.g16	Y2453_MYCLE
BLFC01000647	200000	250001	northoffshore	adig_s0012.g17	SF3B2_HUMAN
BLFC01000647	200000	250001	northoffshore	adig_s0012.g18	CAN9_RAT
BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g219	TCPD_TAKRU
BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g220	DSBD_IDILO
BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g221	SGPP2_HUMAN
BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g222	COQ6_DANRE
BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g223	COMD1_MOUSE
BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g224	ARMT1_BOVIN
BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g225	NAA80_DROME
BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g226	SYE_CLOBL

BLFC01000647	2800000	2850001	northoffshore	adig_s0012.g227	ZN410_BOVIN
BLFC01000715	150000	200001	northoffshore	adig_s0231.g11	CP4V2_MOUSE
BLFC01000718	1150000	1200001	northoffshore	adig_s0112.g31	POL_SFV3L
BLFC01000718	1150000	1200001	northoffshore	adig_s0112.g32	H2AV_STRPU
BLFC01000726	800000	850001	northoffshore	adig_s0119.g22	PYR1_YEAST
BLFC01000726	800000	850001	northoffshore	adig_s0119.g23	MYOD_DROME
BLFC01000726	800000	850001	northoffshore	adig_s0119.g24	KT5AA_DANRE
BLFC01000726	800000	850001	northoffshore	adig_s0119.g25	XBP1_RAT
BLFC01000729	0	150001	northoffshore	adig_s0124.g1	BMP1_HUMAN
BLFC01000729	0	150001	northoffshore	adig_s0124.g2	TLL2_XENLA
BLFC01000729	0	150001	northoffshore	adig_s0124.g3	TLL1_DANRE
BLFC01000729	0	150001	northoffshore	adig_s0124.g4	HEPHL_ACRMI
BLFC01000730	900000	1000001	northoffshore	adig_s0077.g33	MRS2_NEUCR
BLFC01000730	900000	1000001	northoffshore	adig_s0077.g34	MY123_ACTCC
BLFC01000730	900000	1000001	northoffshore	adig_s0077.g35	CFDP2_BOVIN
BLFC01000730	900000	1000001	northoffshore	adig_s0077.g36	ASI1A_DANRE
BLFC01000730	900000	1000001	northoffshore	adig_s0077.g37	YJ9J_YEAST
BLFC01000745	1250000	1300001	northoffshore	adig_s0118.g109	ECE1_MOUSE
BLFC01000745	1600000	1650001	northoffshore	adig_s0118.g137	LSHR_BOVIN
BLFC01000745	1600000	1650001	northoffshore	adig_s0118.g138	EMRE_XENTR
BLFC01000773	1700000	1750001	northoffshore	adig_s0046.g121	RRN3_MOUSE
BLFC01000773	1700000	1750001	northoffshore	adig_s0046.g122	GLO2_MOUSE
BLFC01000773	1700000	1750001	northoffshore	adig_s0046.g123	UBP1_RAT
BLFC01000773	1700000	1750001	northoffshore	adig_s0046.g124	DPOD2_MOUSE
BLFC01000773	1700000	1750001	northoffshore	adig_s0046.g125	KLH20_RAT

BLFC01000773	1700000	1750001	northoffshore	adig_s0046.g126	ACRS2_ALTAL
BLFC01000778	750000	800001	northoffshore	adig_s0107.g47	POL_FFV
BLFC01000778	750000	800001	northoffshore	adig_s0107.g48	PPID_HAEIN
BLFC01000778	750000	800001	northoffshore	adig_s0107.g49	DNAE2_RHOP5
BLFC01000834	1200000	1250001	northoffshore	adig_s0028.g82	SCUB2_DANRE
BLFC01000834	1200000	1250001	northoffshore	adig_s0028.g83	PMPI_CHLTR
BLFC01000834	1200000	1250001	northoffshore	adig_s0028.g84	PA2_CONGI
BLFC01000834	1200000	1250001	northoffshore	adig_s0028.g85	LDLR_RABIT
BLFC01000834	1200000	1250001	northoffshore	adig_s0028.g86	Y1388_SYNY3
BLFC01000834	1200000	1250001	northoffshore	adig_s0028.g87	AA2AR_CANLF
BLFC01000838	1750000	1950001	northoffshore	adig_s0033.g50	VIAAT_XENLA
BLFC01000838	1750000	1950001	northoffshore	adig_s0033.g51	KLF5_HUMAN
BLFC01000838	1750000	1950001	northoffshore	adig_s0033.g52	FUT11_TAKRU
BLFC01000838	1750000	1950001	northoffshore	adig_s0033.g53	TCPR1_MOUSE
BLFC01000838	1750000	1950001	northoffshore	adig_s0033.g54	ZFP2_HUMAN
BLFC01000838	1750000	1950001	northoffshore	adig_s0033.g55	MATN2_HUMAN
BLFC01000838	1750000	1950001	northoffshore	adig_s0033.g56	UBX2A_MOUSE
BLFC01000847	1200000	1350001	northoffshore	adig_s0087.g29	GLBL2_MOUSE
BLFC01000847	1200000	1350001	northoffshore	adig_s0087.g30	ITA2_BOVIN
BLFC01000847	1200000	1350001	northoffshore	adig_s0087.g31	C71Z6_ORYSJ
BLFC01000847	1200000	1350001	northoffshore	adig_s0087.g32	ITA2_BOVIN
BLFC01000857	0	100001	northoffshore	adig_s0070.g1	TRFM_HUMAN
BLFC01000857	0	100001	northoffshore	adig_s0070.g2	TRFE_PAROL
BLFC01000857	0	100001	northoffshore	adig_s0070.g3	TRFM_RABIT
BLFC01000857	0	100001	northoffshore	adig_s0070.g4	TRFE_PAROL

BLFC01000857	0	100001	northoffshore	adig_s0070.g5	TRFE_ANAPL
BLFC01000857	0	100001	northoffshore	adig_s0070.g6	TRFM_RABIT
BLFC01000857	0	100001	northoffshore	adig_s0070.g7	OX2R_HUMAN
BLFC01000889	250000	300001	northoffshore	adig_s0058.g7	WDR7_RAT
BLFC01000909	50000	100001	northoffshore	adig_s0064.g3	XYL1_ASPFU
BLFC01000909	50000	100001	northoffshore	adig_s0064.g4	USOM4_ACRMI
BLFC01000909	50000	100001	northoffshore	adig_s0064.g5	TASOR_HUMAN
BLFC01000909	50000	100001	northoffshore	adig_s0064.g6	PTN23_RAT
BLFC01000909	50000	100001	northoffshore	adig_s0064.g7	IRF2_SIGHI
BLFC01000927	850000	900001	northoffshore	adig_s0079.g65	PGBD5_HUMAN
BLFC01000930	500000	550001	northoffshore	adig_s0088.g21	YIDC_HERAR
BLFC01000930	500000	550001	northoffshore	adig_s0088.g22	MYC2_ORYSJ
BLFC01000047	750000	800001	southoffshore	NA	NA
BLFC01000051	400000	450001	southoffshore	adig_s0009.g18	LAC4_TRAVE
BLFC01000051	400000	450001	southoffshore	adig_s0009.g19	YSM6_CAEEL
BLFC01000055	150000	250001	southoffshore	adig_s0084.g11	ZPP_ACRMI
BLFC01000055	150000	250001	southoffshore	adig_s0084.g12	CIPK9_ORYSJ
BLFC01000055	150000	250001	southoffshore	adig_s0084.g13	SIBA_DICDI
BLFC01000055	150000	250001	southoffshore	adig_s0084.g14	OIT3_HUMAN
BLFC01000055	150000	250001	southoffshore	adig_s0084.g15	SIBA_DICDI
BLFC01000056	1050000	1100001	southoffshore	adig_s0099.g34	ZNT6A_XENLA
BLFC01000056	1050000	1100001	southoffshore	adig_s0099.g35	ZNT6_DANRE
BLFC01000057	650000	700001	southoffshore	adig_s0068.g34	CSR2B_HUMAN
BLFC01000057	650000	700001	southoffshore	adig_s0068.g35	ZAP70_MOUSE
BLFC01000057	650000	700001	southoffshore	adig_s0068.g36	GPSM2_MOUSE

BLFC01000057	650000	700001	southoffshore	adig_s0068.g37	SPAT1_HUMAN
BLFC01000057	650000	700001	southoffshore	adig_s0068.g38	NODB_BRAEL
BLFC01000057	650000	700001	southoffshore	adig_s0068.g39	RDH12_MOUSE
BLFC01000057	650000	700001	southoffshore	adig_s0068.g40	CLPP_MOUSE
BLFC01000074	1200000	1250001	southoffshore	adig_s0057.g31	GST1_ASCSU
BLFC01000074	1200000	1250001	southoffshore	adig_s0057.g32	GST_PLAVI
BLFC01000074	1200000	1250001	southoffshore	adig_s0057.g33	L_TUPVT
BLFC01000074	1200000	1250001	southoffshore	adig_s0057.g34	GST_ANOGA
BLFC01000074	1200000	1250001	southoffshore	adig_s0057.g35	NLRC3_HUMAN
BLFC01000089	100000	200001	southoffshore	NA	NA
BLFC01000100	3250000	3350001	southoffshore	adig_s0008.g145	TTN1_CAEEL
BLFC01000100	3250000	3350001	southoffshore	adig_s0008.g147	FB76_ARATH
BLFC01000100	3250000	3350001	southoffshore	adig_s0008.g148	FRUA_STRMU
BLFC01000100	3250000	3350001	southoffshore	adig_s0008.g149	KSR2_HUMAN
BLFC01000100	3250000	3350001	southoffshore	adig_s0008.g150	SNX27_MOUSE
BLFC01000100	3250000	3350001	southoffshore	adig_s0008.g151	CRTC1_MOUSE
BLFC01000106	450000	500001	southoffshore	adig_s0132.g35	ARPC4_XENLA
BLFC01000106	450000	500001	southoffshore	adig_s0132.g36	FURIN_BOVIN
BLFC01000106	450000	500001	southoffshore	adig_s0132.g37	DCDC2_RAT
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g130	CADH4_CAEEL
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g132	STL2_ARATH
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g133	NSE2_XENLA
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g134	RN217_HUMAN
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g135	FA8A1_HUMAN
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g137	ACOX1_CAVPO

BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g138	JMJD7_MOUSE
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g139	XRCC5_MOUSE
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g140	FZD2_XENLA
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g141	NAS31_CAEEL
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g142	FAXC_XENTR
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g143	TRM11_CHICK
BLFC01000123	1650000	1750001	southoffshore	adig_s0039.g144	GSH1_HUMAN
BLFC01000124	650000	700001	southoffshore	adig_s0098.g54	FUCL6_ANGJA
BLFC01000124	650000	700001	southoffshore	adig_s0098.g55	TF2B_XENLA
BLFC01000124	650000	700001	southoffshore	adig_s0098.g56	EBP_DICDI
BLFC01000124	650000	700001	southoffshore	adig_s0098.g57	ZDH17_HUMAN
BLFC01000124	650000	700001	southoffshore	adig_s0098.g58	ZDH17_RAT
BLFC01000124	650000	700001	southoffshore	adig_s0098.g59	KMT5B_BOVIN
BLFC01000124	650000	700001	southoffshore	adig_s0098.g60	CAPR1_HUMAN
BLFC01000137	650000	850001	southoffshore	adig_s0108.g20	RPOB_AYWBP
BLFC01000137	650000	850001	southoffshore	adig_s0108.g21	VWF_PIG
BLFC01000137	650000	850001	southoffshore	adig_s0108.g22	ALPL_ARATH
BLFC01000137	650000	850001	southoffshore	adig_s0108.g23	ADF1_DROME
BLFC01000137	650000	850001	southoffshore	adig_s0108.g24	ZSC22_HUMAN
BLFC01000137	650000	850001	southoffshore	adig_s0108.g25	FRPD4_MOUSE
BLFC01000137	650000	850001	southoffshore	adig_s0108.g26	CNKR2_RAT
BLFC01000137	650000	850001	southoffshore	adig_s0108.g27	GUN12_ORYSJ
BLFC01000137	650000	850001	southoffshore	adig_s0108.g28	BIPB_BURTA
BLFC01000137	650000	850001	southoffshore	adig_s0108.g29	ZN862_HUMAN
BLFC01000148	600000	800001	southoffshore	adig_s0149.g21	SETD9_HUMAN

BLFC01000148	600000	800001	southoffshore	adig_s0149.g22	EX7L_FUSNN
BLFC01000148	600000	800001	southoffshore	adig_s0149.g23	KTBL1_HUMAN
BLFC01000148	600000	800001	southoffshore	adig_s0149.g24	CAYP2_HUMAN
BLFC01000148	600000	800001	southoffshore	adig_s0149.g25	PPM1A_RAT
BLFC01000148	600000	800001	southoffshore	adig_s0149.g26	PAKC_DICDI
BLFC01000148	600000	800001	southoffshore	adig_s0149.g27	ADA1B_MESAU
BLFC01000148	600000	800001	southoffshore	adig_s0149.g28	UBR7_HUMAN
BLFC01000148	600000	800001	southoffshore	adig_s0149.g29	MOK_HUMAN
BLFC01000148	600000	800001	southoffshore	adig_s0149.g30	LAMA2_HUMAN
BLFC01000148	600000	800001	southoffshore	adig_s0149.g31	DRC7_XENTR
BLFC01000152	2400000	2450001	southoffshore	adig_s0014.g143	MGT4C_HUMAN
BLFC01000152	2400000	2450001	southoffshore	adig_s0014.g144	MGT4C_HUMAN
BLFC01000152	2400000	2450001	southoffshore	adig_s0014.g145	MGT4C_HUMAN
BLFC01000152	2400000	2450001	southoffshore	adig_s0014.g146	AMOT_HUMAN
BLFC01000152	2400000	2450001	southoffshore	adig_s0014.g147	5HT1D_PIG
BLFC01000152	2400000	2450001	southoffshore	adig_s0014.g148	GP161_MOUSE
BLFC01000152	2400000	2450001	southoffshore	adig_s0014.g149	PARL_BOVIN
BLFC01000154	0	50001	southoffshore	adig_s0150.g1	AMPN_CHICK
BLFC01000154	0	50001	southoffshore	adig_s0150.g2	NPFF2_HUMAN
BLFC01000154	0	50001	southoffshore	adig_s0150.g3	GAL2B_DANRE
BLFC01000154	0	50001	southoffshore	adig_s0150.g4	QRFPR_HUMAN
BLFC01000154	0	50001	southoffshore	adig_s0150.g5	QRFPR_BRAFL
BLFC01000154	0	50001	southoffshore	adig_s0150.g6	RGS19_RAT
BLFC01000154	200000	250001	southoffshore	adig_s0150.g20	PXDN_XENTR
BLFC01000154	200000	250001	southoffshore	adig_s0150.g21	PXDN_XENTR

BLFC01000184	1050000	1100001	southoffshore	adig_s0060.g35	TM244_HUMAN
BLFC01000184	1050000	1100001	southoffshore	adig_s0060.g36	NRX1A_MOUSE
BLFC01000184	1050000	1100001	southoffshore	adig_s0060.g37	NRX2A_RAT
BLFC01000184	1050000	1100001	southoffshore	adig_s0060.g38	EPI1_CAEEL
BLFC01000184	1050000	1100001	southoffshore	adig_s0060.g39	MAB21_CAEBR
BLFC01000184	1050000	1100001	southoffshore	adig_s0060.g40	NRX2A_HUMAN
BLFC01000184	1600000	1650001	southoffshore	NA	NA
BLFC01000185	100000	300001	southoffshore	adig_s0062.g10	OCTB2_CHISP
BLFC01000185	100000	300001	southoffshore	adig_s0062.g11	OCTB2_CHISP
BLFC01000185	100000	300001	southoffshore	adig_s0062.g12	ATS6_HUMAN
BLFC01000185	100000	300001	southoffshore	adig_s0062.g13	ATS18_HUMAN
BLFC01000185	100000	300001	southoffshore	adig_s0062.g14	ATS6_HUMAN
BLFC01000185	100000	300001	southoffshore	adig_s0062.g15	ATS6_HUMAN
BLFC01000185	100000	300001	southoffshore	adig_s0062.g16	KLH28_MOUSE
BLFC01000185	100000	300001	southoffshore	adig_s0062.g17	ATS6_HUMAN
BLFC01000185	100000	300001	southoffshore	adig_s0062.g19	UBAC1_XENTR
BLFC01000185	100000	300001	southoffshore	adig_s0062.g21	LRSM1_MOUSE
BLFC01000185	100000	300001	southoffshore	adig_s0062.g22	SPTN1_HUMAN
BLFC01000185	1550000	1600001	southoffshore	adig_s0062.g124	TRAF3_MOUSE
BLFC01000185	1550000	1600001	southoffshore	adig_s0062.g125	GLHR_ANTEL
BLFC01000201	1300000	1350001	southoffshore	adig_s0044.g67	PK1L2_MOUSE
BLFC01000201	1300000	1350001	southoffshore	adig_s0044.g68	VGFAA_DANRE
BLFC01000201	1300000	1350001	southoffshore	adig_s0044.g69	RTRAF_DANRE
BLFC01000201	1300000	1350001	southoffshore	adig_s0044.g70	RTRAF_DANRE
BLFC01000201	1300000	1350001	southoffshore	adig_s0044.g71	BAR3_CHITE

BLFC01000201	2350000	2400001	southoffshore	adig_s0044.g129	LAMA1_MOUSE
BLFC01000208	350000	400001	southoffshore	adig_s0199.g32	ADRB2_TSCTR
BLFC01000208	350000	400001	southoffshore	adig_s0199.g33	TENX_HUMAN
BLFC01000208	350000	400001	southoffshore	adig_s0199.g34	BTBD6_XENLA
BLFC01000208	350000	400001	southoffshore	adig_s0199.g35	LR74A_RAT
BLFC01000208	350000	400001	southoffshore	adig_s0199.g36	FCF1_PONAB
BLFC01000208	350000	400001	southoffshore	adig_s0199.g37	RCA_ARATH
BLFC01000208	350000	400001	southoffshore	adig_s0199.g38	K1755_MOUSE
BLFC01000211	400000	450001	southoffshore	adig_s0090.g24	ERLEC_MOUSE
BLFC01000243	650000	700001	southoffshore	NA	NA
BLFC01000255	450000	500001	southoffshore	adig_s0127.g15	LACXP_LACLL
BLFC01000256	1300000	1400001	southoffshore	adig_s0174.g63	PLXA4_DANRE
BLFC01000286	1950000	2000001	southoffshore	NA	NA
BLFC01000289	650000	700001	southoffshore	adig_s0153.g27	NR2E3_BOVIN
BLFC01000289	650000	700001	southoffshore	adig_s0153.g28	YDAC_BACSU
BLFC01000298	1300000	1350001	southoffshore	adig_s0120.g105	4CL1_TOBAC
BLFC01000298	1300000	1350001	southoffshore	adig_s0120.g106	RGPA1_HUMAN
BLFC01000298	1300000	1350001	southoffshore	adig_s0120.g107	RGPA2_HUMAN
BLFC01000303	500000	550001	southoffshore	adig_s0189.g36	ORCT_DROME
BLFC01000303	500000	550001	southoffshore	adig_s0189.g37	K0513_HUMAN
BLFC01000303	500000	550001	southoffshore	adig_s0189.g38	DLRB2_BOVIN
BLFC01000303	500000	550001	southoffshore	adig_s0189.g39	TM198_XENTR
BLFC01000303	500000	550001	southoffshore	adig_s0189.g40	GRIA2_MOUSE
BLFC01000303	500000	550001	southoffshore	adig_s0189.g41	DUS22_HUMAN
BLFC01000309	1800000	1850001	southoffshore	adig_s0069.g114	HELS_HALMA

BLFC01000309	1800000	1850001	southoffshore	adig_s0069.g115	TEX26_MOUSE
BLFC01000309	1800000	1850001	southoffshore	adig_s0069.g116	GLGA_STRA5
BLFC01000309	1800000	1850001	southoffshore	adig_s0069.g117	PMP8_CHLPN
BLFC01000309	1800000	1850001	southoffshore	adig_s0069.g118	SPT6_USTMA
BLFC01000309	1800000	1850001	southoffshore	adig_s0069.g119	CASK_RAT
BLFC01000309	2150000	2200001	southoffshore	adig_s0069.g137	NDKA2_XENLA
BLFC01000309	2150000	2200001	southoffshore	adig_s0069.g139	RH30_ORYSJ
BLFC01000309	2150000	2200001	southoffshore	adig_s0069.g140	CO5A1_MOUSE
BLFC01000309	2150000	2200001	southoffshore	adig_s0069.g141	PTH_LISMH
BLFC01000309	2150000	2200001	southoffshore	adig_s0069.g142	MANA1_BACSU
BLFC01000317	1000000	1050001	southoffshore	adig_s0091.g49	FA5V_OXYSU
BLFC01000317	1000000	1050001	southoffshore	adig_s0091.g50	SF3B3_MOUSE
BLFC01000348	2050000	2150001	southoffshore	NA	NA
BLFC01000375	100000	450001	southoffshore	adig_s0052.g7	CSEN_RAT
BLFC01000375	100000	450001	southoffshore	adig_s0052.g8	TEDC2_HUMAN
BLFC01000375	100000	450001	southoffshore	adig_s0052.g9	KLHDB_DROER
BLFC01000375	100000	450001	southoffshore	adig_s0052.g10	CG026_DANRE
BLFC01000375	100000	450001	southoffshore	adig_s0052.g11	NCKX5_HUMAN
BLFC01000375	100000	450001	southoffshore	adig_s0052.g12	NCKX3_HUMAN
BLFC01000375	100000	450001	southoffshore	adig_s0052.g13	NCKX5_DANRE
BLFC01000375	100000	450001	southoffshore	adig_s0052.g14	PGBD4_HUMAN
BLFC01000375	100000	450001	southoffshore	adig_s0052.g15	NCKX5_DANRE
BLFC01000375	100000	450001	southoffshore	adig_s0052.g16	LC7L2_HUMAN
BLFC01000375	100000	450001	southoffshore	adig_s0052.g17	RLF33_ARATH
BLFC01000375	100000	450001	southoffshore	adig_s0052.g18	CO6A1_HUMAN

BLFC01000375	100000	450001	southoffshore	adig_s0052.g19	RBSA1_BURCM
BLFC01000375	100000	450001	southoffshore	adig_s0052.g21	TM220_XENLA
BLFC01000375	100000	450001	southoffshore	adig_s0052.g22	ACON_MOUSE
BLFC01000375	100000	450001	southoffshore	adig_s0052.g25	LST8_DANRE
BLFC01000375	100000	450001	southoffshore	adig_s0052.g26	ASTE1_HUMAN
BLFC01000375	100000	450001	southoffshore	adig_s0052.g27	ASTE1_HUMAN
BLFC01000393	1950000	2000001	southoffshore	adig_s0102.g133	USOM1_ACRMI
BLFC01000393	1950000	2000001	southoffshore	adig_s0102.g134	USOM1_ACRMI
BLFC01000404	1250000	1300001	southoffshore	adig_s0020.g113	PEX14_MOUSE
BLFC01000404	1250000	1300001	southoffshore	adig_s0020.g114	NOC2L_MOUSE
BLFC01000404	1250000	1300001	southoffshore	adig_s0020.g115	PCL9_YEAST
BLFC01000407	1950000	2000001	southoffshore	adig_s0145.g119	RNT_SHEB8
BLFC01000454	300000	400001	southoffshore	adig_s0096.g7	BMPH_STRPU
BLFC01000454	300000	400001	southoffshore	adig_s0096.g8	E2F3_HUMAN
BLFC01000454	300000	400001	southoffshore	adig_s0096.g9	GANP_MOUSE
BLFC01000454	300000	400001	southoffshore	adig_s0096.g10	PGCB_MOUSE
BLFC01000468	300000	400001	southoffshore	adig_s0172.g12	DCR1A_CHICK
BLFC01000468	300000	400001	southoffshore	adig_s0172.g13	CCNJ_XENLA
BLFC01000468	300000	400001	southoffshore	adig_s0172.g15	SOX8_TETNG
BLFC01000468	300000	400001	southoffshore	adig_s0172.g16	NA
BLFC01000468	300000	400001	southoffshore	adig_s0172.g17	CC142_MOUSE
BLFC01000468	300000	400001	southoffshore	adig_s0172.g19	MF13A_CHICK
BLFC01000468	300000	400001	southoffshore	adig_s0172.g21	CFA43_MOUSE
BLFC01000480	800000	900001	southoffshore	adig_s0100.g67	GDL23_ARATH
BLFC01000480	800000	900001	southoffshore	adig_s0100.g68	TATA_AROAE

BLFC01000480	800000	900001	southoffshore	adig_s0100.g69	MRAY_LACS1
BLFC01000480	800000	900001	southoffshore	adig_s0100.g71	CNTP5_CHICK
BLFC01000480	800000	900001	southoffshore	adig_s0100.g72	CNTP5_CHICK
BLFC01000480	800000	900001	southoffshore	adig_s0100.g73	CNTP5_CHICK
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g38	CPS_HUNT2
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g39	ATPG_SACEN
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g40	RP1_PAPHA
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g41	MA2C1_RAT
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g42	MA2C1_MOUSE
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g43	OX1R_BOVIN
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g44	NMBR_MOUSE
BLFC01000542	200000	300001	southoffshore	adig_s0177.g14	G2E3_HUMAN
BLFC01000557	550000	600001	southoffshore	adig_s0152.g15	PYRR_NATTJ
BLFC01000557	550000	600001	southoffshore	adig_s0152.g16	ZMYM2_HUMAN
BLFC01000557	550000	600001	southoffshore	adig_s0152.g17	ALPL_ARATH
BLFC01000565	300000	350001	southoffshore	adig_s0198.g21	NOD2_MOUSE
BLFC01000583	50000	100001	southoffshore	adig_s0268.g2	RBM38_HUMAN
BLFC01000596	550000	600001	southoffshore	adig_s0006.g36	MNTH_STAS1
BLFC01000596	550000	600001	southoffshore	adig_s0006.g37	STT3_DICDI
BLFC01000596	550000	600001	southoffshore	adig_s0006.g38	NA
BLFC01000596	550000	600001	southoffshore	adig_s0006.g39	MBTP1_HUMAN
BLFC01000596	1400000	1450001	southoffshore	adig_s0006.g102	PMD1_LOCMI
BLFC01000596	1400000	1450001	southoffshore	adig_s0006.g103	LAMC2_HORSE
BLFC01000596	1400000	1450001	southoffshore	adig_s0006.g104	AOSL_PLEHO
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g220	LRC8A_HUMAN

BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g221	LRC8A_RAT
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g222	MOK11_SCHPO
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g223	AF17_HUMAN
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g224	ISPD_RALSO
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g225	LRC8A_HUMAN
BLFC01000599	750000	800001	southoffshore	adig_s0135.g44	GRB2_XENTR
BLFC01000599	750000	800001	southoffshore	adig_s0135.g45	UBE2S_NEMVE
BLFC01000599	750000	800001	southoffshore	adig_s0135.g46	SMBT1_HUMAN
BLFC01000599	750000	800001	southoffshore	adig_s0135.g47	Y733_CORDI
BLFC01000600	400000	450001	southoffshore	adig_s0005.g51	RBP2_PANTR
BLFC01000600	400000	450001	southoffshore	adig_s0005.g52	TIR_BRADI
BLFC01000600	400000	450001	southoffshore	adig_s0005.g53	MCM7_HUMAN
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g338	SHE2_ASHGO
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g339	RET_HUMAN
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g340	ASPM_MACMU
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g342	FRIS_LYMST
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g341	FRIS_LYMST
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g343	FRIS_LYMST
BLFC01000632	550000	600001	southoffshore	adig_s0004.g41	KCMB3_RAT
BLFC01000632	550000	600001	southoffshore	adig_s0004.g42	SODC_CAEBR
BLFC01000632	2950000	3000001	southoffshore	adig_s0004.g211	CADN_ACRMI
BLFC01000632	2950000	3000001	southoffshore	adig_s0004.g212	NK2R_RABIT
BLFC01000632	2950000	3000001	southoffshore	adig_s0004.g213	NPFF2_HUMAN
BLFC01000645	400000	450001	southoffshore	adig_s0166.g32	PL8L1_HUMAN
BLFC01000645	400000	450001	southoffshore	adig_s0166.g35	ADA12_HUMAN

BLFC01000647	1450000	1500001	southoffshore	adig_s0012.g104	TORSO_BOMMO
BLFC01000647	1450000	1500001	southoffshore	adig_s0012.g105	CHIA_BOVIN
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g243	TPC1_RAT
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g244	PMGT1_PONAB
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g245	PUS10_HUMAN
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g246	NRP2_HUMAN
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g247	Y1425_MYCTU
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g248	Y1425_MYCTU
BLFC01000660	700000	750001	southoffshore	adig_s0133.g29	EPHA7_CHICK
BLFC01000660	700000	750001	southoffshore	adig_s0133.g30	MFS12_HUMAN
BLFC01000660	700000	750001	southoffshore	adig_s0133.g32	MFS12_MOUSE
BLFC01000692	0	50001	southoffshore	adig_s0191.g1	DNLI4_YEAST
BLFC01000692	0	50001	southoffshore	adig_s0191.g2	MLXIP_MOUSE
BLFC01000734	800000	850001	southoffshore	adig_s0081.g71	CENPO_HUMAN
BLFC01000734	800000	850001	southoffshore	adig_s0081.g72	MAB21_DROME
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g116	LEXA_CHRSD
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g117	AMT1_CAEEL
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g118	BMAL1_HUMAN
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g119	INV2_DAUCA
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g120	CHKA_HUMAN
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g121	RL27A_XENLA
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g123	MFS6L_DANRE
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g124	OSBL8_MOUSE
BLFC01000756	600000	650001	southoffshore	adig_s0089.g37	KS6A1_CAEEL
BLFC01000763	350000	400001	southoffshore	adig_s0200.g12	HMU_HALWD

BLFC01000763	350000	400001	southoffshore	adig_s0200.g13	PLB_DRYCN
BLFC01000763	350000	400001	southoffshore	adig_s0200.g14	ENDD1_MOUSE
BLFC01000765	100000	300001	southoffshore	adig_s0157.g6	LIN41_MOUSE
BLFC01000765	100000	300001	southoffshore	adig_s0157.g7	CRERF_DROME
BLFC01000765	100000	300001	southoffshore	adig_s0157.g10	SLN13_RAT
BLFC01000765	100000	300001	southoffshore	adig_s0157.g11	KDM1B_HUMAN
BLFC01000766	900000	950001	southoffshore	adig_s0059.g67	MLRP1_ACRMI
BLFC01000766	900000	950001	southoffshore	adig_s0059.g68	WSDU1_RAT
BLFC01000774	1650000	1700001	southoffshore	adig_s0106.g64	IGBP1_HUMAN
BLFC01000774	1650000	1700001	southoffshore	adig_s0106.g65	PPCS_HUMAN
BLFC01000774	1650000	1700001	southoffshore	adig_s0106.g66	NPT2B_RAT
BLFC01000774	1650000	1700001	southoffshore	adig_s0106.g67	NPT2B_RAT
BLFC01000778	700000	800001	southoffshore	adig_s0107.g45	FXL14_HUMAN
BLFC01000778	700000	800001	southoffshore	adig_s0107.g46	WNT5A_HUMAN
BLFC01000778	700000	800001	southoffshore	adig_s0107.g47	POL_FFV
BLFC01000778	700000	800001	southoffshore	adig_s0107.g48	PPID_HAEIN
BLFC01000778	700000	800001	southoffshore	adig_s0107.g49	DNAE2_RHOP5
BLFC01000818	2100000	2150001	southoffshore	adig_s0037.g145	GBP2_RAT
BLFC01000818	2100000	2150001	southoffshore	adig_s0037.g146	GRIK2_XENLA
BLFC01000818	2100000	2150001	southoffshore	adig_s0037.g147	GLR37_ARATH
BLFC01000818	2100000	2150001	southoffshore	adig_s0037.g148	GFPT1_MOUSE
BLFC01000827	1250000	1300001	southoffshore	adig_s0018.g63	YL091_MIMIV
BLFC01000829	450000	500001	southoffshore	adig_s0242.g39	KUP_GEOUR
BLFC01000829	450000	500001	southoffshore	adig_s0242.g40	CMOB_HAEIN
BLFC01000829	450000	500001	southoffshore	adig_s0242.g42	AMPE_BOVIN

BLFC01000829	450000	500001	southoffshore	adig_s0242.g43	AMPE_BOVIN
BLFC01000834	1150000	1200001	southoffshore	adig_s0028.g79	KCNG2_CHICK
BLFC01000834	1150000	1200001	southoffshore	adig_s0028.g80	DER_NATTJ
BLFC01000834	3550000	3600001	southoffshore	adig_s0028.g255	LOXH1_MOUSE
BLFC01000834	3550000	3600001	southoffshore	adig_s0028.g256	OPN4B_XENLA
BLFC01000834	3550000	3600001	southoffshore	adig_s0028.g257	CATA_PIG
BLFC01000850	1200000	1350001	southoffshore	adig_s0076.g65	PSRP_SHEDO
BLFC01000850	1200000	1350001	southoffshore	adig_s0076.g66	ZSC29_HUMAN
BLFC01000850	1200000	1350001	southoffshore	adig_s0076.g67	GPC5D_HUMAN
BLFC01000850	1200000	1350001	southoffshore	adig_s0076.g68	CED12_CAEBR
BLFC01000857	0	150001	southoffshore	adig_s0070.g1	TRFM_HUMAN
BLFC01000857	0	150001	southoffshore	adig_s0070.g2	TRFE_PAROL
BLFC01000857	0	150001	southoffshore	adig_s0070.g3	TRFM_RABIT
BLFC01000857	0	150001	southoffshore	adig_s0070.g4	TRFE_PAROL
BLFC01000857	0	150001	southoffshore	adig_s0070.g5	TRFE_ANAPL
BLFC01000857	0	150001	southoffshore	adig_s0070.g6	TRFM_RABIT
BLFC01000857	0	150001	southoffshore	adig_s0070.g7	OX2R_HUMAN
BLFC01000857	0	150001	southoffshore	adig_s0070.g8	KHDR3_MOUSE
BLFC01000857	650000	700001	southoffshore	adig_s0070.g40	DDL_THEYD
BLFC01000857	650000	700001	southoffshore	adig_s0070.g41	DDL_SYNJB
BLFC01000857	650000	700001	southoffshore	adig_s0070.g42	DDL_HYDCU
BLFC01000857	650000	700001	southoffshore	adig_s0070.g43	CV039_XENLA
BLFC01000857	650000	700001	southoffshore	adig_s0070.g44	TXD17_MOUSE
BLFC01000877	650000	700001	southoffshore	adig_s0111.g42	UFSP2_HUMAN
BLFC01000877	650000	700001	southoffshore	adig_s0111.g43	CGAS_NEMVE

BLFC01000877	650000	700001	southoffshore	adig_s0111.g44	NPHP3_HUMAN
BLFC01000877	650000	700001	southoffshore	adig_s0111.g45	UBE2A_MOUSE
BLFC01000877	650000	700001	southoffshore	adig_s0111.g46	S2543_DANRE
BLFC01000877	650000	700001	southoffshore	adig_s0111.g47	ADT_CHLRE
BLFC01000929	100000	150001	southoffshore	adig_s0013.g6	VWF_RAT
BLFC01000929	100000	150001	southoffshore	adig_s0013.g7	TF26_SCHPO
BLFC01000929	100000	150001	southoffshore	adig_s0013.g8	LAC10_ARATH
BLFC01000929	100000	150001	southoffshore	adig_s0013.g9	EIF3J_YEAST
BLFC01000929	100000	150001	southoffshore	adig_s0013.g10	VITRN_MOUSE
BLFC01000929	100000	150001	southoffshore	adig_s0013.g11	VITRN_BOVIN
BLFC01000954	750000	800001	southoffshore	adig_s0181.g74	CAHD1_HUMAN
BLFC01000954	750000	800001	southoffshore	adig_s0181.g75	CAHD1_HUMAN
BLFC01000954	750000	800001	southoffshore	adig_s0181.g76	CAHD1_MOUSE
BLFC01000954	1100000	1150001	southoffshore	adig_s0181.g97	AL_DROME
BLFC01000954	1100000	1150001	southoffshore	adig_s0181.g98	SESQ1_HUMAN
BLFC01000954	1100000	1150001	southoffshore	adig_s0181.g99	INMT_RABIT
BLFC01000954	1100000	1150001	southoffshore	adig_s0181.g101	AGRL2_RAT
BLFC01000480	800000	900001	southoffshore	adig_s0100.g70	NA
BLFC01000480	800000	900001	southoffshore	adig_s0100.g71	Q0V8S9
BLFC01000480	800000	900001	southoffshore	adig_s0100.g72	Q0V8S9
BLFC01000480	800000	900001	southoffshore	adig_s0100.g73	Q0V8S9
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g37	NA
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g38	NA
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g39	NA
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g40	NA

BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g41	P21139
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g42	Q91W89
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g43	Q0GBZ5
BLFC01000524	1050000	1250001	southoffshore	adig_s0040.g44	O54799
BLFC01000542	200000	300001	southoffshore	adig_s0177.g14	NA
BLFC01000557	550000	600001	southoffshore	adig_s0152.g15	NA
BLFC01000557	550000	600001	southoffshore	adig_s0152.g16	Q5RDJ2
BLFC01000557	550000	600001	southoffshore	adig_s0152.g17	Q9M2U3
BLFC01000565	300000	350001	southoffshore	adig_s0198.g21	Q8K3Z0
BLFC01000583	50000	100001	southoffshore	adig_s0268.g2	Q9H0Z9
BLFC01000583	50000	100001	southoffshore	adig_s0268.g3	P40123
BLFC01000596	550000	600001	southoffshore	adig_s0006.g36	NA
BLFC01000596	550000	600001	southoffshore	adig_s0006.g37	NA
BLFC01000596	550000	600001	southoffshore	adig_s0006.g38	NA
BLFC01000596	550000	600001	southoffshore	adig_s0006.g39	Q14703
BLFC01000596	1400000	1450001	southoffshore	adig_s0006.g102	NA
BLFC01000596	1400000	1450001	southoffshore	adig_s0006.g103	NA
BLFC01000596	1400000	1450001	southoffshore	adig_s0006.g104	O16025
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g220	Q8IWT6
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g221	Q4V8I7
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g222	NA
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g223	NA
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g224	NA
BLFC01000596	3450000	3500001	southoffshore	adig_s0006.g225	Q8IWT6
BLFC01000599	750000	800001	southoffshore	adig_s0135.g44	Q66II3

BLFC01000599	750000	800001	southoffshore	adig_s0135.g45	A7SE05
BLFC01000599	750000	800001	southoffshore	adig_s0135.g46	Q9UHJ3
BLFC01000599	750000	800001	southoffshore	adig_s0135.g47	NA
BLFC01000599	750000	800001	southoffshore	adig_s0135.g48	A7SE07
BLFC01000600	400000	450001	southoffshore	adig_s0005.g51	H2QII6
BLFC01000600	400000	450001	southoffshore	adig_s0005.g52	NA
BLFC01000600	400000	450001	southoffshore	adig_s0005.g53	P33993
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g338	NA
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g339	P07949
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g340	P62292
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g342	P42577
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g341	P42577
BLFC01000600	3650000	3700001	southoffshore	adig_s0005.g343	P42577
BLFC01000632	550000	600001	southoffshore	adig_s0004.g41	NA
BLFC01000632	550000	600001	southoffshore	adig_s0004.g42	A8XCP3
BLFC01000632	2950000	3000001	southoffshore	adig_s0004.g211	B3EWZ3
BLFC01000632	2950000	3000001	southoffshore	adig_s0004.g212	P79218
BLFC01000632	2950000	3000001	southoffshore	adig_s0004.g213	Q9Y5X5
BLFC01000645	400000	450001	southoffshore	adig_s0166.g32	A1L4L8
BLFC01000645	400000	450001	southoffshore	adig_s0166.g33	Q9NZF1
BLFC01000645	400000	450001	southoffshore	adig_s0166.g34	Q6DK93
BLFC01000645	400000	450001	southoffshore	adig_s0166.g35	O43184
BLFC01000647	1450000	1500001	southoffshore	adig_s0012.g104	D2IYS2
BLFC01000647	1450000	1500001	southoffshore	adig_s0012.g105	Q95M17
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g243	Q9WTN5

BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g244	Q5RCB9
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g245	Q3MIT2
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g246	O60462
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g247	P9WKC1
BLFC01000647	3000000	3050001	southoffshore	adig_s0012.g248	P9WKC1
BLFC01000660	700000	750001	southoffshore	adig_s0133.g29	NA
BLFC01000660	700000	750001	southoffshore	adig_s0133.g30	Q6NUT3
BLFC01000660	700000	750001	southoffshore	adig_s0133.g31	Q6NUT3
BLFC01000660	700000	750001	southoffshore	adig_s0133.g32	Q3U481
BLFC01000692	0	50001	southoffshore	adig_s0191.g1	NA
BLFC01000692	0	50001	southoffshore	adig_s0191.g2	NA
BLFC01000734	800000	850001	southoffshore	adig_s0081.g70	Q9HAR2
BLFC01000734	800000	850001	southoffshore	adig_s0081.g71	NA
BLFC01000734	800000	850001	southoffshore	adig_s0081.g72	Q9U3W6
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g116	NA
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g117	P54145
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g118	O00327
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g119	NA
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g120	P35790
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g121	P47830
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g122	Q99PP7
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g123	NA
BLFC01000745	1350000	1450001	southoffshore	adig_s0118.g124	B9EJ86
BLFC01000756	600000	650001	southoffshore	adig_s0089.g36	Q9Z1G4
BLFC01000756	600000	650001	southoffshore	adig_s0089.g37	Q21734

BLFC01000763	350000	400001	southoffshore	adig_s0200.g12	NA
BLFC01000763	350000	400001	southoffshore	adig_s0200.g13	Q5U2V4
BLFC01000763	350000	400001	southoffshore	adig_s0200.g14	NA
BLFC01000765	100000	300001	southoffshore	adig_s0157.g4	Q96AY4
BLFC01000765	100000	300001	southoffshore	adig_s0157.g5	F1NY98
BLFC01000765	100000	300001	southoffshore	adig_s0157.g6	NA
BLFC01000765	100000	300001	southoffshore	adig_s0157.g7	NA
BLFC01000765	100000	300001	southoffshore	adig_s0157.g8	NA
BLFC01000765	100000	300001	southoffshore	adig_s0157.g9	Q55E58
BLFC01000765	100000	300001	southoffshore	adig_s0157.g10	NA
BLFC01000765	100000	300001	southoffshore	adig_s0157.g11	Q8NB78
BLFC01000766	900000	950001	southoffshore	adig_s0059.g67	B3EWZ5
BLFC01000766	900000	950001	southoffshore	adig_s0059.g68	Q5FVN8
BLFC01000774	1650000	1700001	southoffshore	adig_s0106.g64	P78318
BLFC01000774	1650000	1700001	southoffshore	adig_s0106.g65	Q9HAB8
BLFC01000774	1650000	1700001	southoffshore	adig_s0106.g66	Q9JJ09
BLFC01000774	1650000	1700001	southoffshore	adig_s0106.g67	Q9JJ09
BLFC01000778	700000	800001	southoffshore	adig_s0107.g45	Q8N1E6
BLFC01000778	700000	800001	southoffshore	adig_s0107.g46	P41221
BLFC01000778	700000	800001	southoffshore	adig_s0107.g47	NA
BLFC01000778	700000	800001	southoffshore	adig_s0107.g48	NA
BLFC01000778	700000	800001	southoffshore	adig_s0107.g49	NA
BLFC01000818	2100000	2150001	southoffshore	adig_s0037.g145	Q63663
BLFC01000818	2100000	2150001	southoffshore	adig_s0037.g146	Q91755
BLFC01000818	2100000	2150001	southoffshore	adig_s0037.g147	Q9SDQ4

BLFC01000818	2100000	2150001	southoffshore	adig_s0037.g148	P47856
BLFC01000827	1250000	1300001	southoffshore	adig_s0018.g63	NA
BLFC01000829	450000	500001	southoffshore	adig_s0242.g37	Q5I0K7
BLFC01000829	450000	500001	southoffshore	adig_s0242.g38	Q1LWX3
BLFC01000829	450000	500001	southoffshore	adig_s0242.g39	NA
BLFC01000829	450000	500001	southoffshore	adig_s0242.g40	NA
BLFC01000829	450000	500001	southoffshore	adig_s0242.g41	Q9DCT6
BLFC01000829	450000	500001	southoffshore	adig_s0242.g42	Q32LQ0
BLFC01000829	450000	500001	southoffshore	adig_s0242.g43	Q32LQ0
BLFC01000834	1150000	1200001	southoffshore	adig_s0028.g79	O73606
BLFC01000834	1150000	1200001	southoffshore	adig_s0028.g80	NA
BLFC01000834	1150000	1200001	southoffshore	adig_s0028.g81	P21399
BLFC01000834	3550000	3600001	southoffshore	adig_s0028.g255	C8YR32
BLFC01000834	3550000	3600001	southoffshore	adig_s0028.g256	O57422
BLFC01000834	3550000	3600001	southoffshore	adig_s0028.g257	O62839
BLFC01000850	1200000	1350001	southoffshore	adig_s0076.g65	NA
BLFC01000850	1200000	1350001	southoffshore	adig_s0076.g66	Q8IWY8
BLFC01000850	1200000	1350001	southoffshore	adig_s0076.g67	NA
BLFC01000850	1200000	1350001	southoffshore	adig_s0076.g68	NA
BLFC01000857	0	150001	southoffshore	adig_s0070.g1	P08582
BLFC01000857	0	150001	southoffshore	adig_s0070.g2	O93429
BLFC01000857	0	150001	southoffshore	adig_s0070.g3	O97490
BLFC01000857	0	150001	southoffshore	adig_s0070.g4	O93429
BLFC01000857	0	150001	southoffshore	adig_s0070.g5	P56410
BLFC01000857	0	150001	southoffshore	adig_s0070.g6	O97490

BLFC01000857	0	150001	southoffshore	adig_s0070.g7	O43614
BLFC01000857	0	150001	southoffshore	adig_s0070.g8	Q9R226
BLFC01000857	650000	700001	southoffshore	adig_s0070.g40	B5YFT1
BLFC01000857	650000	700001	southoffshore	adig_s0070.g41	Q2JJV2
BLFC01000857	650000	700001	southoffshore	adig_s0070.g42	Q31I42
BLFC01000857	650000	700001	southoffshore	adig_s0070.g43	A2BD89
BLFC01000857	650000	700001	southoffshore	adig_s0070.g44	Q9CQM5
BLFC01000877	650000	700001	southoffshore	adig_s0111.g42	Q9NUQ7
BLFC01000877	650000	700001	southoffshore	adig_s0111.g43	A7SFB5
BLFC01000877	650000	700001	southoffshore	adig_s0111.g44	Q80XJ3
BLFC01000877	650000	700001	southoffshore	adig_s0111.g45	Q9Z255
BLFC01000877	650000	700001	southoffshore	adig_s0111.g46	Q5U3V7
BLFC01000877	650000	700001	southoffshore	adig_s0111.g47	P27080
BLFC01000929	100000	150001	southoffshore	adig_s0036.g2	NA
BLFC01000954	750000	800001	southoffshore	NA	NA
BLFC01000954	1100000	1150001	southoffshore	NA	NA

Supplementary Table 3.6 Summary of enriched GO terms and supporting statistics for genes overlapping selective sweeps

GO.ID	Term	Annotated	Significant	Expected	Ontology	Pop	Num Sweeps	Num Background Regions	Region p-value
GO:0042981	regulation of apoptotic process	78	7	1.42	BP	inshore	2	9	0.00443564
GO:0000981	DNA-binding transcription factor activity	108	11	2.01	MF	inshore	4	15	1.98E-05
GO:0020037	heme binding	92	7	1.71	MF	inshore	4	18	4.32E-05
GO:0055076	transition metal ion homeostasis	11	4	0.2	BP	northoffshore	2	6	0.00226627
GO:0006875	cellular metal ion homeostasis	16	4	0.29	BP	northoffshore	2	6	0.00226627
GO:0007186	G protein-coupled receptor signaling pathway	581	22	10.4	BP	northoffshore	11	55	4.73E-11
GO:0070588	calcium ion transmembrane transport	34	4	0.61	BP	northoffshore	3	8	1.02E-04
GO:0004930	G protein-coupled receptor activity	564	22	8.72	MF	northoffshore	11	55	4.73E-11
GO:0008083	growth factor activity	43	4	0.67	MF	northoffshore	2	7	0.00314694
GO:0016021	integral component of membrane	1207	36	23.64	CC	northoffshore	17	99	2.08E-15
GO:0016879	ligase activity, forming carbon-nitrogen bonds	17	4	0.29	MF	southoffshore	2	7	0.00322491
GO:0004930	G protein-coupled receptor activity	564	19	9.55	MF	southoffshore	12	55	2.23E-12

Supplmentary Table 3.7 List of genes overlapping with selective sweeps and supporting enrichment of GO terms listed in Supplementary Table 3.6

GO.ID	Ontology	Population	Gene id	Uniprot id
GO:0042981	BP	inshore	adig_s0038.g84	CASP8_RAT
GO:0042981	BP	inshore	adig_s0038.g85	DPOL_PPV01
GO:0042981	BP	inshore	adig_s0038.g87	CAPSD_GMDNV
GO:0042981	BP	inshore	adig_s0038.g88	LIPB_EHRCR
GO:0042981	BP	inshore	adig_s0038.g89	ABCAD_MOUSE
GO:0042981	BP	inshore	adig_s0038.g90	DNHD1_HUMAN
GO:0042981	BP	inshore	adig_s0041.g87	B2CL1_HUMAN
GO:0000981	MF	inshore	adig_s0005.g262	ERR3_PONAB
GO:0000981	MF	inshore	adig_s0025.g59	HNK2_XENLA
GO:0000981	MF	inshore	adig_s0069.g184	PAX2_MOUSE
GO:0000981	MF	inshore	adig_s0069.g185	DMX1B_DANRE
GO:0000981	MF	inshore	adig_s0069.g186	DMX1B_DANRE
GO:0000981	MF	inshore	adig_s0069.g187	DMX1B_DANRE
GO:0000981	MF	inshore	adig_s0069.g188	DMX1B_DANRE
GO:0000981	MF	inshore	adig_s0069.g189	DMX1B_DANRE
GO:0000981	MF	inshore	adig_s0069.g190	PITX2_HUMAN
GO:0000981	MF	inshore	adig_s0073.g154	BARH1_HUMAN
GO:0000981	MF	inshore	adig_s0073.g155	DLX4A_DANRE
GO:0020037	MF	inshore	adig_s0032.g129	C27C1_LITCT
GO:0020037	MF	inshore	adig_s0046.g22	CP17A_ICTPU
GO:0020037	MF	inshore	adig_s0114.g7	SUOX_HUMAN
GO:0020037	MF	inshore	adig_s0150.g21	PXDN_XENTR
GO:0020037	MF	inshore	adig_s0150.g23	PERC_DROME

GO:0020037	MF	inshore	adig_s0150.g24	PXDN_HUMAN
GO:0020037	MF	inshore	adig_s0150.g25	PXDN_XENTR
GO:0055076	BP	northoffshore	adig_s0005.g341	FRIS_LYMST
GO:0055076	BP	northoffshore	adig_s0005.g342	FRIS_LYMST
GO:0055076	BP	northoffshore	adig_s0005.g343	FRIS_LYMST
GO:0055076	BP	northoffshore	adig_s0012.g223	COMD1_MOUSE
GO:0006875	BP	northoffshore	adig_s0005.g341	FRIS_LYMST
GO:0006875	BP	northoffshore	adig_s0005.g342	FRIS_LYMST
GO:0006875	BP	northoffshore	adig_s0005.g343	FRIS_LYMST
GO:0006875	BP	northoffshore	adig_s0118.g138	EMRE_XENTR
GO:0007186	BP	northoffshore	adig_s0004.g194	NPFF2_HUMAN
GO:0007186	BP	northoffshore	adig_s0004.g198	NPFF2_RAT
GO:0007186	BP	northoffshore	adig_s0004.g199	QRFPR_BRAFL
GO:0007186	BP	northoffshore	adig_s0004.g200	NPFF2_MOUSE
GO:0007186	BP	northoffshore	adig_s0004.g202	NK2R_RABIT
GO:0007186	BP	northoffshore	adig_s0004.g205	NK2R_RABIT
GO:0007186	BP	northoffshore	adig_s0004.g212	NK2R_RABIT
GO:0007186	BP	northoffshore	adig_s0004.g213	NPFF2_HUMAN
GO:0007186	BP	northoffshore	adig_s0008.g33	ADRB2_MESAU
GO:0007186	BP	northoffshore	adig_s0008.g34	ADRB2_MESAU
GO:0007186	BP	northoffshore	adig_s0028.g87	AA2AR_CANLF
GO:0007186	BP	northoffshore	adig_s0031.g136	GPR63_HUMAN
GO:0007186	BP	northoffshore	adig_s0031.g138	NPY1R_CANLF
GO:0007186	BP	northoffshore	adig_s0044.g14	ADB4C_MELGA
GO:0007186	BP	northoffshore	adig_s0054.g60	GR101_LYMST

GO:0007186	BP	northoffshore	adig_s0070.g7	OX2R_HUMAN
GO:0007186	BP	northoffshore	adig_s0075.g75	DRD1L_OREMO
GO:0007186	BP	northoffshore	adig_s0075.g76	ADRB1_XENLA
GO:0007186	BP	northoffshore	adig_s0092.g60	MTR1A_SHEEP
GO:0007186	BP	northoffshore	adig_s0092.g62	GABR2_RAT
GO:0007186	BP	northoffshore	adig_s0118.g137	LSHR_BOVIN
GO:0007186	BP	northoffshore	adig_s0127.g10	OPN4B_GADMO
GO:0070588	BP	northoffshore	adig_s0004.g211	CADN_ACRMI
GO:0070588	BP	northoffshore	adig_s0017.g112	TRPC5_RABIT
GO:0070588	BP	northoffshore	adig_s0017.g113	TRPC5_RABIT
GO:0070588	BP	northoffshore	adig_s0118.g138	EMRE_XENTR
GO:0004930	MF	northoffshore	adig_s0004.g194	NPFF2_HUMAN
GO:0004930	MF	northoffshore	adig_s0004.g198	NPFF2_RAT
GO:0004930	MF	northoffshore	adig_s0004.g199	QRFPR_BRAFL
GO:0004930	MF	northoffshore	adig_s0004.g200	NPFF2_MOUSE
GO:0004930	MF	northoffshore	adig_s0004.g202	NK2R_RABIT
GO:0004930	MF	northoffshore	adig_s0004.g205	NK2R_RABIT
GO:0004930	MF	northoffshore	adig_s0004.g212	NK2R_RABIT
GO:0004930	MF	northoffshore	adig_s0004.g213	NPFF2_HUMAN
GO:0004930	MF	northoffshore	adig_s0008.g33	ADRB2_MESAU
GO:0004930	MF	northoffshore	adig_s0008.g34	ADRB2_MESAU
GO:0004930	MF	northoffshore	adig_s0028.g87	AA2AR_CANLF
GO:0004930	MF	northoffshore	adig_s0031.g136	GPR63_HUMAN
GO:0004930	MF	northoffshore	adig_s0031.g138	NPY1R_CANLF
GO:0004930	MF	northoffshore	adig_s0044.g14	ADB4C_MELGA

GO:0004930	MF	northoffshore	adig_s0054.g60	GR101_LYMST
GO:0004930	MF	northoffshore	adig_s0070.g7	OX2R_HUMAN
GO:0004930	MF	northoffshore	adig_s0075.g75	DRD1L_OREMO
GO:0004930	MF	northoffshore	adig_s0075.g76	ADRB1_XENLA
GO:0004930	MF	northoffshore	adig_s0092.g60	MTR1A_SHEEP
GO:0004930	MF	northoffshore	adig_s0092.g62	GABR2_RAT
GO:0004930	MF	northoffshore	adig_s0118.g137	LSHR_BOVIN
GO:0004930	MF	northoffshore	adig_s0127.g10	OPN4B_GADMO
GO:0008083	MF	northoffshore	adig_s0120.g46	FGF1_NOTVI
GO:0008083	MF	northoffshore	adig_s0163.g80	FGF1_MESAU
GO:0008083	MF	northoffshore	adig_s0163.g81	FGF1_NOTVI
GO:0008083	MF	northoffshore	adig_s0163.g82	FGF1_PIG
GO:0016021	CC	northoffshore	adig_s0004.g194	NPFF2_HUMAN
GO:0016021	CC	northoffshore	adig_s0004.g195	S26A6_MOUSE
GO:0016021	CC	northoffshore	adig_s0004.g198	NPFF2_RAT
GO:0016021	CC	northoffshore	adig_s0004.g199	QRFPR_BRAFL
GO:0016021	CC	northoffshore	adig_s0004.g200	NPFF2_MOUSE
GO:0016021	CC	northoffshore	adig_s0004.g202	NK2R_RABIT
GO:0016021	CC	northoffshore	adig_s0004.g205	NK2R_RABIT
GO:0016021	CC	northoffshore	adig_s0004.g212	NK2R_RABIT
GO:0016021	CC	northoffshore	adig_s0004.g213	NPFF2_HUMAN
GO:0016021	CC	northoffshore	adig_s0004.g214	ACHA6_HUMAN
GO:0016021	CC	northoffshore	adig_s0004.g215	ACH10_CHICK
GO:0016021	CC	northoffshore	adig_s0004.g216	ACHA7_MACMU
GO:0016021	CC	northoffshore	adig_s0004.g219	CERS1_MOUSE

GO:0016021	CC	northoffshore	adig_s0008.g33	ADRB2_MESAU
GO:0016021	CC	northoffshore	adig_s0008.g34	ADRB2_MESAU
GO:0016021	CC	northoffshore	adig_s0010.g97	NCL_ARATH
GO:0016021	CC	northoffshore	adig_s0011.g92	PTHD3_MOUSE
GO:0016021	CC	northoffshore	adig_s0014.g149	PARL_BOVIN
GO:0016021	CC	northoffshore	adig_s0028.g87	AA2AR_CANLF
GO:0016021	CC	northoffshore	adig_s0031.g136	GPR63_HUMAN
GO:0016021	CC	northoffshore	adig_s0031.g138	NPY1R_CANLF
GO:0016021	CC	northoffshore	adig_s0033.g53	TCPR1_MOUSE
GO:0016021	CC	northoffshore	adig_s0044.g14	ADB4C_MELGA
GO:0016021	CC	northoffshore	adig_s0054.g60	GR101_LYMST
GO:0016021	CC	northoffshore	adig_s0054.g61	PROM1_MOUSE
GO:0016021	CC	northoffshore	adig_s0061.g52	AA2AR_HORSE
GO:0016021	CC	northoffshore	adig_s0070.g7	OX2R_HUMAN
GO:0016021	CC	northoffshore	adig_s0075.g73	MRP4_HUMAN
GO:0016021	CC	northoffshore	adig_s0075.g75	DRD1L_OREMO
GO:0016021	CC	northoffshore	adig_s0075.g76	ADRB1_XENLA
GO:0016021	CC	northoffshore	adig_s0092.g60	MTR1A_SHEEP
GO:0016021	CC	northoffshore	adig_s0092.g62	GABR2_RAT
GO:0016021	CC	northoffshore	adig_s0118.g137	LSHR_BOVIN
GO:0016021	CC	northoffshore	adig_s0118.g138	EMRE_XENTR
GO:0016021	CC	northoffshore	adig_s0127.g10	OPN4B_GADMO
GO:0016021	CC	northoffshore	adig_s0248.g10	APH1A_MOUSE
GO:0016879	MF	southoffshore	adig_s0039.g144	GSH1_HUMAN
GO:0016879	MF	southoffshore	adig_s0070.g40	DDL_THEYD

GO:0016879	MF	southoffshore	adig_s0070.g41	DDL_SYNJB
GO:0016879	MF	southoffshore	adig_s0070.g42	DDL_HYDCU
GO:0004930	MF	southoffshore	adig_s0004.g212	NK2R_RABIT
GO:0004930	MF	southoffshore	adig_s0004.g213	NPFF2_HUMAN
GO:0004930	MF	southoffshore	adig_s0014.g147	5HT1D_PIG
GO:0004930	MF	southoffshore	adig_s0014.g148	GP161_MOUSE
GO:0004930	MF	southoffshore	adig_s0028.g256	OPN4B_XENLA
GO:0004930	MF	southoffshore	adig_s0037.g147	GLR37_ARATH
GO:0004930	MF	southoffshore	adig_s0040.g43	OX1R_BOVIN
GO:0004930	MF	southoffshore	adig_s0040.g44	NMBR_MOUSE
GO:0004930	MF	southoffshore	adig_s0062.g10	OCTB2_CHISP
GO:0004930	MF	southoffshore	adig_s0062.g11	OCTB2_CHISP
GO:0004930	MF	southoffshore	adig_s0062.g125	GLHR_ANTEL
GO:0004930	MF	southoffshore	adig_s0070.g7	OX2R_HUMAN
GO:0004930	MF	southoffshore	adig_s0149.g27	ADA1B_MESAU
GO:0004930	MF	southoffshore	adig_s0150.g2	NPFF2_HUMAN
GO:0004930	MF	southoffshore	adig_s0150.g3	GAL2B_DANRE
GO:0004930	MF	southoffshore	adig_s0150.g4	QRFPR_HUMAN
GO:0004930	MF	southoffshore	adig_s0150.g5	QRFPR_BRAFL
GO:0004930	MF	southoffshore	adig_s0181.g101	AGRL2_RAT
GO:0004930	MF	southoffshore	adig_s0199.g32	ADRB2_TSCTR

Supplementary Table 3.8 Loci under selection for which it was possible to estimate the TMRCA of selected haplotypes

Label*	Locus	Scaffold id	Start	End	Population	Mean TMRCA	TMRCA sd.
1	BLFC01000706_750000_1000001	BLFC01000706	750000	1000001	inshore	1022.23113	91.2890253
2	BLFC01000706_250000_500001	BLFC01000706	250000	500001	inshore	1261.67504	82.694645
3	BLFC01000593_400000_450001	BLFC01000593	400000	450001	inshore	1426.54211	213.574323
4	BLFC01000770_3100000_3150001	BLFC01000770	3100000	3150001	inshore	1734.97414	192.162473
5	BLFC01000407_1800000_2050001	BLFC01000407	1800000	2050001	inshore	2156.90928	178.667718
6	BLFC01000745_300000_400001	BLFC01000745	300000	400001	inshore	2235.52935	181.601899
7	BLFC01000326_1550000_2150001	BLFC01000326	1550000	2150001	inshore	2310.23142	150.967836
8	BLFC01000718_1250000_1400001	BLFC01000718	1250000	1400001	inshore	2893.6158	167.980191
9	BLFC01000632_2550000_3150001	BLFC01000632	2550000	3150001	northoffshore	3343.65771	310.960348
10	BLFC01000639_1300000_1550001	BLFC01000639	1300000	1550001	inshore	3654.60716	347.217466
11	BLFC01000413_150000_200001	BLFC01000413	150000	200001	inshore	3727.11823	435.044562
12	BLFC01000051_600000_650001	BLFC01000051	600000	650001	inshore	3834.19725	416.145668
13	BLFC01000690_2600000_2850001	BLFC01000690	2600000	2850001	inshore	4165.01231	331.422217
14	BLFC01000348_2400000_2450001	BLFC01000348	2400000	2450001	inshore	4398.82955	226.798594

15	BLFC01000690_3200000_3250001	BLFC01000690	3200000	3250001	inshore	5006.94791	335.667046
16	BLFC01000289_450000_500001	BLFC01000289	450000	500001	inshore	5111.6236	164.305263
17	BLFC01000765_100000_300001	BLFC01000765	100000	300001	southoffshore	5148.35059	414.801289
18	BLFC01000375_100000_450001	BLFC01000375	100000	450001	southoffshore	5681.39468	296.443366
19	BLFC01000770_850000_900001	BLFC01000770	850000	900001	inshore	5890.26726	520.085508
20	BLFC01000439_1750000_1800001	BLFC01000439	1750000	1800001	inshore	6366.77741	848.649697
21	BLFC01000154_250000_300001	BLFC01000154	250000	300001	inshore	6922.94326	469.302021
22	BLFC01000770_1200000_1250001	BLFC01000770	1200000	1250001	inshore	8440.10993	337.639839
23	BLFC01000074_100000_250001	BLFC01000074	100000	250001	inshore	8556.93901	405.978561
24	BLFC01000309_2200000_2300001	BLFC01000309	2200000	2300001	inshore	9108.53823	287.7192
25	BLFC01000185_900000_1050001	BLFC01000185	900000	1050001	inshore	10949.7931	365.050554
26	BLFC01000265_200000_250001	BLFC01000265	200000	250001	inshore	11543.6007	456.952226
27	BLFC01000600_2900000_2950001	BLFC01000600	2900000	2950001	inshore	14385.3398	283.857051
28	BLFC01000600_3700000_3800001	BLFC01000600	3700000	3800001	inshore	30747.9399	319.647159

*Column "label" is the numerical label shown in Figure 4B; TMRCA mean and sd values assume a generation time of 5 years

Supplementary Table 3.9 The likelihoods of four models with different divergence senarios

Topology model	No. of parameters	Log10(Lhood)	AIC
(NO, SO), IN	7	-3,174,851	14,620,743
(IN, SO), NO	7	-3,177,387	14,632,422
(IN, NO), SO	7	-3,177,414	14,632,545
(IN, NO, SO)	5	-3,176,949	14,630,399

Supplementary Table 3.10 The parameters estimated by six models with the parameter confident interval of model IMc

Model paramA10:G29	Model SI	Model IM	Model IMc (95% confidence intervals)	Model SC	Model EM	Model AM
-Log10(Lhood)	3,173,814	3,168,990	3,168,978	3,169,011	3,169,213	3,169,275
No. of parameters	10	14	13	14	14	11
AIC*	14,615,974	14,593,768	14,593,708	14,593,864	14,594,793	14,595,074
Model normalised relative likelihood*	0	0	1	0	0	0
Ancestral population size	367,863	393,110	389,722 [386,421-396,843]	389,045	390,786	399,267
Ancestral offshore population size	11,845	3,851	12,319[5,110-19,792]	11,860	12,503	17,320
inshore population size(ancestral-recent)	2,157	2,429	2,345 [1,695-3,368]	2,517	2,068	843
	82,397	362,597	217,578 [150,542-650,673]	231,841	440,714	258,507
north offshore population size (ancestral-recent)	1,009	8,888	11,061 [9,279-16,807]	11,454	11,804	975
	47,774	779,789	700,401 [438,943-1,039,143]	863,534	612,212	404,582
south offshore population size (ancestral-recent)	705	8,945	10,267 [7,435-12,965]	9,669	8,099	728
	314,156	429,610	334,711 [272,905-975,782]	267,146	659,425	27,922

Divergence time in generation (inshore)	311	956	1,275 [1,012-1,586]	1,258	1,354	706
Divergence time in generation (offshore)	122	836	904 [801-978]	917	967	118
Time of changed migration in generation	-	-	-	903	102	-
Migration rate (fration of individuals per generation * 2Ne) between offshore ancestral and inshore	-	2.12E-08	-	-	-	6.92E-04
Migration rate (fration of individuals per generation * 2Ne) between north offshore and inshore populations	-	1.91E-04	1.85E-04 [1.77E-04 - 1.919E-04]	1.86E-04	1.92E-04	-
Migration rate (fration of individuals per generation * 2Ne) between south offshore and inshore populations	-	1.89E-04	1.90E-04 [1.77E-04 - 1.918E-04]	1.83E-04	1.89E-04	-
Migration rate (fration of individuals per generation * 2Ne) between offshore populations	-	9.80E-05	1.76E-04 [4E-07 - 1.79E-04]	1.56E-04	1.43E-04	-

Parameters are defined for Supplementary Fig 13

*AIC=2d-2ln(Lhood)

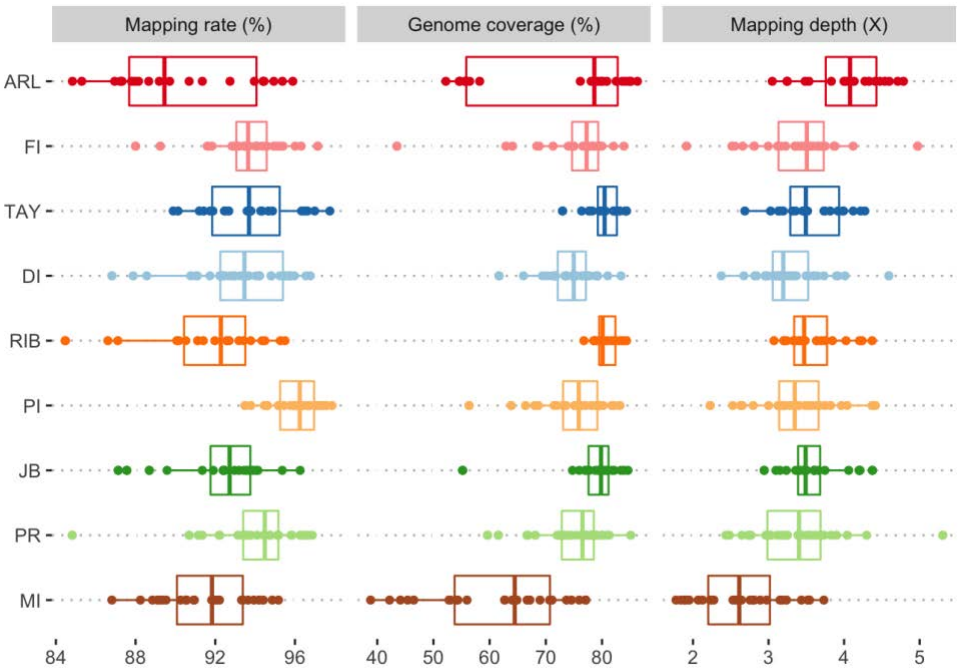
*Weighted AIC was calculated as Laurent Excoffier et al. 2013

Supplementary Table 4.1 Running time of a pairwise *S* distance calculation of *Symbiodinium* genome and *Durusdinium* genome using the original version alignment-free tools and d2ssect.

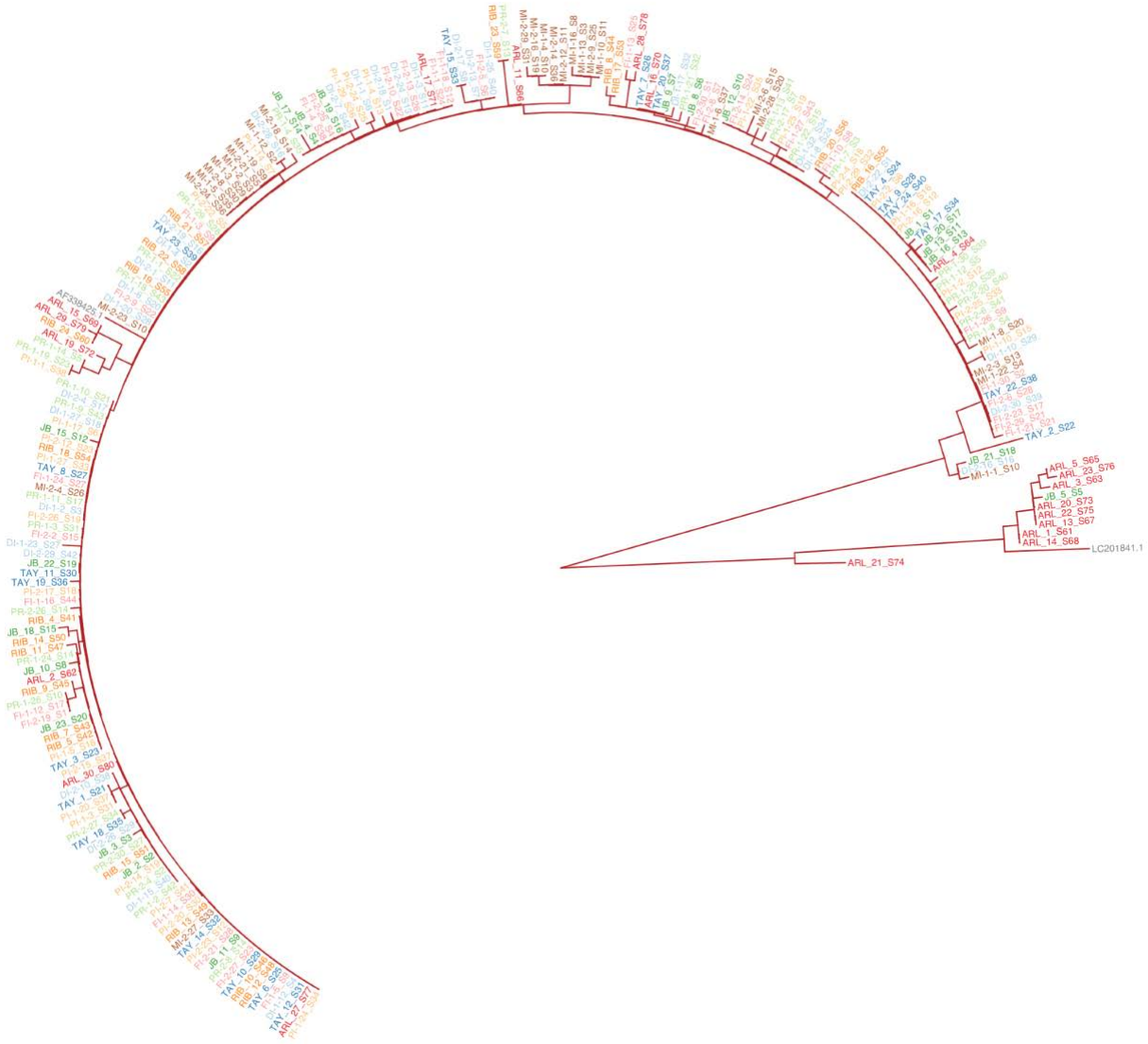
Steps in alignment-free-tools	**Time used in AF-tools**	**Time used in d2ssect**
s1 char_freq	00:00:39	/
s2_nkz_A	03:37:02	
s2_nkz_D	04:29:19	
s3_cal_d2s	06:59:41	
Generating matrix	00:00:01	
Total time	15:06:42	00:41:39
Speed up	1.00 x	22.21 x

Supplementary Table 4.2 The total running time used in analysing different simulated datasets with the different number of threads.

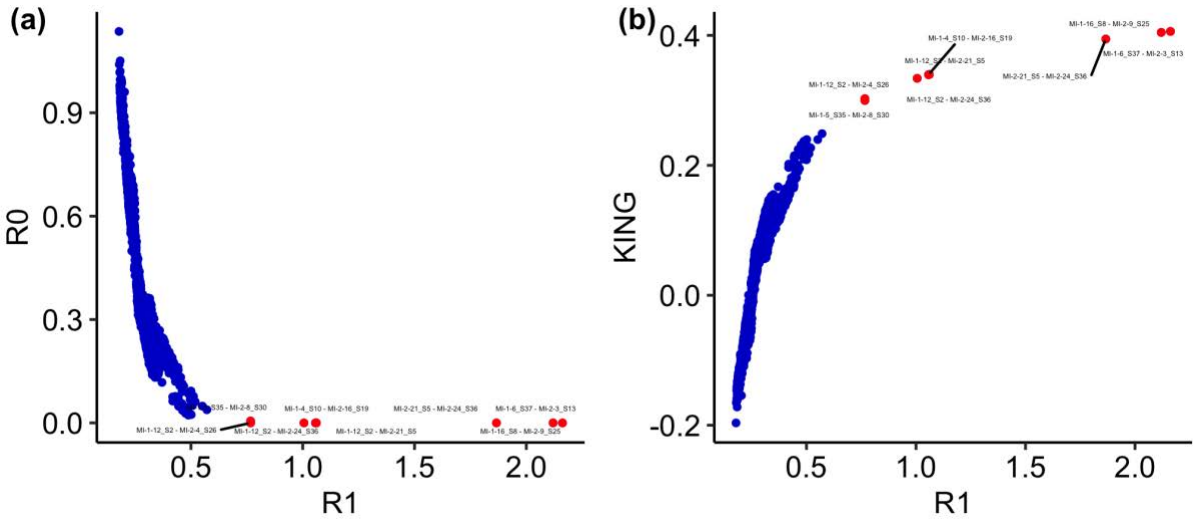
Threads	**1**		**10**		**20**		**30**	
	N=3	N=5	N=3	N=5	N=3	N=5	N=3	N=5
10K	89.952	179.558	10.22	17.649	6.144	11.226	5.716	8.685
100K	702.43	1483.147	95.288	164.87	55.16	107.098	51.18	78.681
500K	3753.049	6931.74	448.458	802.302	260.493	522.019	249.687	396.855



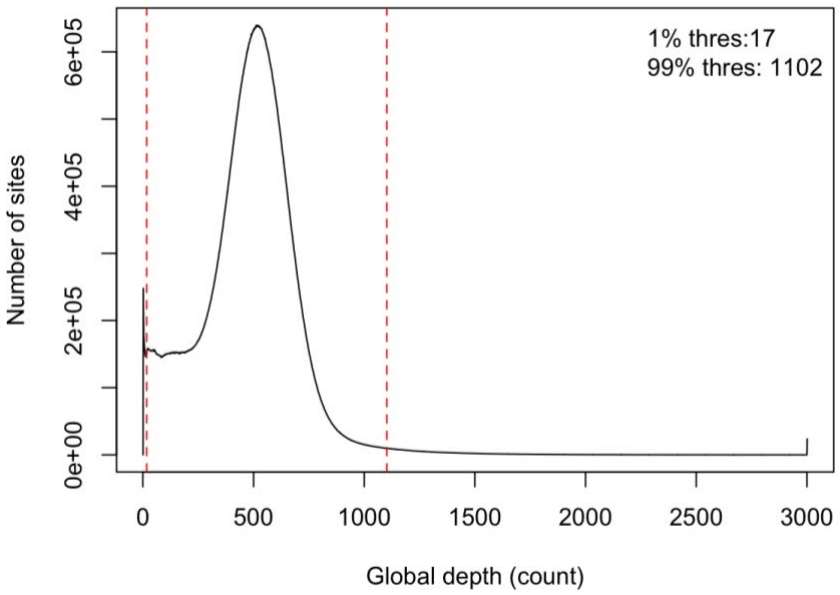
Supplementary Fig 2.1 The distribution of mapping rates and genome coverage for all 228 *A. tenuis* samples in 9 reef locations. Genome coverage was calculated as the proportion of genome bases in the reference assembly mapped by at least one read.



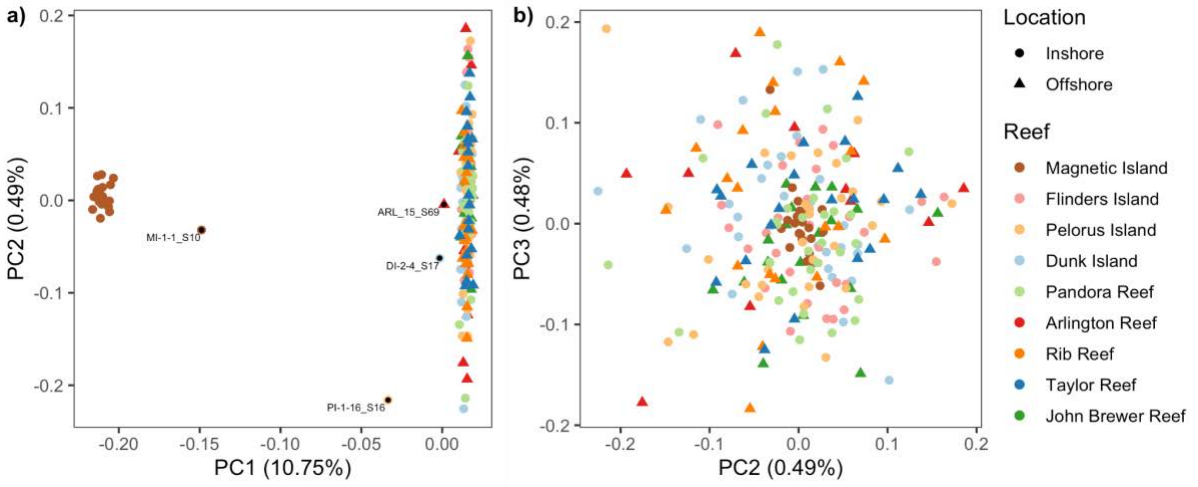
Supplementary Fig 2.2 The maximum likelihood tree of consensus sequences of the mitochondrial genome in 228 *A. tenuis* samples. Tip labels were colored in Supplementary Figure 2.1 for each reef location, the reference mitogenome sequences of *A. tenuis* (AF338425.1) and A. echinata (LC201841.1) were colored in grey.



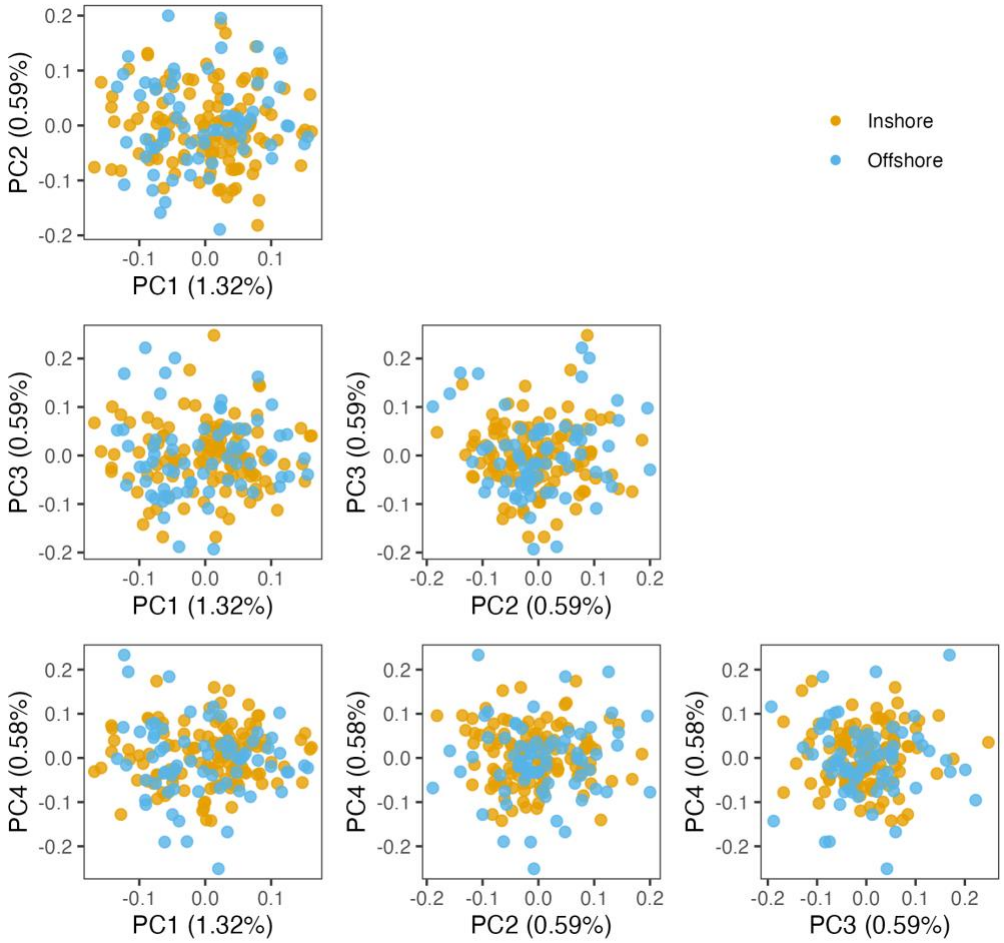
Supplementary Fig 2.3 The R1-R0 and R1-KING scatterplots for all *A.tenuis* samples calculated by NgsRelate. Each dot represents an individual sample and red dots represent samples identified as closed relatives.



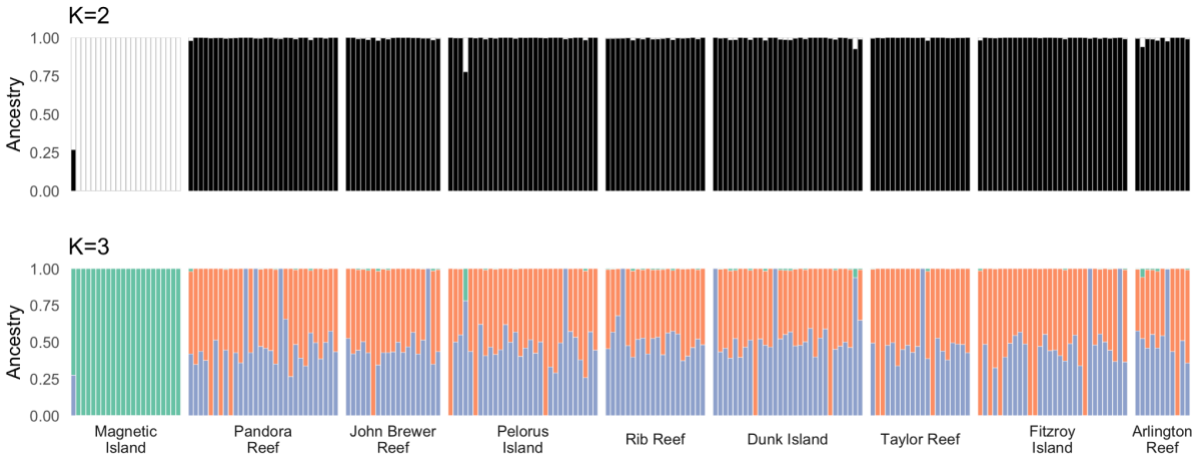
Supplementary Fig 2.4 In reference filtering, we excluded sites with extreme sequencing depth calculated by ANGSD for all samples except outliers and closely related individuals. The numbers and red dash lines represent the estimated lower and upper 1% percentile global-depth threshold.



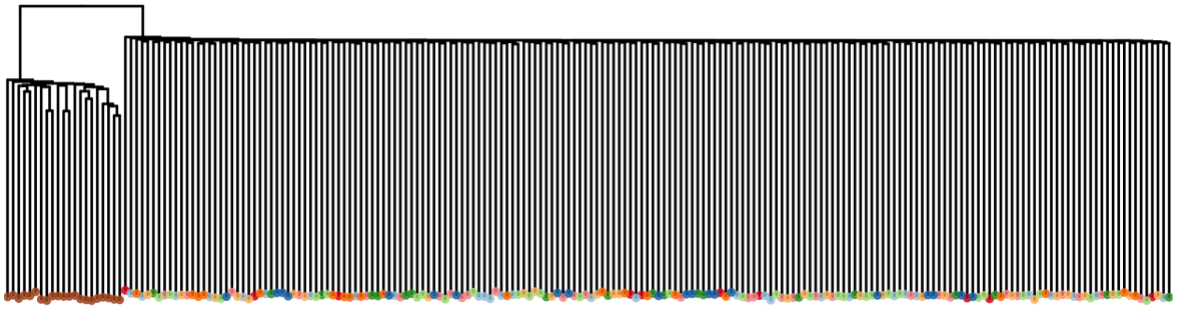
Supplementary Fig 2.5 Principle component analysis of *A. tenuis* based on genotype likelihoods of all variants. Plots display a) PC1 against PC2 and b) PC2 against PC3 with points colored by reefs and shaped by location. The text labels in a) indicate the samples identified as hybrids.



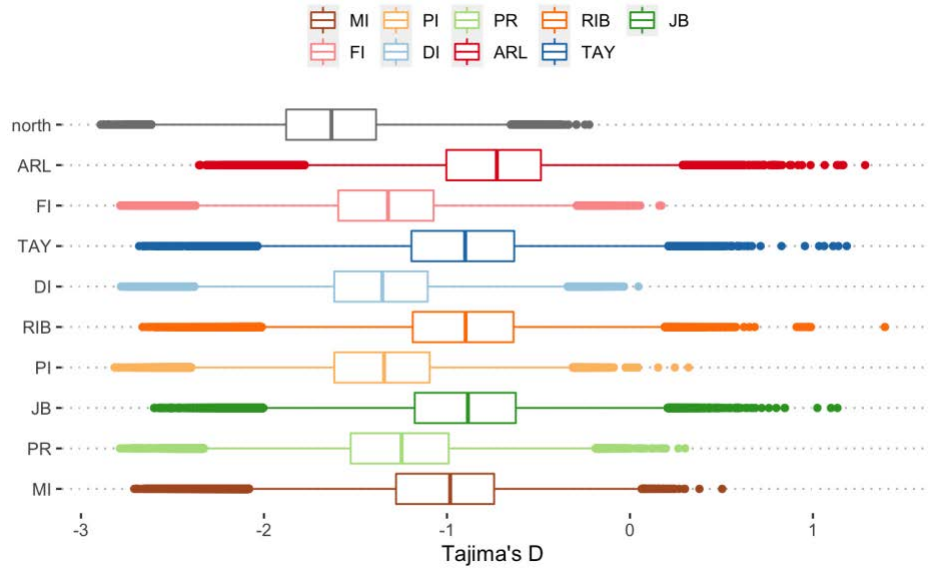
Supplementary Fig 2.6 The PCA plots of the North GBR samples which are labelled as inshore and offshore. From PC1 to PC4, there is no obvious clustering pattern reflecting the difference between inshore and offshore. This suggests there is no systematic bias in our analysis based on variants in all samples.



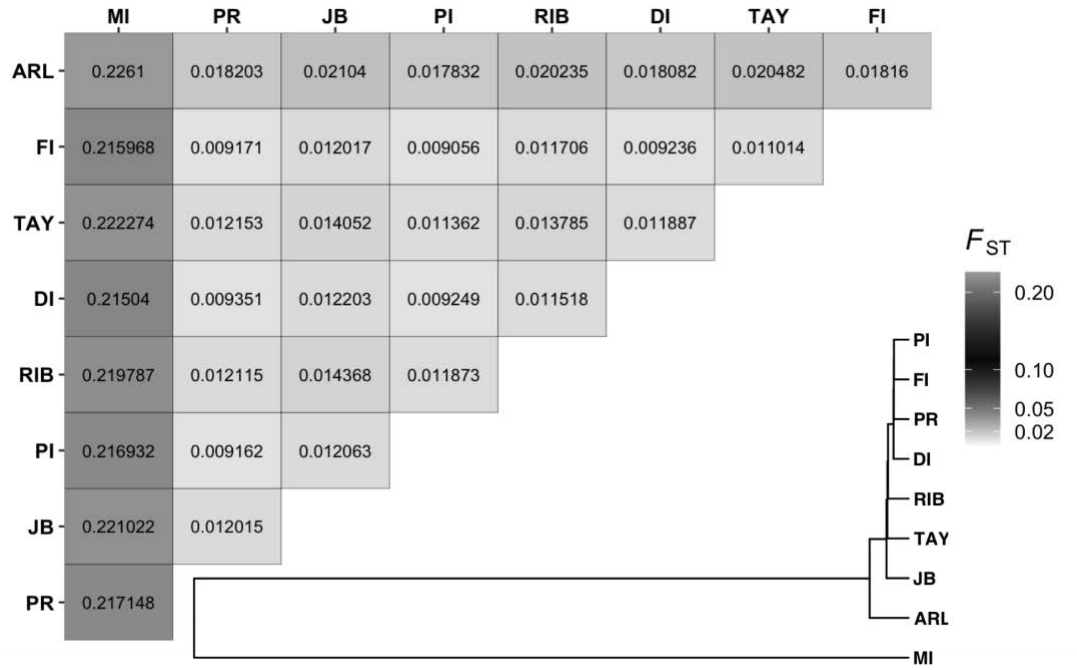
Supplementary Fig 2.7 Ancestry proportions estimated in NGSadmix for K=2 (top) and K=3 (bottom). The mixed bars represent individuals with mixed ancestry profiles with different proportions.



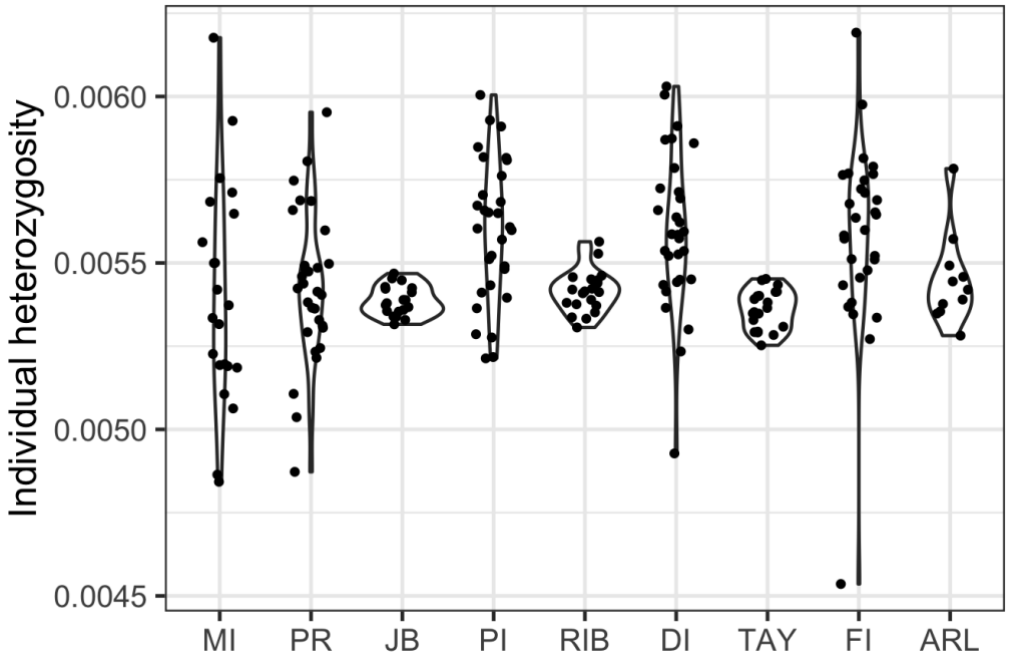
Supplementary Fig 2.8 The neighbour-joining (NJ) tree based on an identity-by-state (IBS) matrix of 212 individuals calculated in ANGSD.



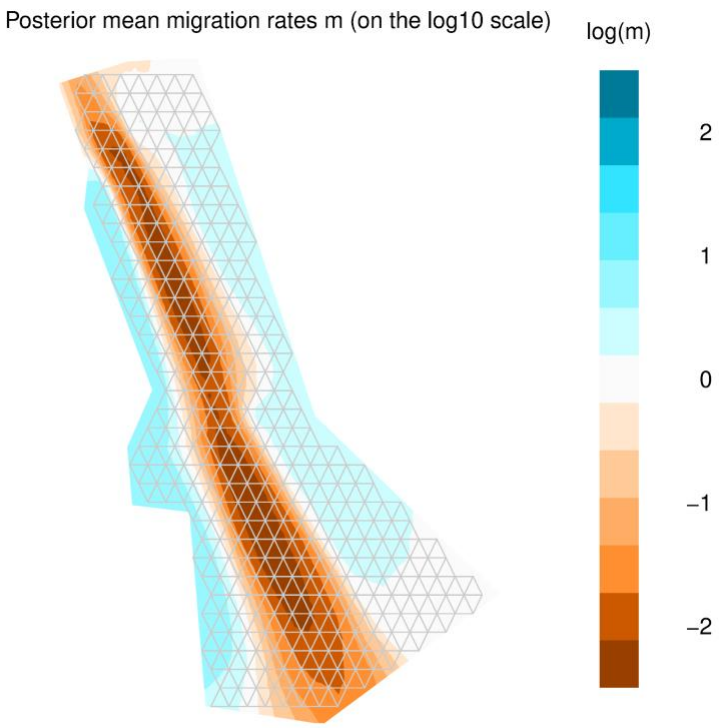
Supplementary Fig 2.9 The distribution of genome-wide estimate of Tajima's D in each reef.



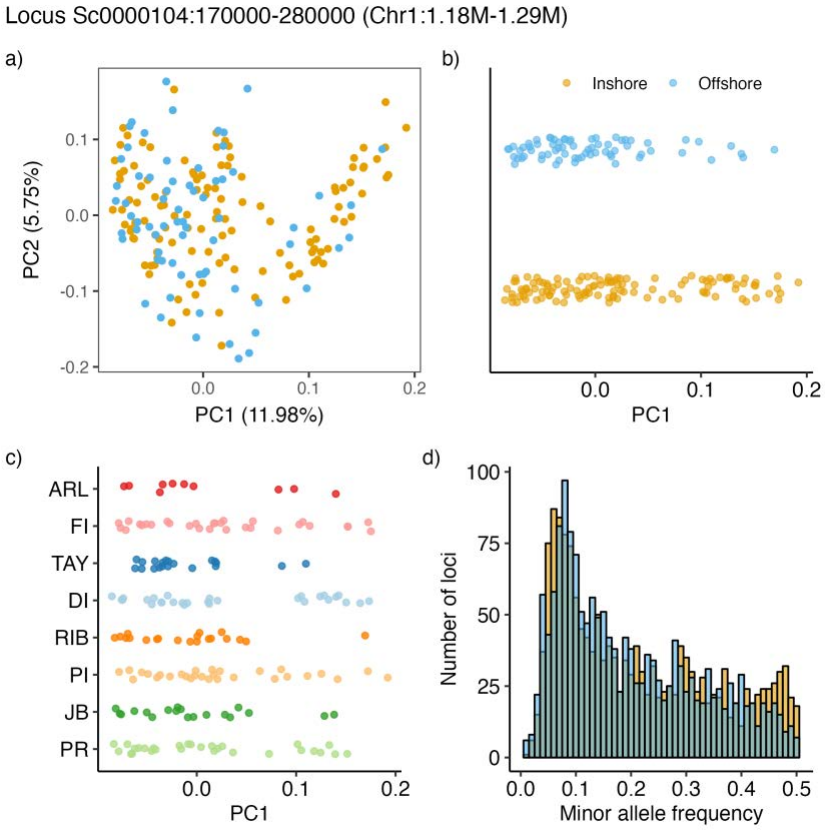
Supplementary Fig 2.10 Pairwise F_{ST} between populations and the clustering of the Fst matrix



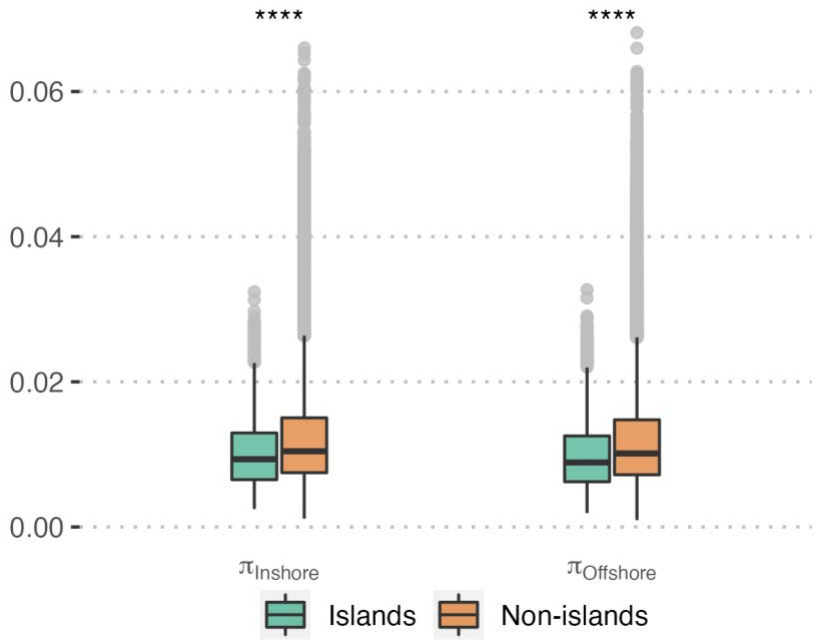
Supplementary Fig 2.11 The individual heterozygosity of samples at each reef location. The individual heterozygosity was calculated as the number of heterozygous sites divided by the total number of sites in each sample. Although similar overall values were observed in reefs, the spread of individual heterozygosity is much wider in inshore reefs than in offshore reefs.



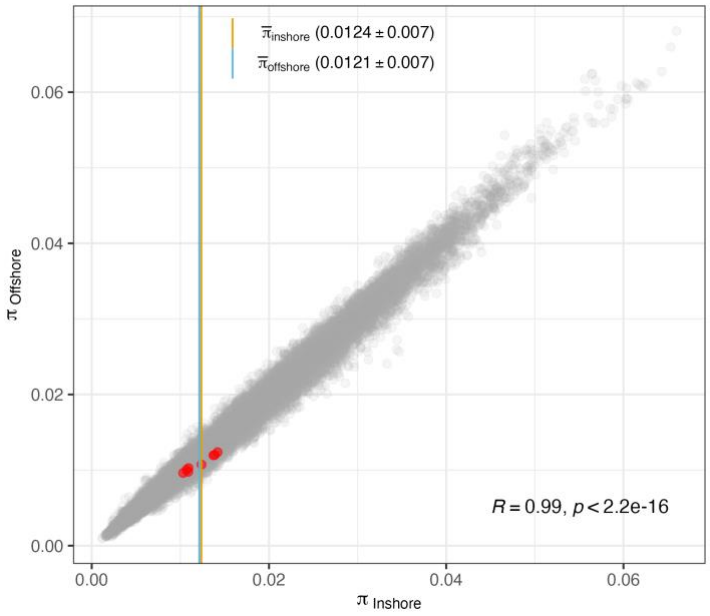
Supplementary Fig 2.12 The visualisation of the estimated relative effective migration surface based on the IBS matrix.



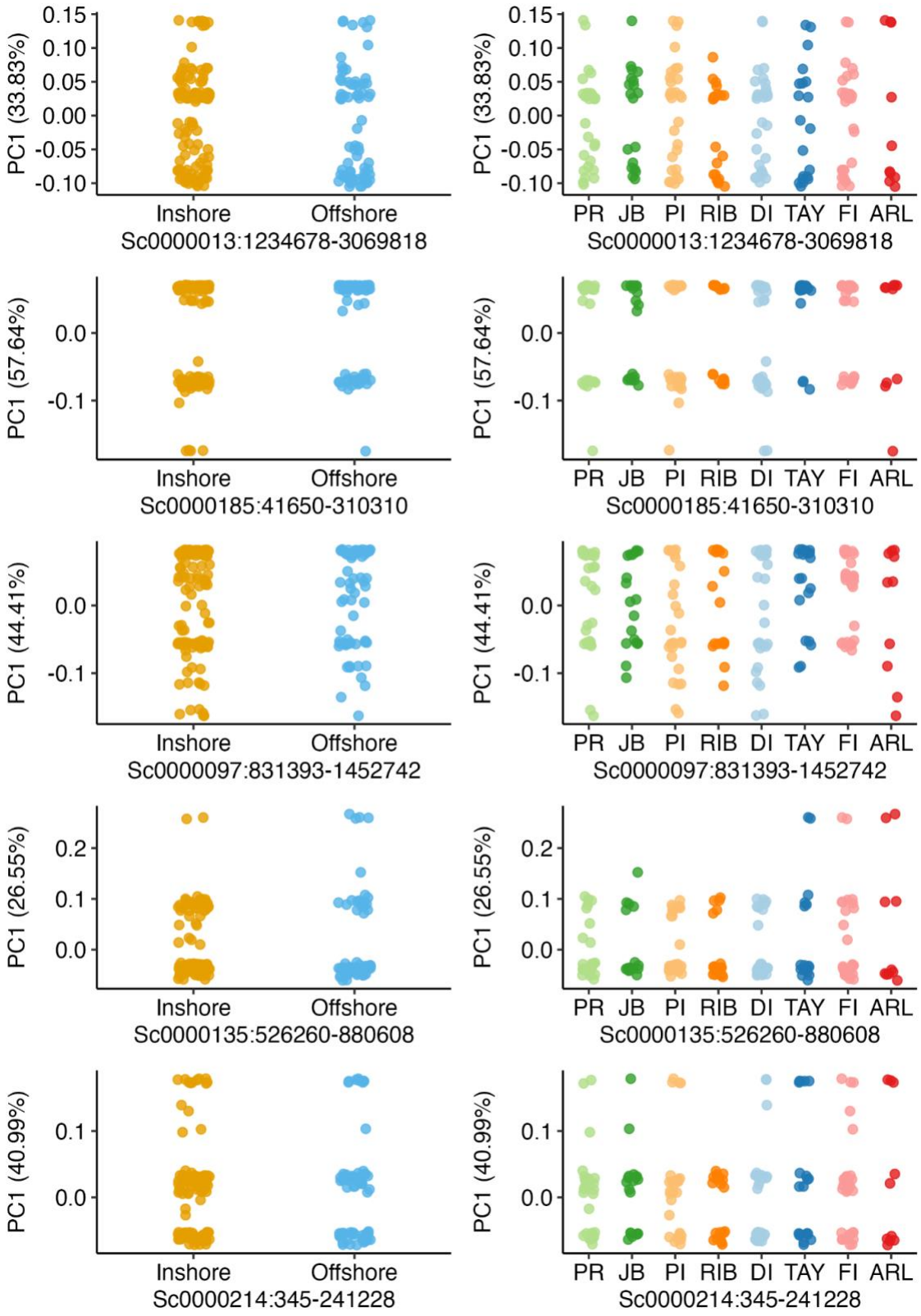
Supplementary Fig 2.13 Characterising highly differentiated region on Locus Sc0000104. a) the population structure inferred by PCAangsd. b) PC1 loading of each sample by inshore/offshore location. c) PC1 loading of each sample by reef location. d) The minor allele frequency between inshore and offshore of SNPs within the locus.



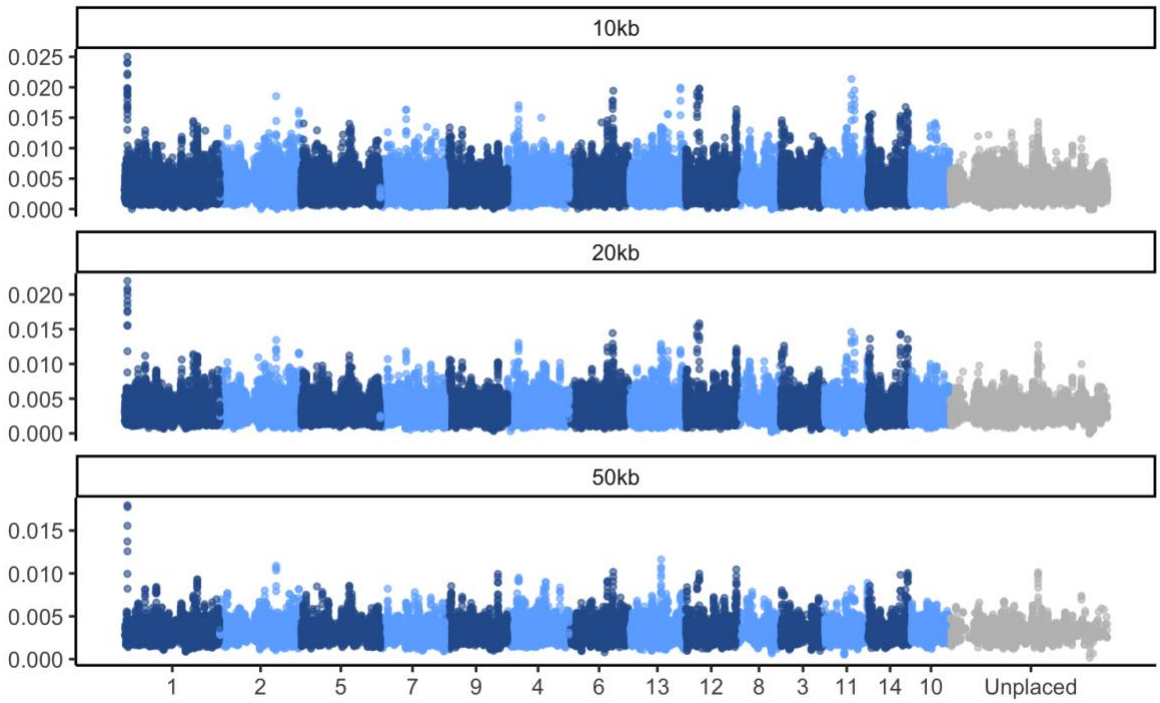
Supplementary Fig 2.14 The genetic diversity distribution of the inshore reefs and offshore reefs at the genomic islands and non-island regions. The significance of the mean difference was tested with the Wilcoxon test (P<0.05).



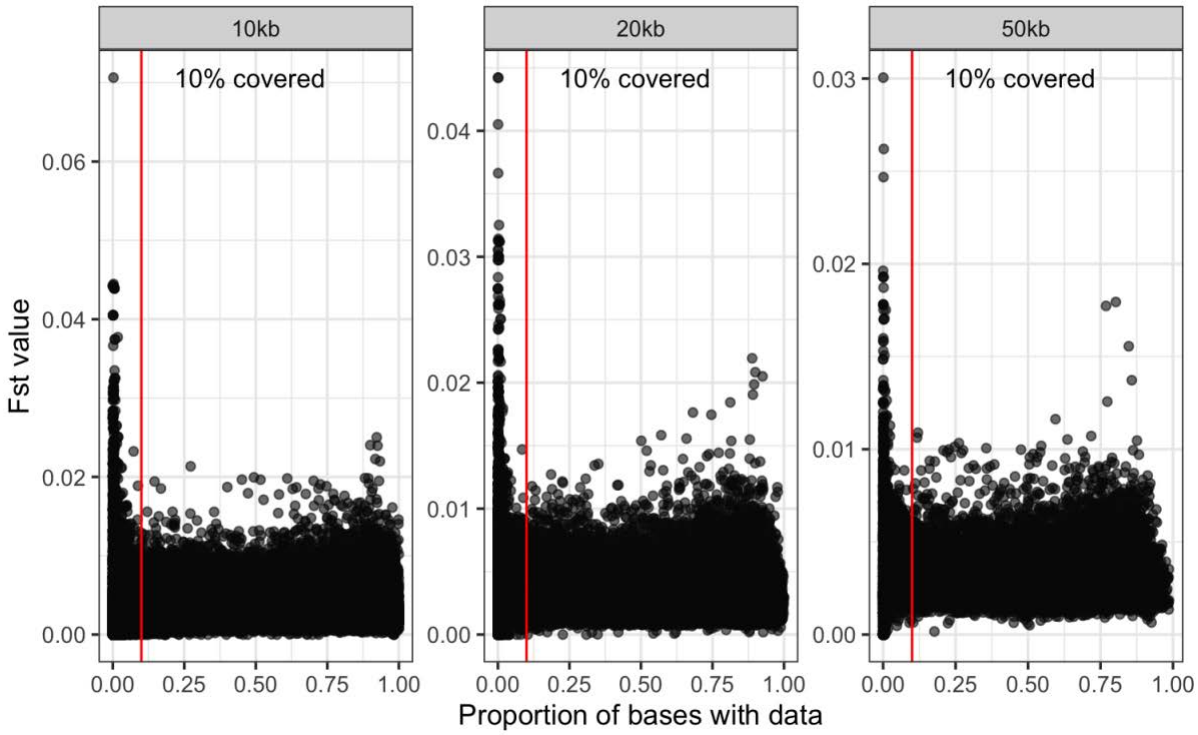
Supplementary Fig 2.15 The correlation of nucleotide diversity in the north inshore reefs and offshore reefs in 20kb sliding windows with red dots present the genomic regions with F_{ST} outliers. The genome-wide averages were indicated with vertical lines.



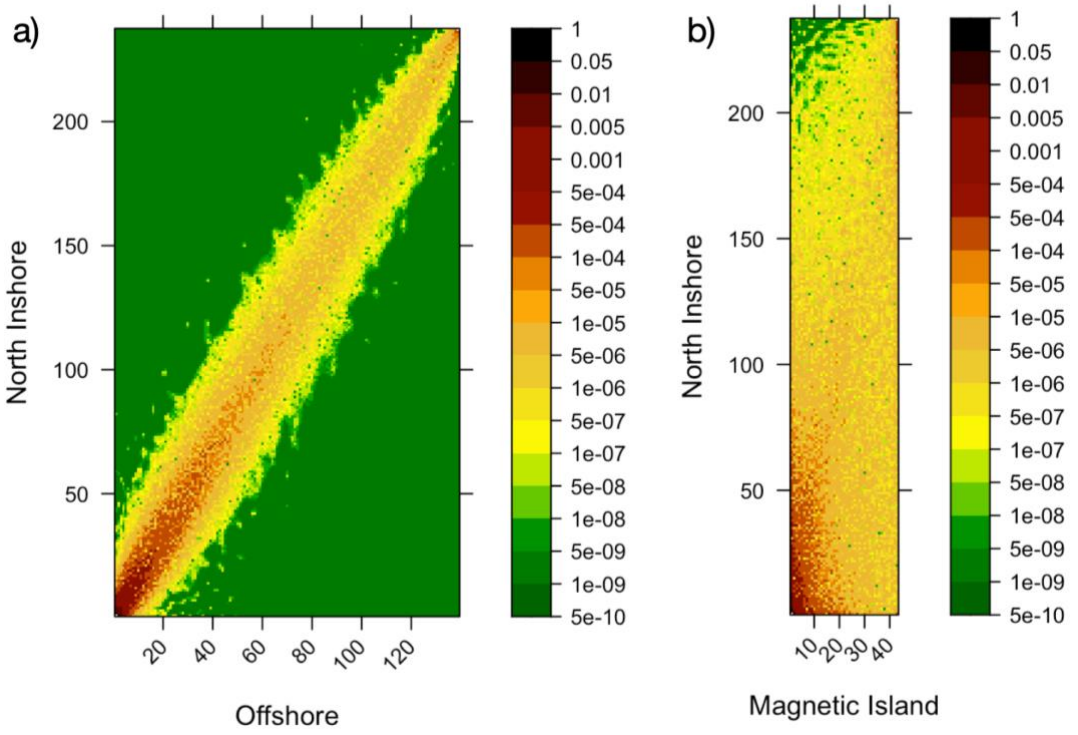
Supplementary Fig 2.16 The PC1 loading of each sample within each region visualised by inshore/offshore or reef locations



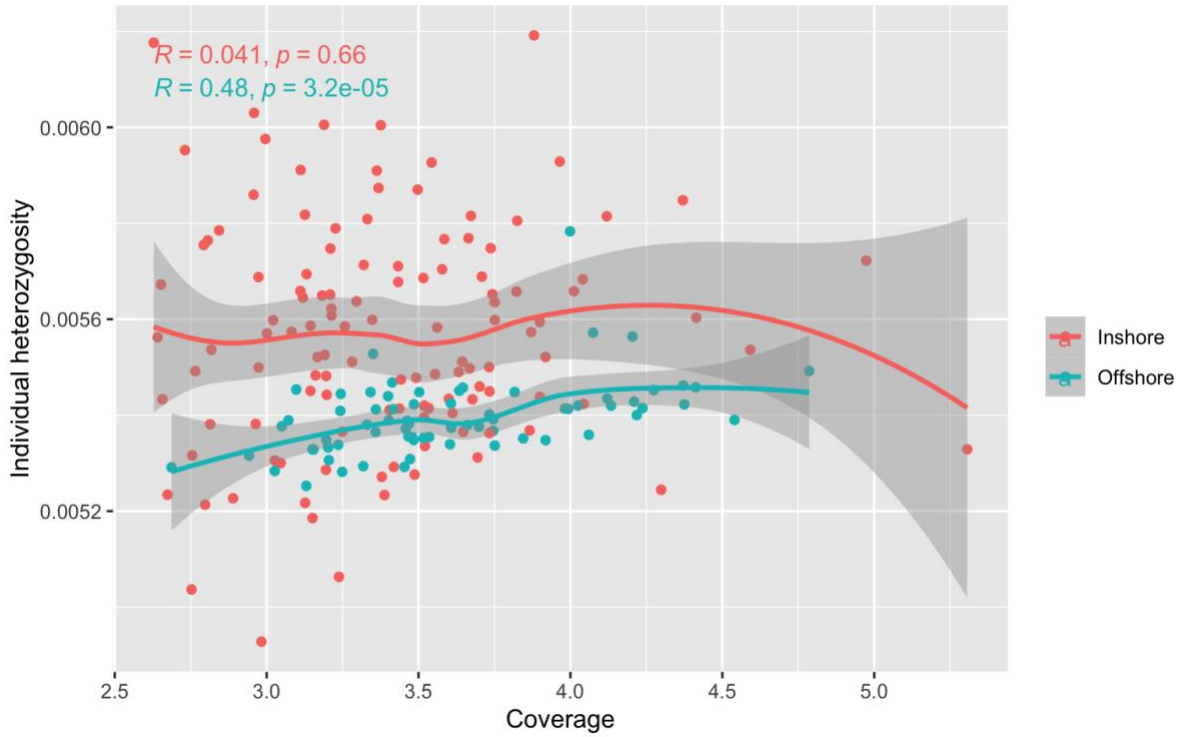
Supplementary Fig 2.17 The genome-wide pattern of F_{ST} divergence using different sliding-window sizes. Each figure displays the genome-wide pattern of F_{ST} between inshore and offshore reefs with 10kb, 20kb, and 50kb window sizes. Windows with less than 10% of sites covered were excluded in plots.



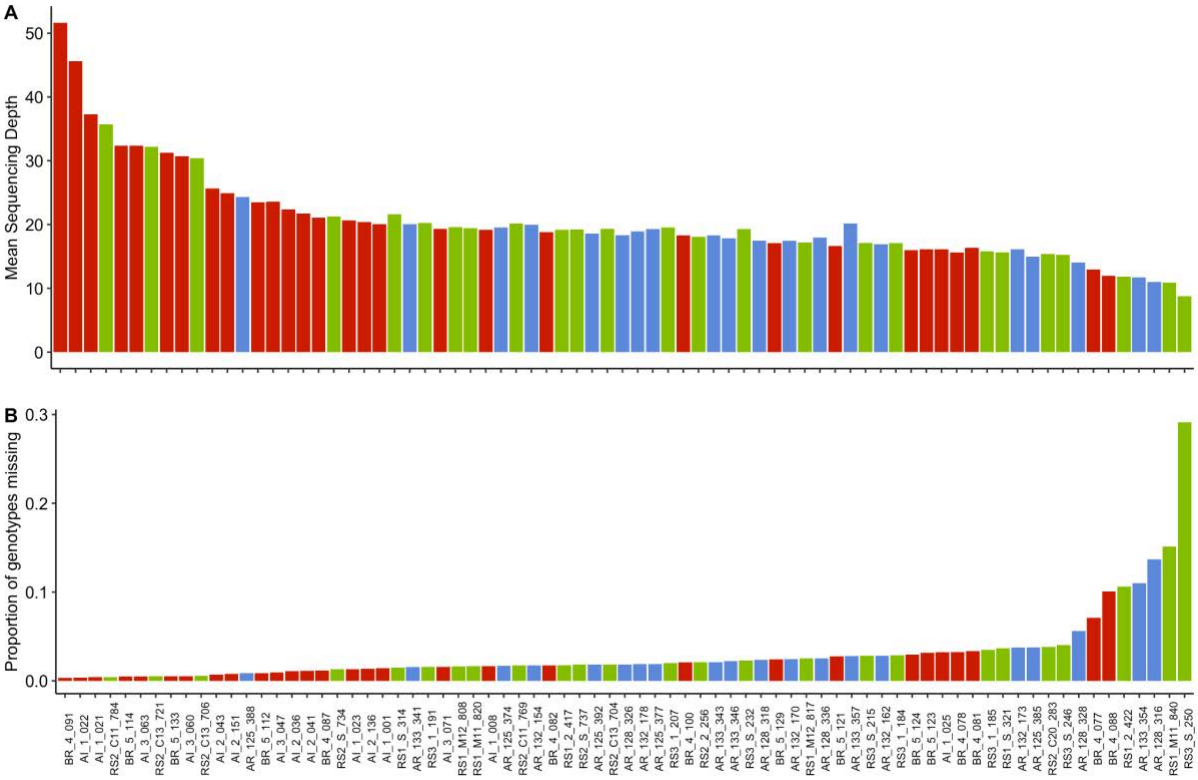
Supplementary Fig 2.18 The distribution of F_{ST} estimates in windows with different proportions of missing data with 10kb, 20kb and 50kb window sizes. Red vertical lines represent the 10% threshold.



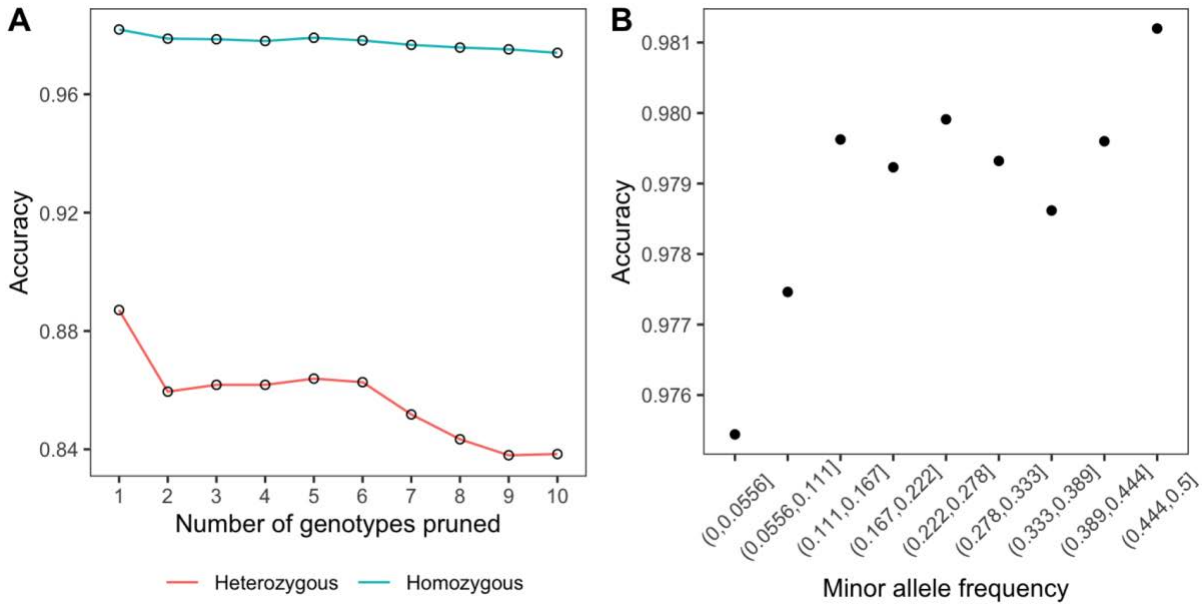
Supplementary Fig 2.19 The joint allele-frequency spectrum (AFS) for a) north inshore and offshore samples; b) Magnetic Island and north inshore samples. Color showing the count of minor alleles for genome-wide SNPs from each group.



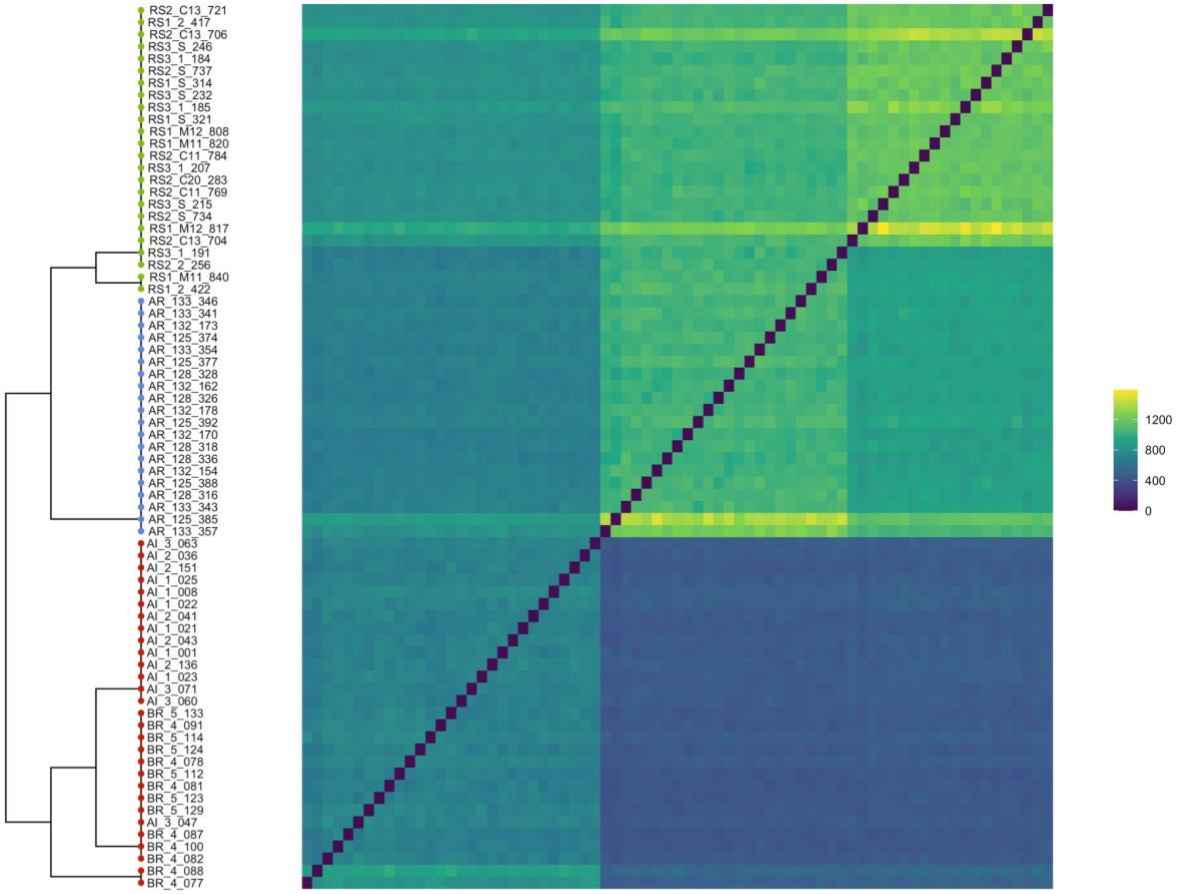
Supplementary Fig 2.20 The relationship between sequencing coverage and individual heterozygosity in each sample from inshore and offshore.



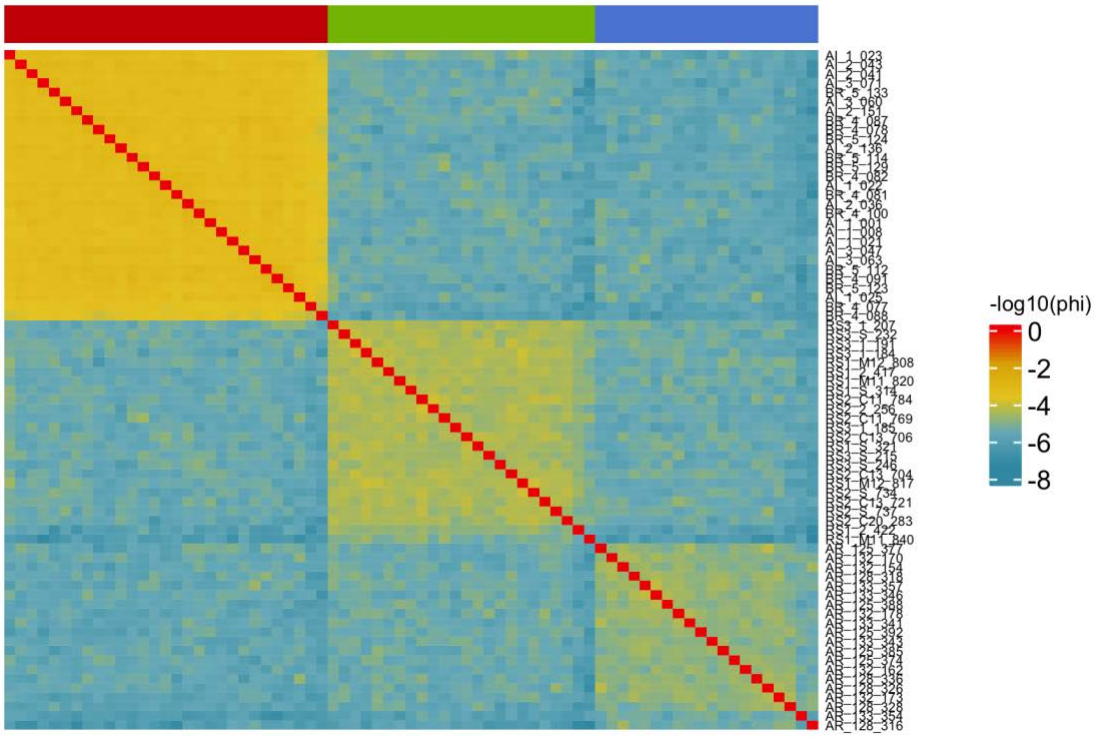
Supplementary Fig 3.1 Mapping depth (A) and genotype missingness (B) of all samples. The colour of dots represents the sample origin population, inshore (red), north offshore (blue), and south offshore (green).



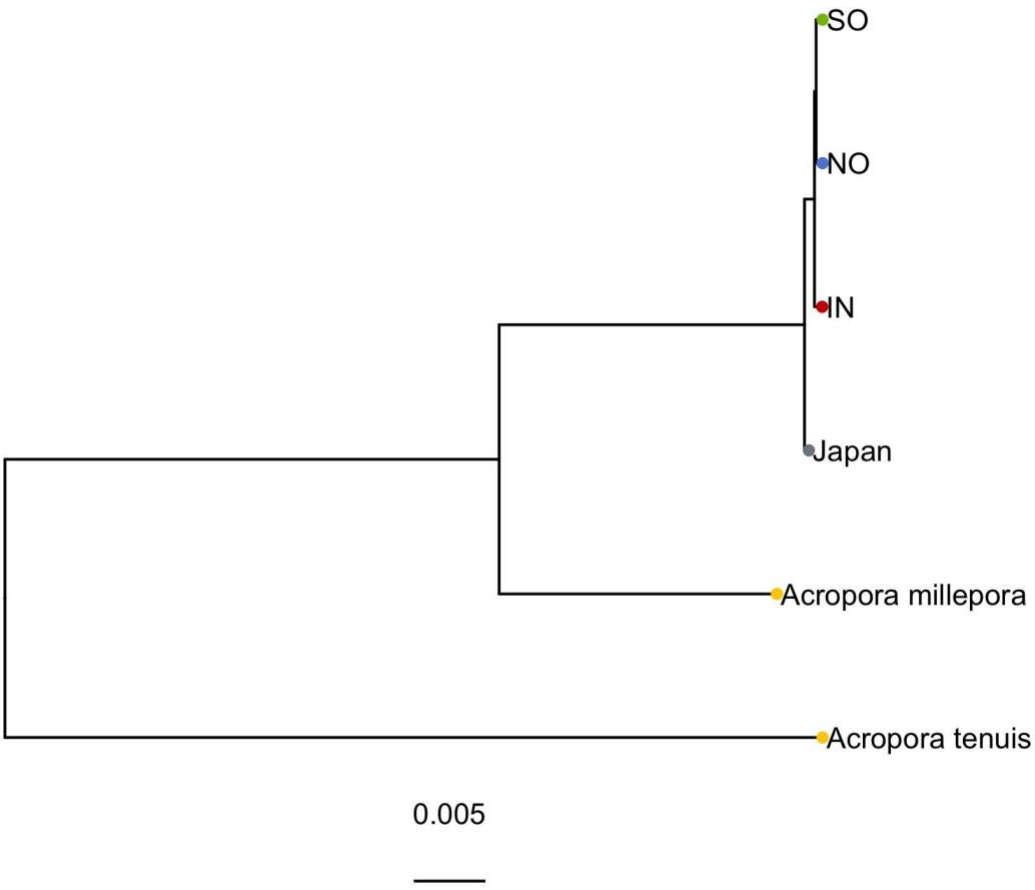
Supplementary Fig 3.2 A. The estimated imputation accuracy at homozygous and heterozygous sites as a function of the number of missing genotypes. B. Estimated imputation accuracy as a function of minor allele frequency



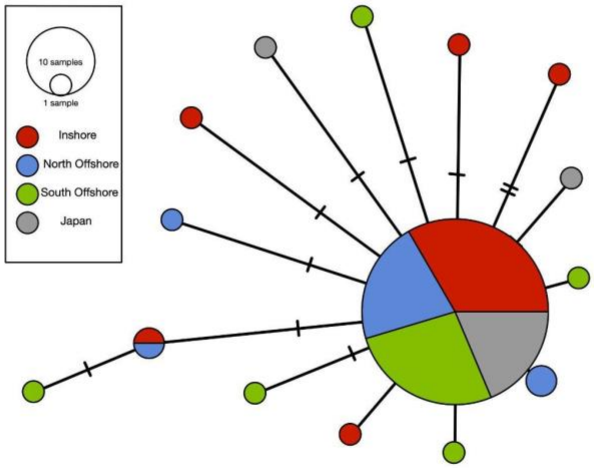
Supplementary Fig 3.3 Sample tree and coancestry heatmap inferred by fineStructure. The tree shows nodes with greater than 99% bootstrap support and samples are coloured according to the location using the standard scheme (Figure 3.1 in the main text). Heatmap shows coancestry between a donor sample (column) and recipient sample (row). Black line along the diagonal shows where column and row samples are the same



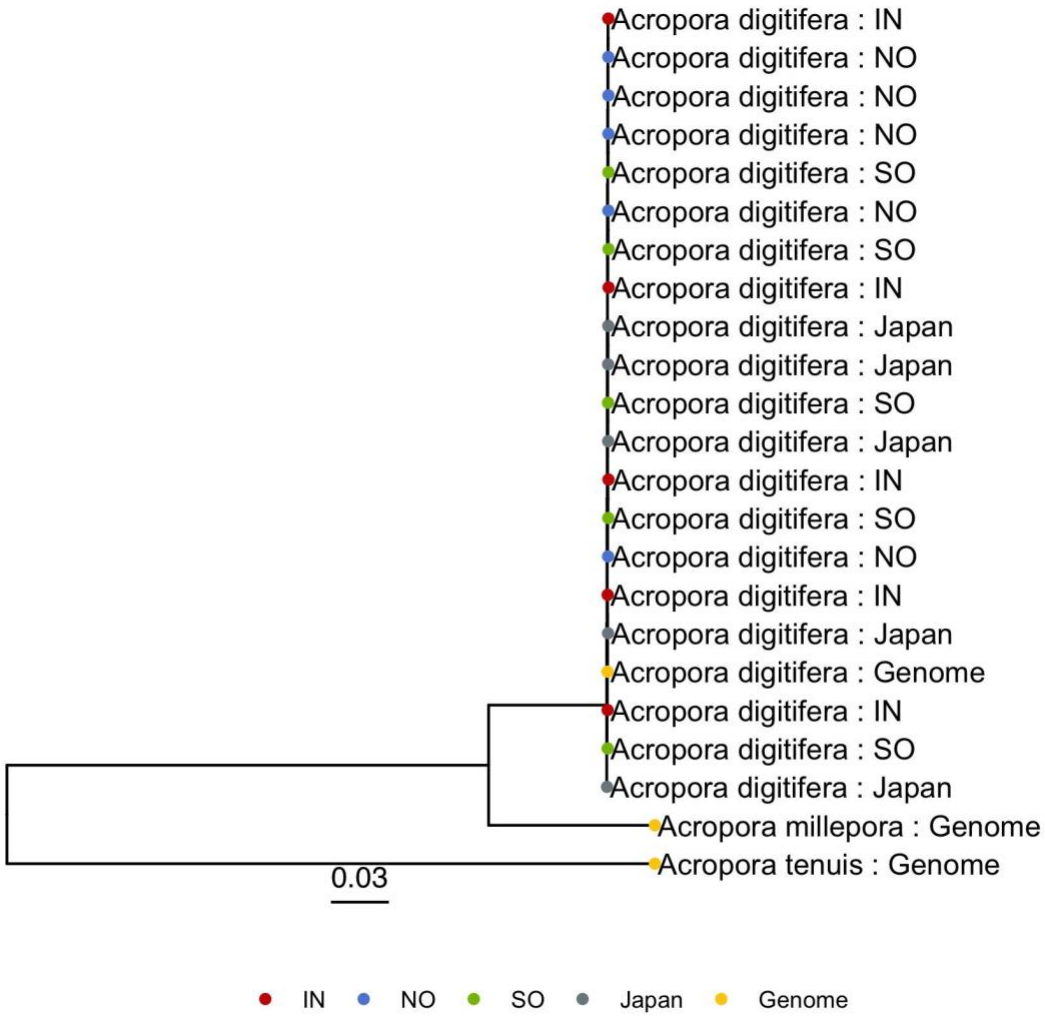
Supplementary Fig 3.4 Clustered heatmap based on pairwise relatedness (phi) inferred by the relative proportion of IBD (identical by descent) segments. Rows and columns both represent samples and use the same clustering. Sample labels are shown for rows and column colours indicate their location using our standard colour scheme (Figure 3.1 main text).



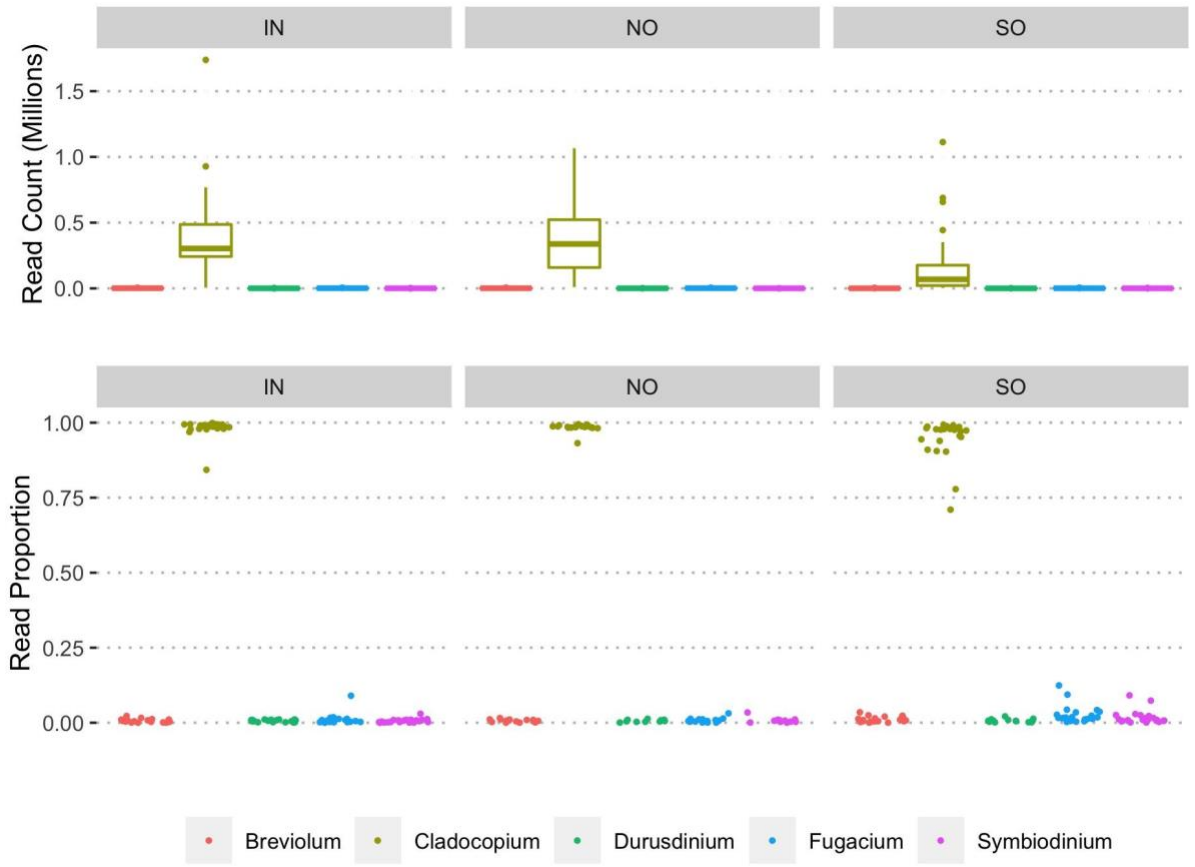
Supplementary Fig 3.5 Phylogenetic tree showing relationships between *A. digitifera* populations from the Kimberley, Western Australia and Ryukyu Archipelago Japan, as well as outgroup species *A. tenuis* and *A. millepora*. Samples are coloured according to location using our standard colour scheme (Figure 3.1 main text) with the addition of grey to indicate samples from Japan and yellow to indicate whole genome sequences. Branch lengths incorporate both fixed substitutions and shifts in allele frequency but have been rescaled to represent substitutions per site. All clades have 100% bootstrap support based on 1000 ultrafast bootstrap replicates.



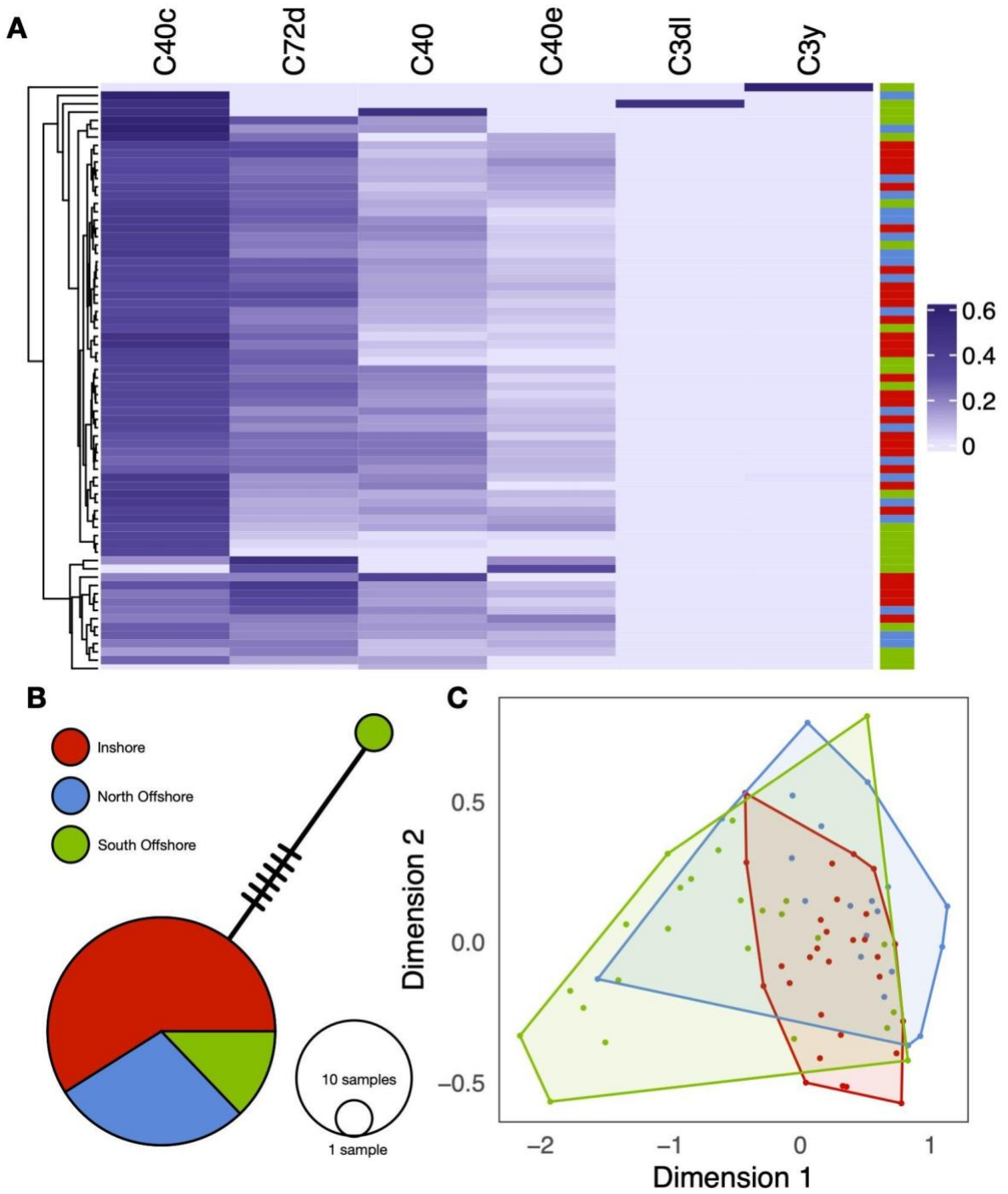
Supplementary Fig 3.6 Haplotype network based on mitochondrial genomes of *A. digitifera* samples from the Kimberley region, Western Australia, and the Ryukyu Archipelago, Japan. Cross bars on edges indicate the number of mutations separating haplotypes while the size of nodes indicates the number of samples with the same haplotype.



Supplementary Fig 3.7 Maximum likelihood tree inferred from consensus UCE and Exon sequences. Sequences include those obtained from published reference genomes for *Acropora millepora* (GCF_004143615.1; (Ying et al. 2019)), *Acropora digitifera* (GCA_014634065.1, (Shinzato et al. 2020)), and *Acropora tenuis* (http://reefgenomics.org/aten/; (Cooke et al. 2020)) as well as representative population genomic samples from our study (NO, SO, IN) and from Japan (NCBI Bioproject PRJDB4188; (Shinzato et al. 2015)). Japanese samples correspond to SRA accessions (DRR099286,DRR099287,DRR099303,DRR099291,DRR099351). Samples from our study included (IN: AI_2_151, AI_3_071, AI_2_043, AI_3_047, BR_5_129; NO: AR_128_336, AR_132_154, AR_133_343, AR_132_170, AR_125_374; SO: RS1_M12_817,RS3_1_207, RS2_C13_721, RS1_S_321, RS3_1_191).

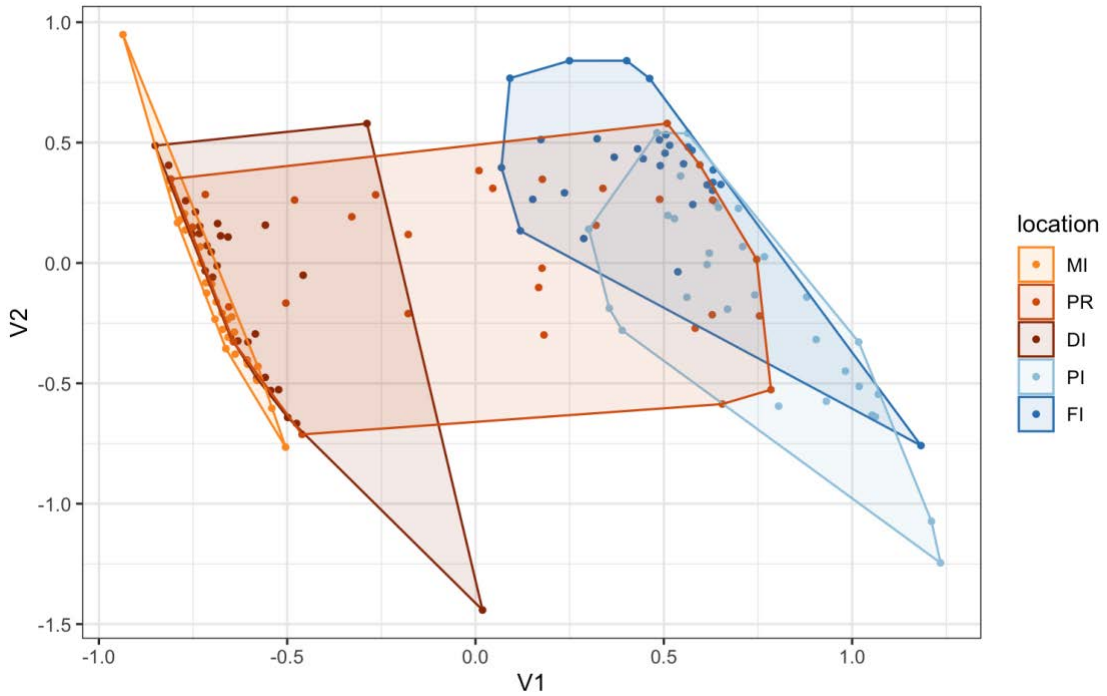


Supplementary Fig 3.8 Summary of reads classified as Symbiodiniaceae using Kraken. Both top and bottom plots show the spread of values measured across individual samples. Read counts are shown as sample totals (top) and as proportion of the sample total (bottom) across five genera of Symbiodiniaceae. Samples are plotted separately for each location (IN: Inshore, NO: North Offshore, SO: South Offshore).

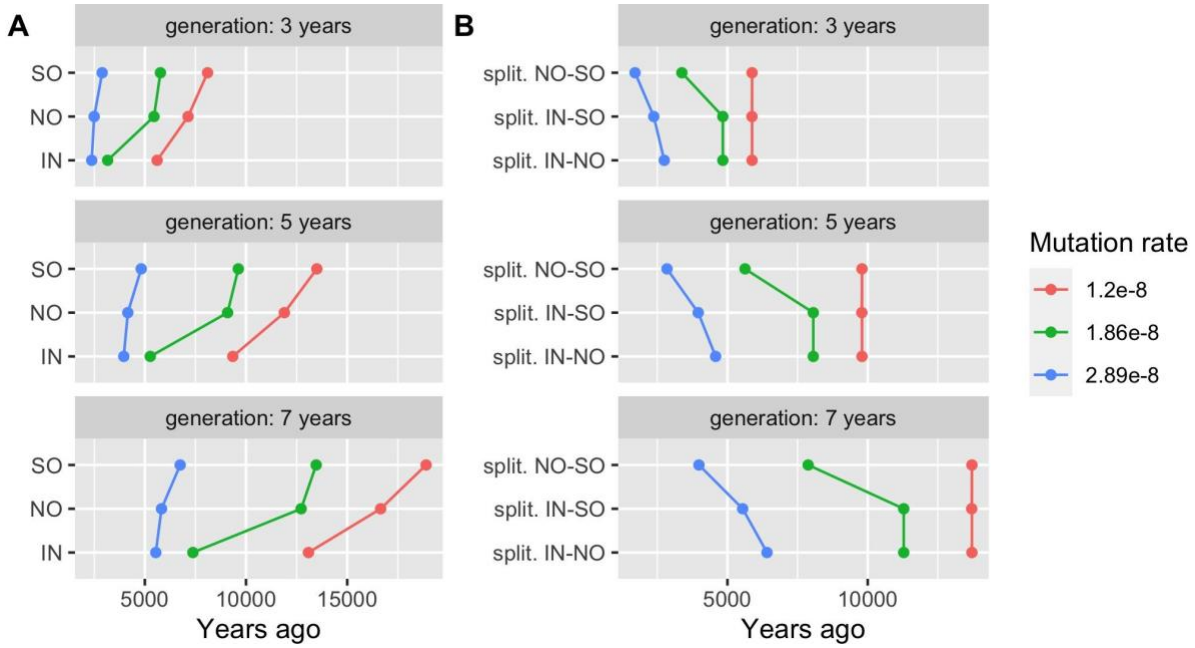


Supplementary Fig 3.9 Diversity of sequences related to the dominant symbiont genus, *Cladocopium*. A. Heatmap of read counts mapping to symportal ITS2 reference type sequences. Rows represent coral samples, and the columns show the detected ITS2 types from read mapping. Coloured strip on the right indicates the location of origin for each sample using the colour scheme shown in B. B. Haplotype network based on mitochondrial sequences for 41 samples for which sufficient reads were available to allow consensus calling. Edge cross bars indicate the number of mutations separating haplotypes and the size of nodes indicates the number of samples with a given haplotype. C. Multidimensional scaling (MDS) plot based on pairwise distances between samples calculated using D_2^S statistics. D_2^S statistics are calculated based on *k*-mer counts in reads of *Cladocopium* origin (that map to the *Cladocopium goreaui*

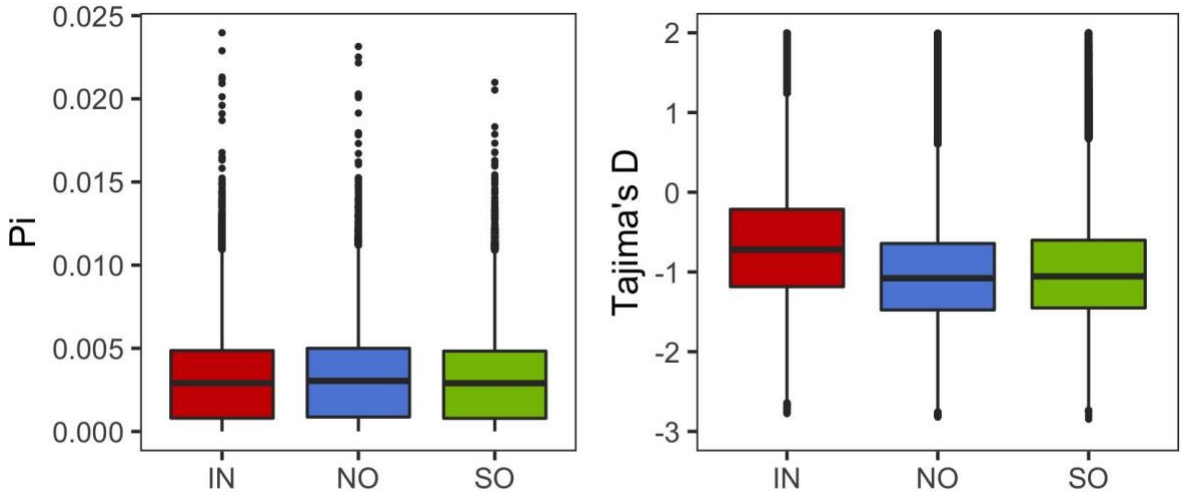
genome). Convex hulls enclose points from each location and are coloured according to our standard location color scheme (see B).



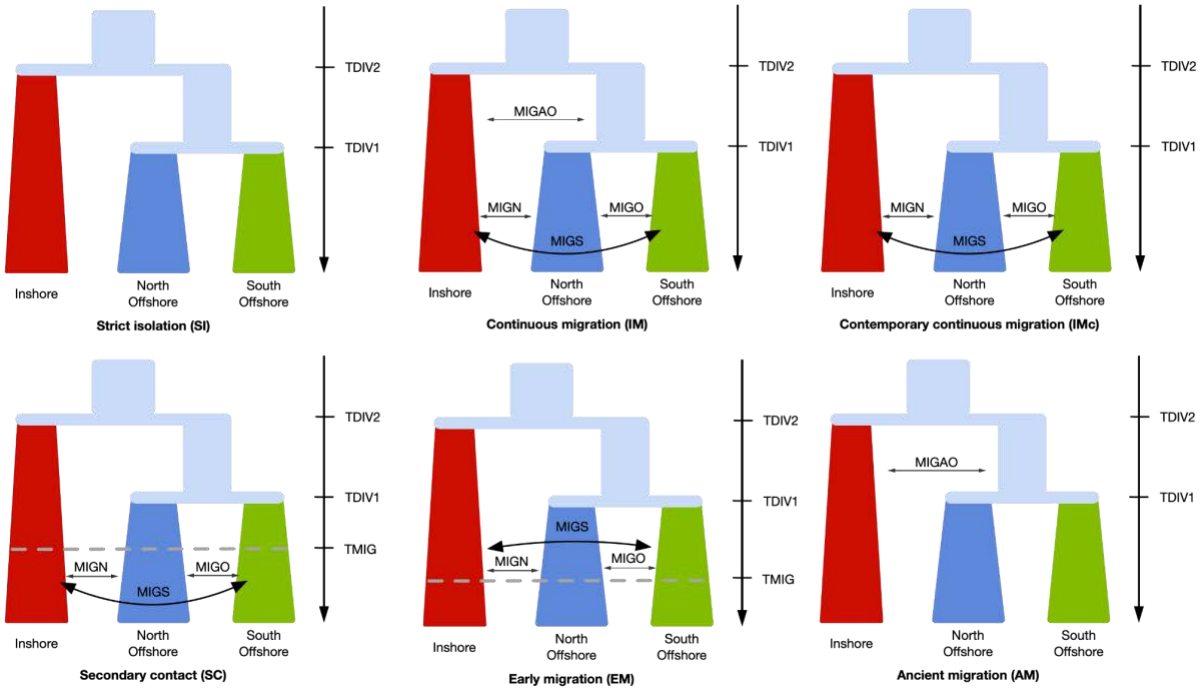
Supplementary Fig 3.10 Multidimensional scaling (MDS) plot based on pairwise distances between Symbiodiniacea reads extracted from whole genome sequencing of *Acropora tenuis* samples from the Great Barrier Reef (Cooke et al. 2020). Distances are calculated using D_2^S statistics as in Figure 3.1E in the main text. Samples are labelled by location (n=30 per site) and colour coded according to the scheme used in (Cooke et al. 2020) Red/Brown colours represent plume locations and blue colours represent marine. Location codes are, MI (Magnetic Island), PR (Pandora Reef), DI (Dunk Island), PI (Pelorus Island), FI (Fitzroy Island). Note that partitioning of clusters shown in this plot recapitulates patterns shown in the mitochondrial haplotype network presented in Figure 2C from (Cooke et al. 2020)



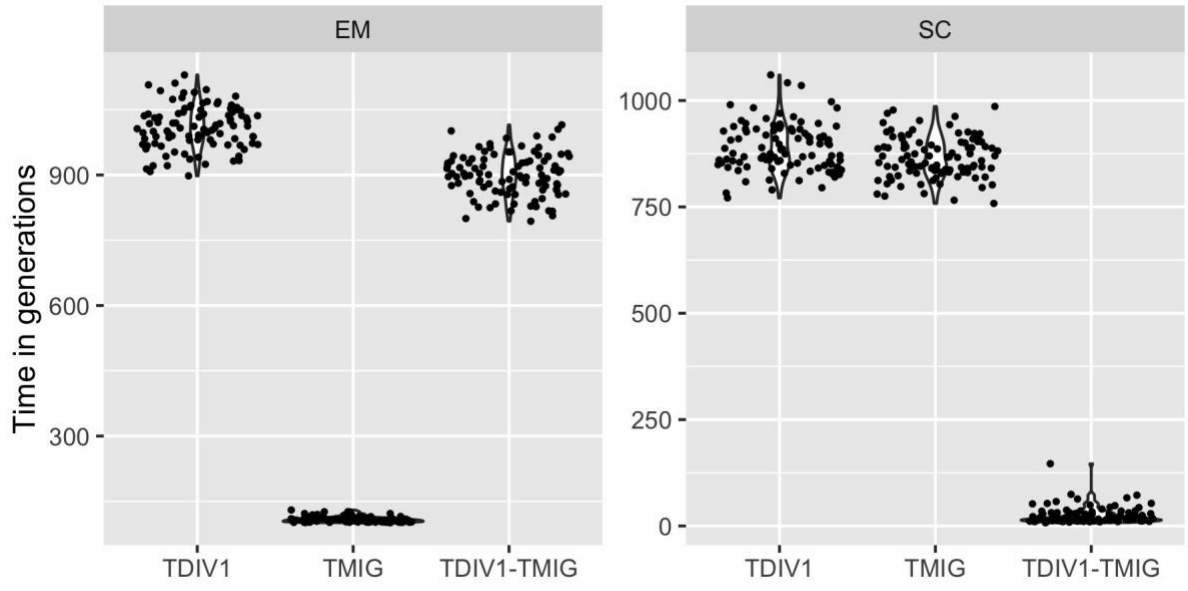
Supplementary Fig 3.11 Variation in the estimated timing of key demographic events under different mutation rates (as mutations per base per generation) and generation times. All estimates were obtained using SMC++. A. the bottleneck time of inshore (IN), north offshore (NO), and south offshore (SO) *A. digitifera* populations. B. the split time between each pair of populations.



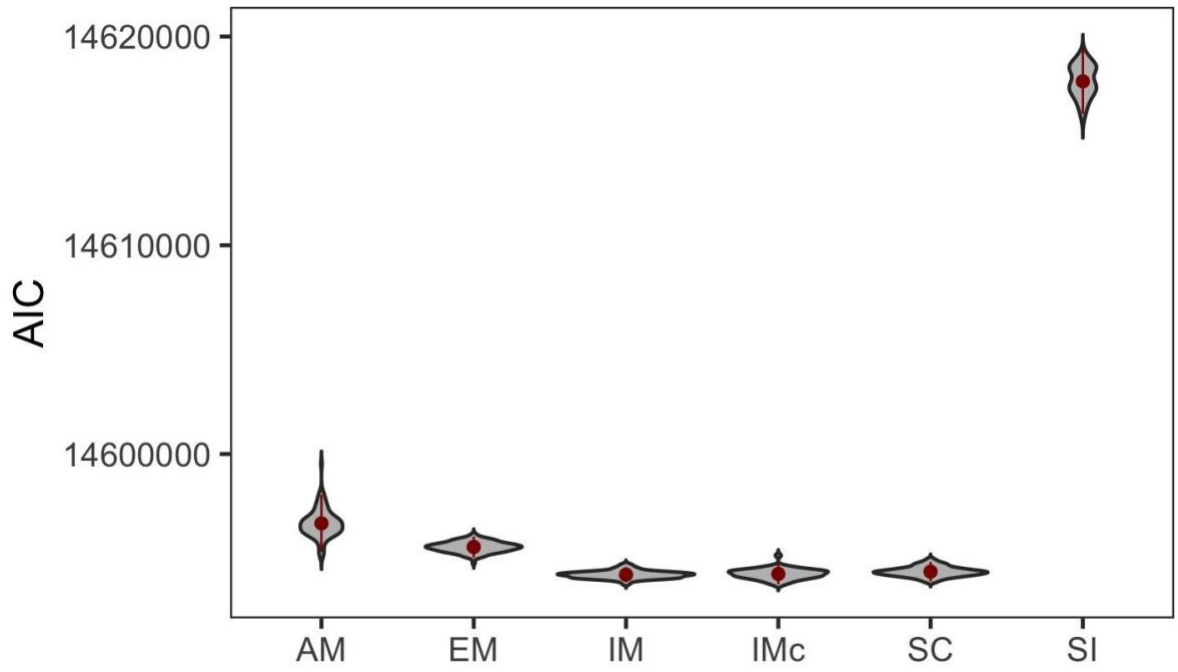
Supplementary Fig 3.12 Boxplots showing the genome-wide distribution of nucleotide diversity (Pi; left) and of Tajima's D (right). Both plots show results for each of the three populations separately and use our standard color scheme to denote location (Figure 3.1 main text).



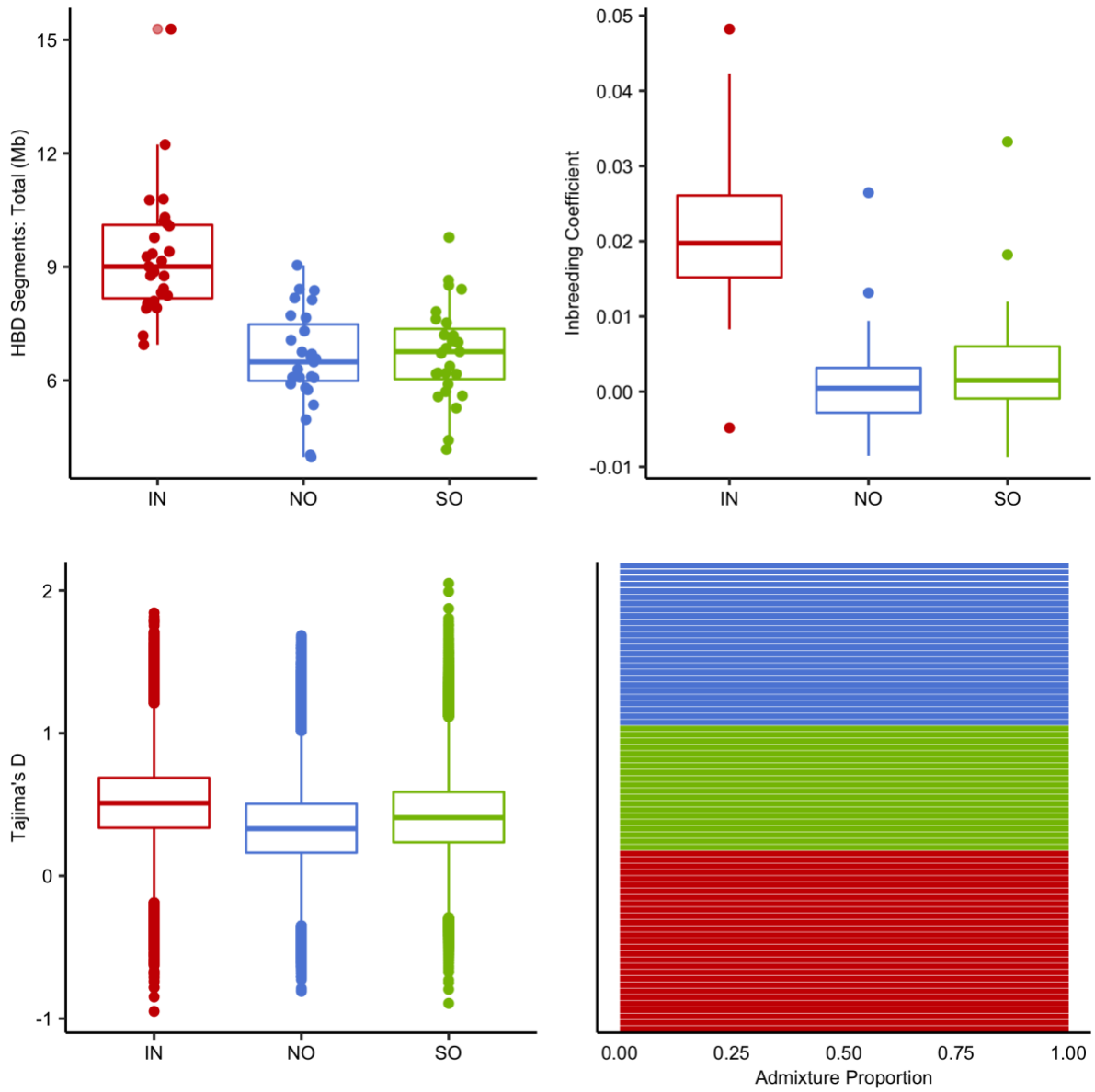
Supplementary Fig 3.13 Schematic diagram of six alternative demographic models used in fastsimcoal2. Time is shown from most ancient (top) to present day (bottom). All models were allowed to have an exponential growth rate. The parameters TDIV1 and TDIV2 represent the time of offshore-offshore divergence and inshore-offshore divergence, respectively. Moving forward in time, TMIG represents the time at which migration starts. In models with TMIG there is no migration prior to TMIG. In models without TMIG the migration parameters persist only during one of the time intervals delineated by TDIV1 and TDIV2.



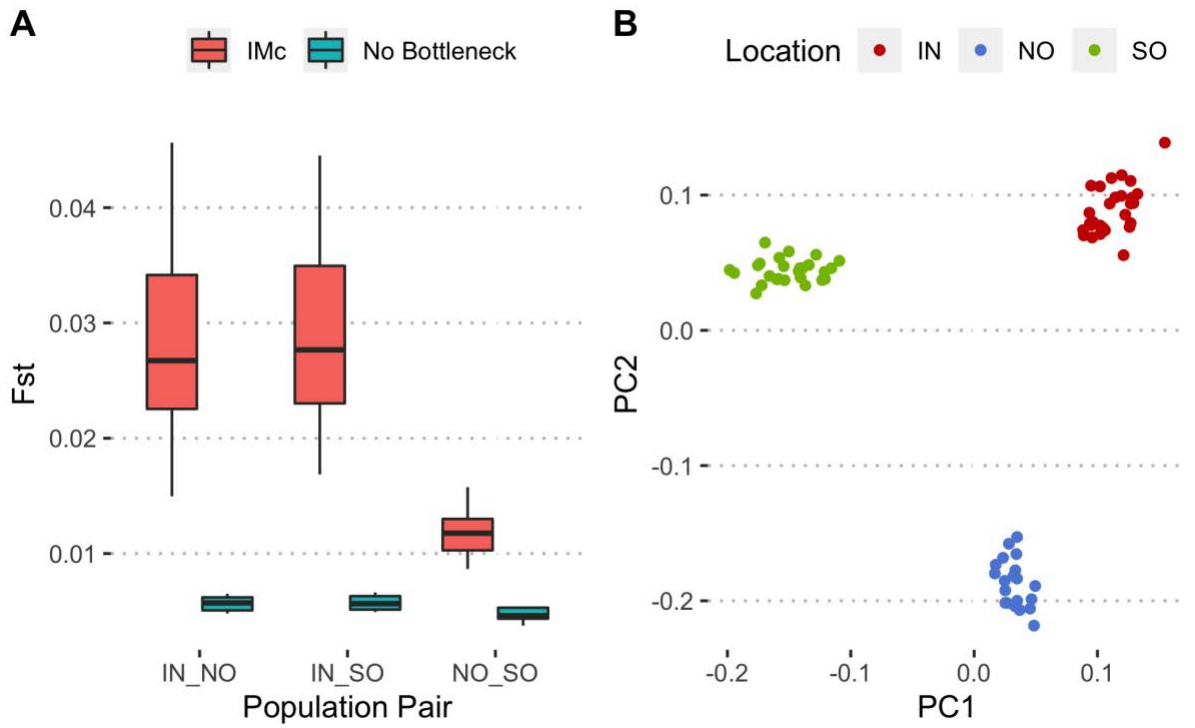
Supplementary Fig 3.14 Estimates from fastsimcoal2 for the time of offshore-offshore divergence (TDIV1) and the migration start time (TMIG) in models EM and SC (see Supplementary Fig 3.12). Points are randomly jittered on the x-axis to avoid overplotting and show the distribution over 100 independent fastsimcoal2 runs.



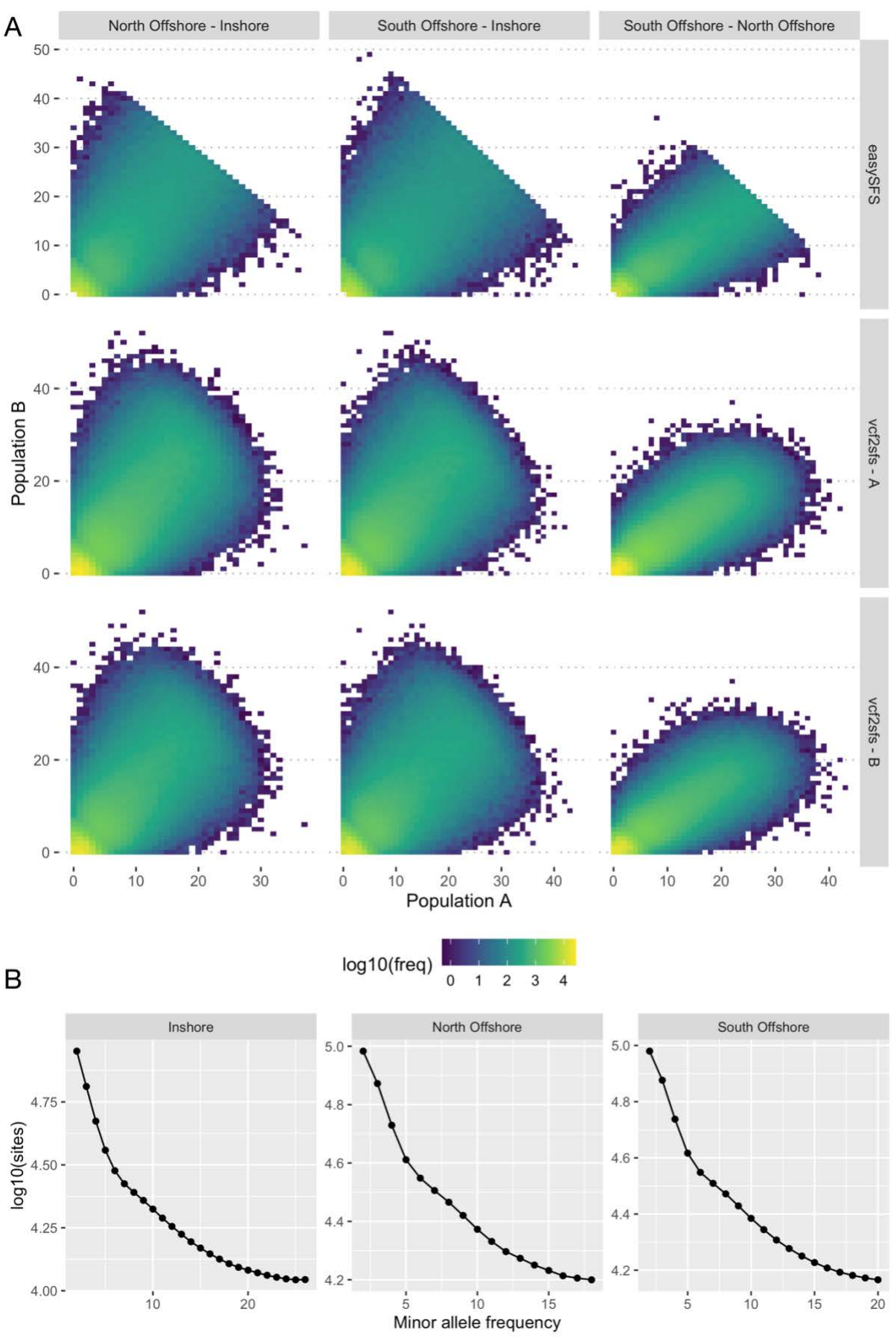
Supplementary Fig 3.15 Distributions of AIC values from 100 independent fastsimcoal2 runs for each model described in Supplementary Fig 3.12.



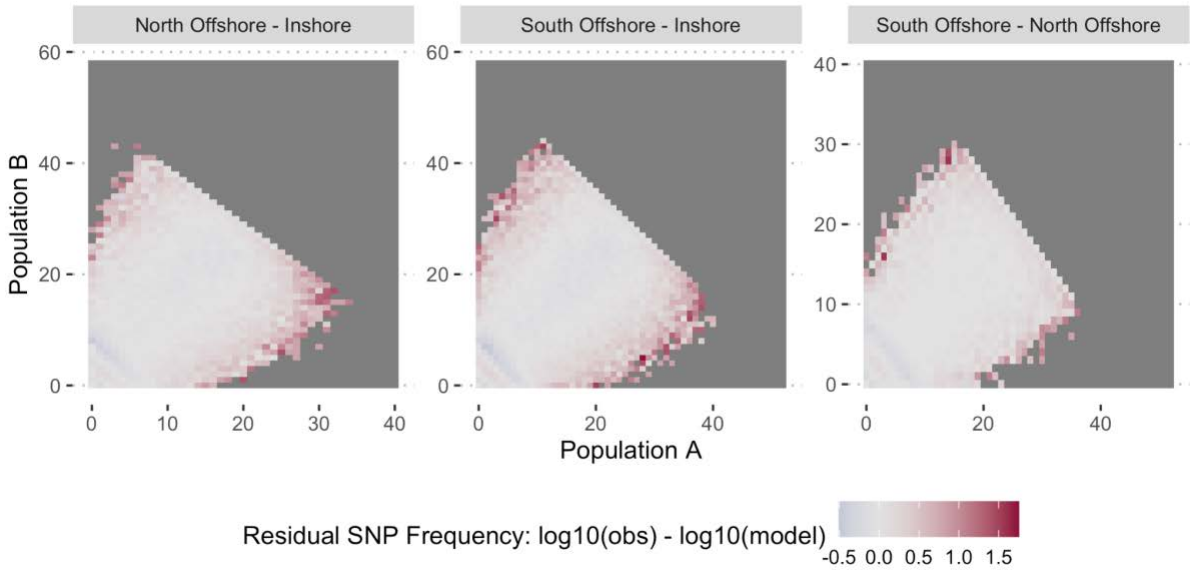
Supplementary Fig 3.16 Population genetic statistics and Admixture proportions calculated based on simulated data under the best fitting model from fastsimcoal2 (model IMc in Supplementary Table 3.10). Sample locations are named and coloured according to our standard scheme (main text Figure 3.1). Boxplots show the distribution of values from 50 simulation runs with fastsimcoal2 based on independent draws across the error range of model parameters. A single representative Admixture plot is shown as all simulations produced near-identical plots.



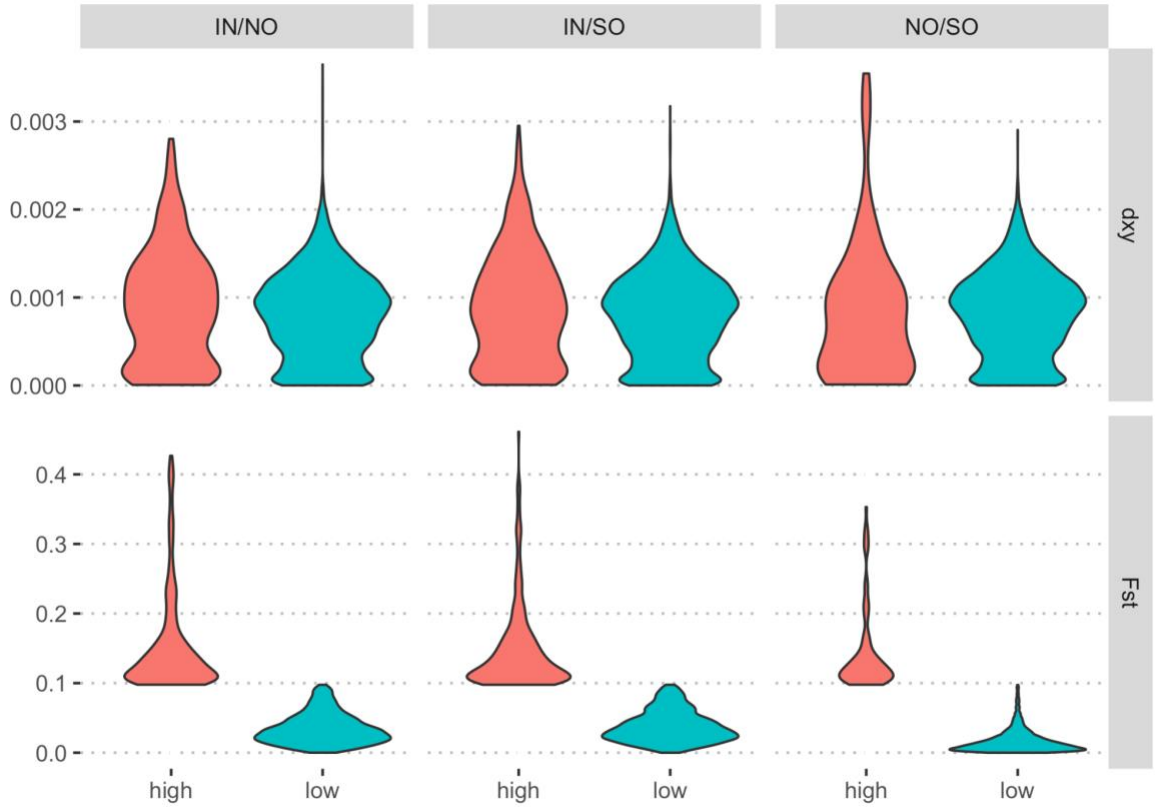
Supplementary Fig 3.17 Influence of the bottleneck on population structure and divergence. A. Boxplots of pairwise Fst (Hudson) for the full IMc model compared with a model with constant population size (No Bottleneck). Spread of values is from 10 independent simulations for each model. B. PCA showing population structure under the No Bottleneck model. PCA shows results for one simulation. Other replicates displayed qualitatively similar results.


A
North Offshore - Inshore
South Offshore - Inshore
South Offshore - North Offshore
easySFS
vcf2sfs - A
vcf2sfs - B
Population B
Population A
log10(freq)
0 1 2 3 4
B
Inshore
North Offshore
South Offshore
log10(sites)
Minor allele frequency

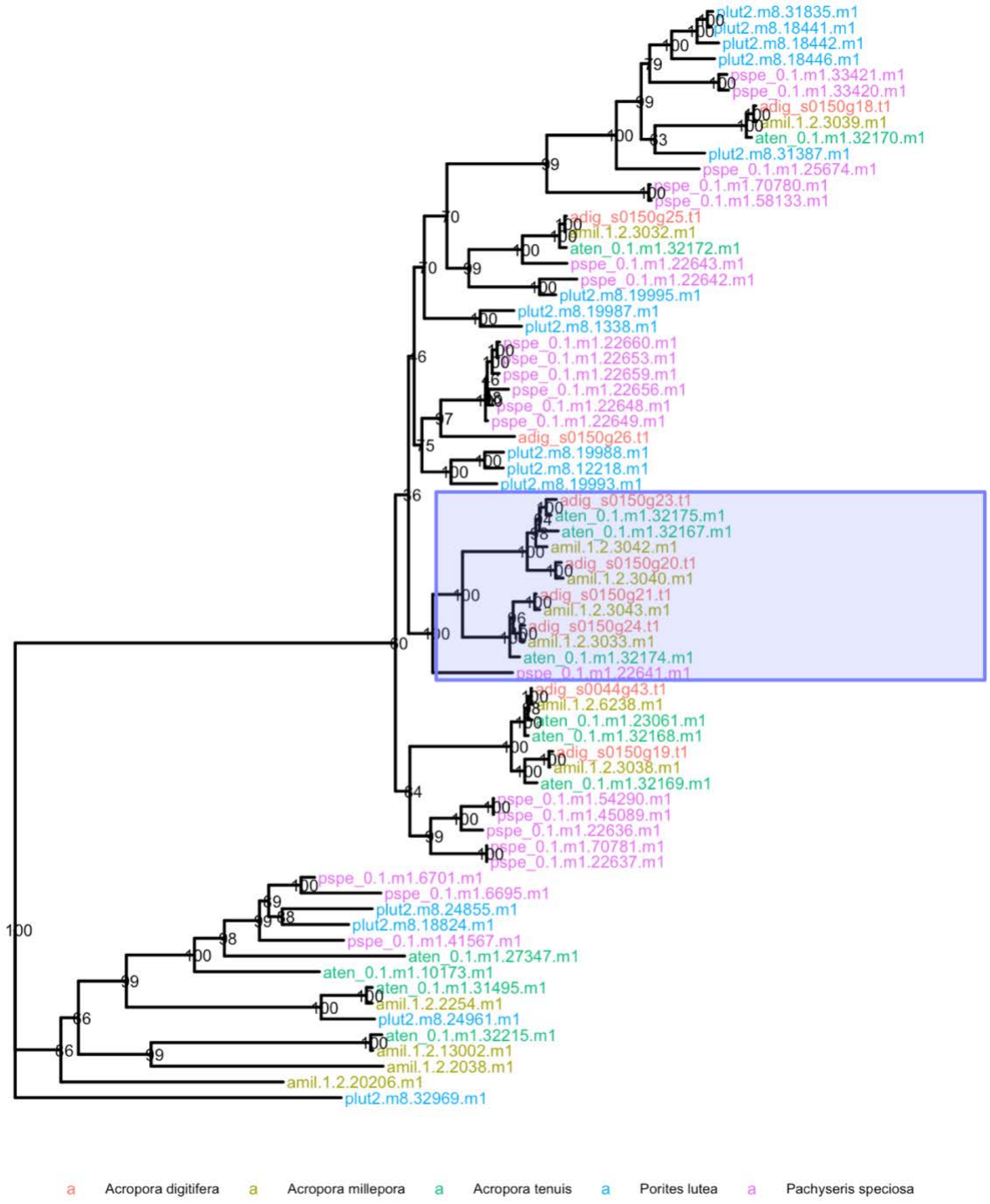
Supplementary Fig 3.18 A. Two dimensional folded site frequency spectra for all pairs of Western Australian *A. digitifera* populations. Labels (top) are written as population A - population B, where the allele frequency in A is shown on the horizontal axis and the allele frequency in B is shown on the vertical axis. Spectra resulting from three different vcf to SFS conversion methods and input datasets are shown. easySFS is the SFS used throughout the manuscript. vcf2sfs - A is generated based on phased input data using scripts vcf2sfs, foldSFS and SFSTools.R (as described in methods). vcf2sfs - B was generated using the same pipeline as vcf2sfs - A but after the removal of sites within 500kb of putative sweep regions. Monomorphic sites (lower left corner) have been removed to improve the colour scale. Note that highly diverged barrier loci should be visible as high frequency bands along horizontal/vertical axes under a scenario of ancient divergence with secondary contact and heterogenous gene flow. B. 1D projections of the SFS generated using easySFS. Note that a broad “shoulder” is visible in all plots between MAF 5-10 and this is also visible as a dark diagonal band in the lower left corner of all plots in A.



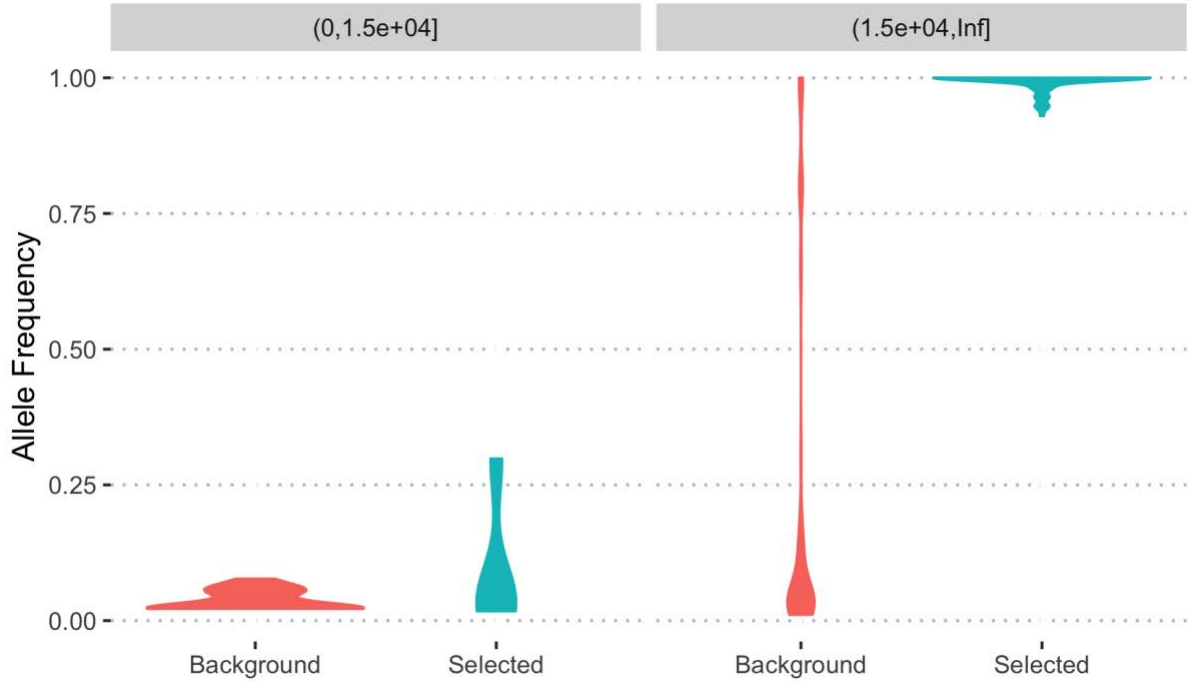
Supplementary Fig 3.19 Residuals plots showing difference between observed SFS and modeled values based on the best fit to the IMc model. Labels (top) are written as population A - population B, where the allele frequency in A is shown on the horizontal axis and the allele frequency in B is shown on the vertical axis. White regions show areas of good fit. The lower left of all plots shows alternating grey and red bands reflecting deviations between modelled and observed spectra in this region. Note that the edges of the SFS are supported by the least data resulting in more noise (red; observed values too high or grey; observed values too low resulting in missing data).



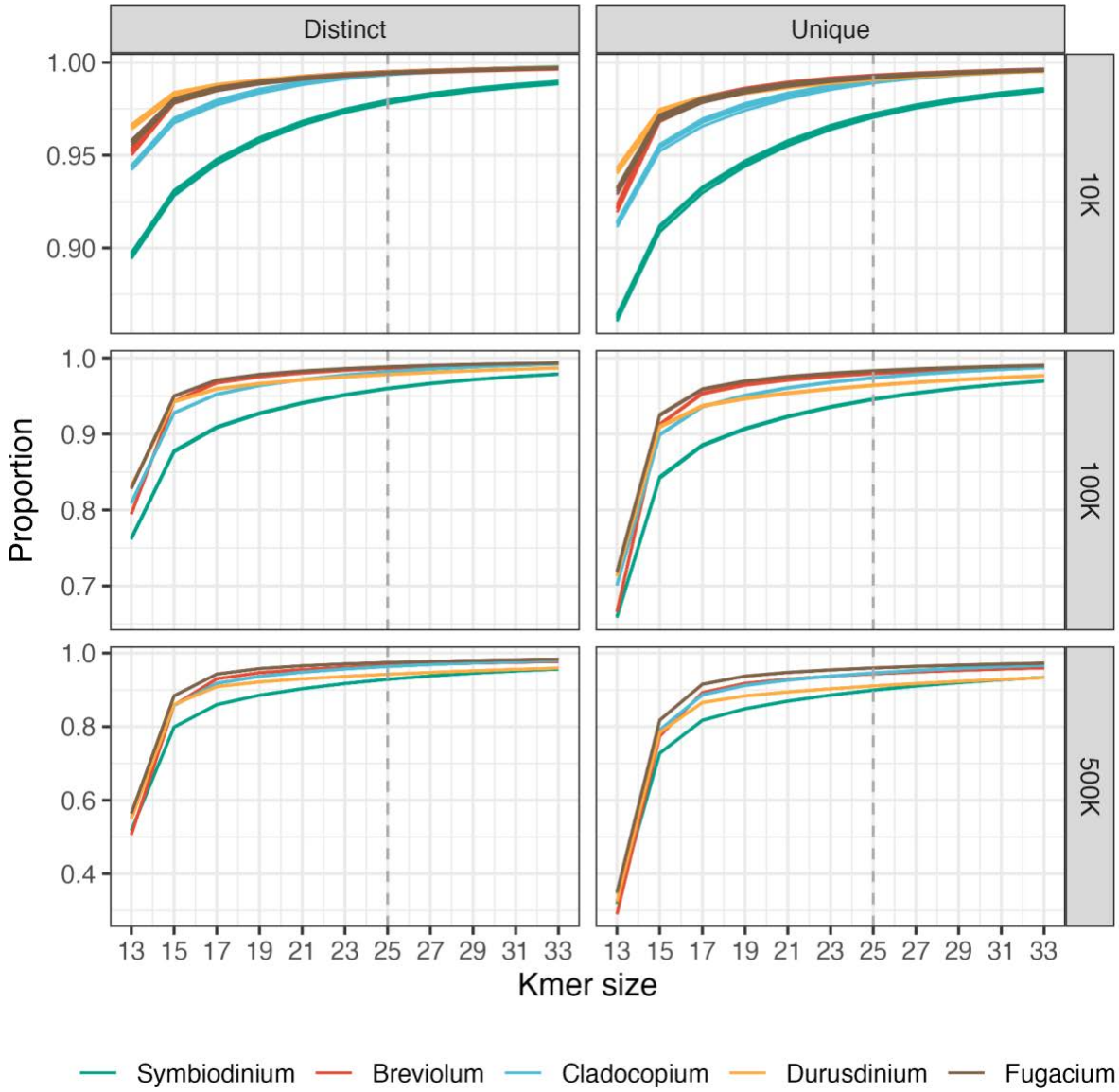
Supplementary Fig 3.20 Comparison of the distribution of F_{ST} and D_{XY} calculated in 50kb sliding windows across the genome. All windows are classified as having either "high" F_{ST} (top 5%) or "low" (remainder). Violin plots show the distribution of F_{ST} and D_{XY} for all windows separated into these categories, and for all pairs of populations. A Wilcoxon-Rank test performed using the wilcox.test function in R for pairs IN/NO and IN/SO was significant (p=0.008; p=0.02 respectively), indicating an increase in D_{XY} in regions of elevated F_{ST}, the magnitude of change was small (11.8%, 10% of the mean respectively) for these pairs.



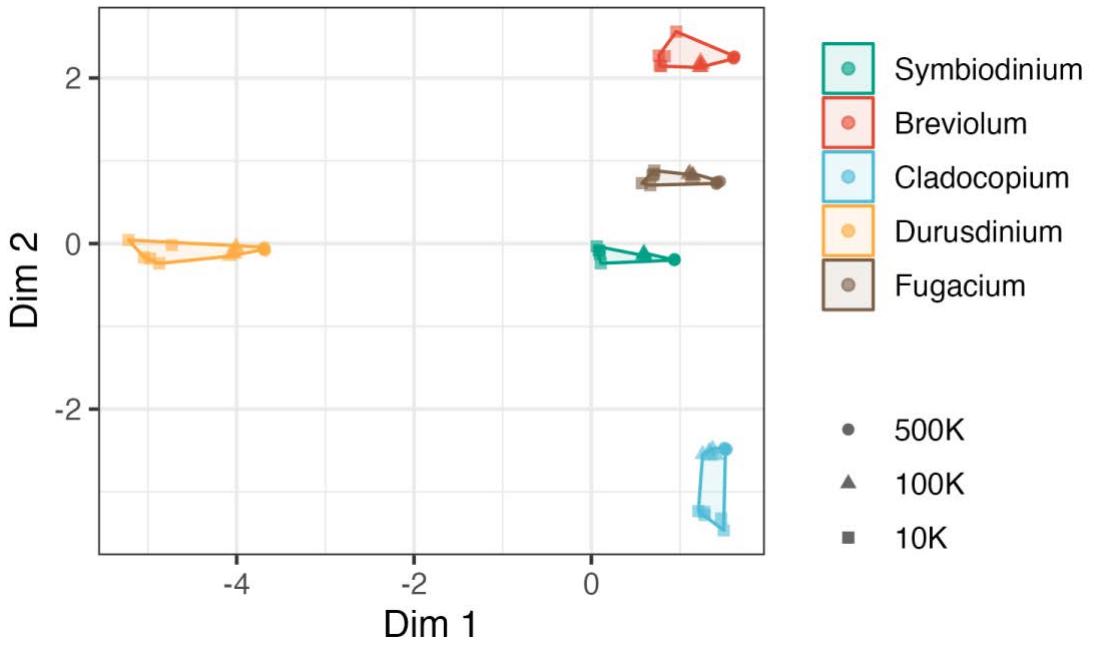
Supplementary Fig 3.21 Phylogenetic relationships among haem peroxidases in representative coral genomes. Species chosen include three representatives of the genus *Acropora* and two outgroups, Porites lutea and Pachyseris speciosa. Highlighted clade includes four genes from the peroxinectin locus in *A. digitifera* that was examined in detail in the main text. All genes within the highlighted clade form clusters (closely spaced within the genome) in their respective species. The phylogeny shown is a subtree of the full phylogeny of haem peroxidases that includes all 8 members of the co-located peroxinectin cluster in *A. digitifera* as well as an outgroup used to root the tree. Nodes show bootstrap values based on 1000 ultrafast bootstrap replicates in IQ-Tree.



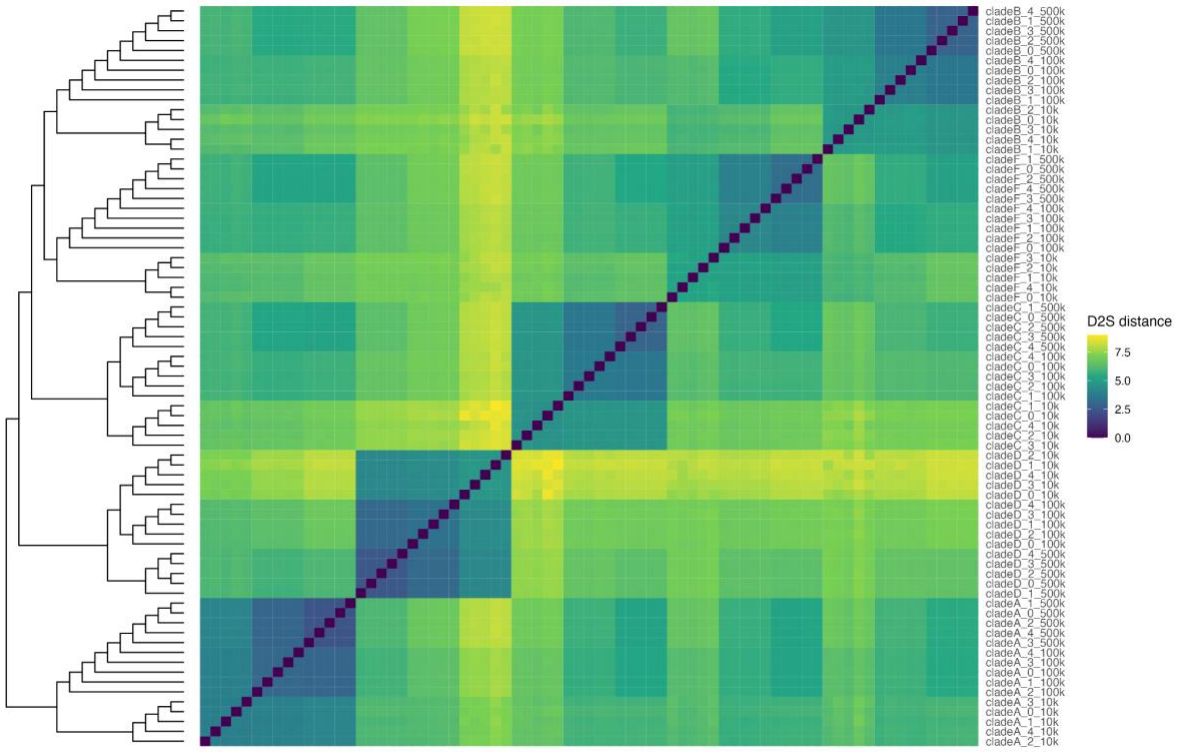
Supplementary Fig 3.22 Relationship between frequency and age for derived alleles at 178 SNPs within the gene s0150.g24. All SNPs for which GEVA was able to calculate an age are included except those that were monomorphic compared with the reference. Distributions of allele frequencies are shown using violin plots and split by age class (0-15Kya : left) and (>15Kya : right). For each allele its frequency in background haplotypes (red) is calculated separately from in selected haplotypes (blue).



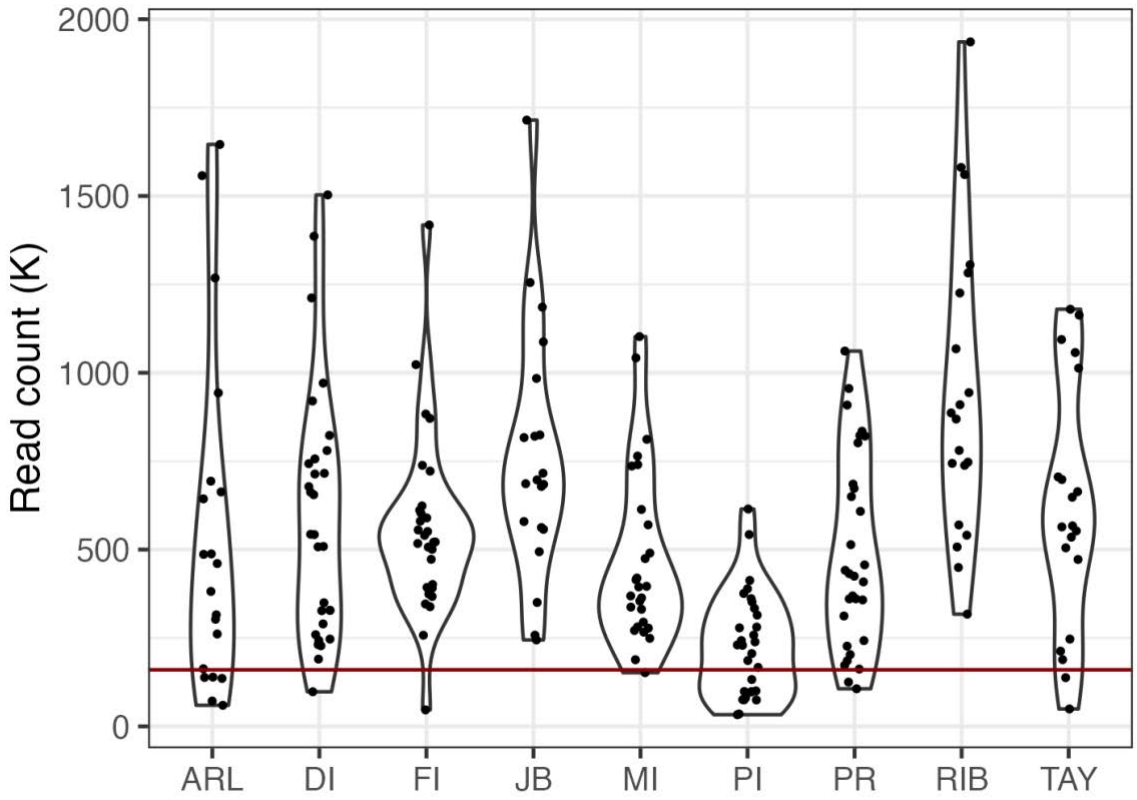
Supplementary Fig 4.1 The proportion of distinct *k*-mers and the proportion of unique *k*-mers for the overall dataset at odd-numbered k ranging from 13 to 33. The grey dashed vertical lines suggest the chosen optimal k size of 25.



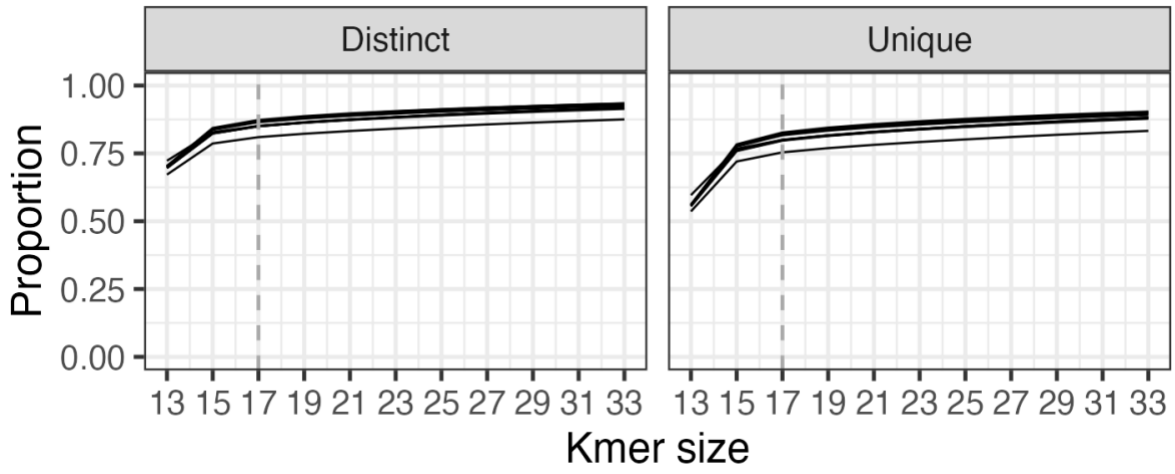
Supplementary Fig 4.2 The MDS clustering of all simulated datasets with different sequencing coverage. Different shapes were used to indicate the sequence coverages of 10K, 100K, and 500K short reads.



Supplementary Fig 4.3 The heatmap of pairwise *S* distance among samples of simulated data. The dendrogram clustering of samples on the left is based on the neighbour-joining (NJ) tree rooted with *Symbiodinium*. The samples from *Symbiodinium*, *Breviolum*, *Cladocopium*, *Durusdinium*, and *Fugacium* were labelled as Clade A, B, C, D, and F, respectively for short.



Supplementary Fig 4.4 Number of reads in thousand for each sample originated from *Cladocopium* in *Acropora tenuis* corals at distinct locations (Chapter 2).



Supplementary Fig 4.5 The proportion of distinct *k*-mers and the proportion of unique *k*-mers for 10 randomly picked samples from all samples at odd-numbered k ranging from 13 to 33. The grey dashed vertical lines suggest the chosen optimal k size of 17.

Supplementary Note

Additional methods for Chapter 3

Sample collection and sequencing

75 samples from adult corals across our three study locations were selected from a larger pool of 564 samples collected as part of a separate study across a wider geographic area that was primarily based on DArT sequencing (Adam et al. 2022). At these three study locations, small nubbins of *A. digitifera*, approximately 1-6 cm3 were collected in November 2017 (Rowley Shoals, Ashmore Reef, Adele Island and Beagle Reef) and March 2018 (Rowley Shoals) across 21 sites and stored in 100% ethanol. These samples were later subsampled to be sent to Diversity Array Technology Pty Ltd. (DArT P/L) for further processing. DNA extractions were performed for all samples by DArT and the remaining DNA (not used for DArT) was sent to the QB3 UC Berkeley sequencing centre for whole genome sequencing. Samples for whole genome sequencing were selected randomly from samples previously sequenced by DArT and after excluding 7 that failed initial quality checks an additional 7 replacement samples were also randomly selected. Initial sequencing was performed on a single NovaSeq S4 flowcell to obtain ~3 billion 2x150bp paired-end reads across all samples. Additional sequencing was then performed on a second NovaSeq S4 flowcell for 33 samples because they failed to achieve the target depth of 10x in the first batch. Samples included in the second batch of sequencing were spread across all sites in the study (Supplementary Table 3.1) and we did not observe any population structure attributable to batch in fineSTRUCTURE analyses (Supplementary Fig 3.3).

Identification of mislabelled sample

Initial population structure analyses revealed a single sample (BR_5_121) coded as inshore that clustered with south offshore samples. To check whether this was a genuine example of migration or a mislabelled sample we combined our whole genome sequencing data with all raw reads from the DArT dataset. Raw DArT reads were first mapped to the genome using bwa mem (v0.7.17) and variants called with freebayes (v1.3.2-dirty; (Garrison and Marth 2012)) with min-mapping-quality set to 30 and min-base-quality set to 20. The resulting vcf file was then filtered to retain only variants with maf>0.1, min depth of 8x and min mean depth of 15. This vcf was then merged with the filtered vcf file from whole genome analyses retaining only variant sites common to both approaches. Using this combined vcf we then calculated the relatedness using the relatedness2 statistic implemented in VCFtools (v0.1.16) between all pairs of samples and found that all but two pairs had relatedness values < 0.1. The remaining two pairs had relatedness values (>0.48) indicative of clones or identical

samples. One of these pairs was irrelevant to the current analyses as it concerned two DArT samples only. The remaining pair indicated a match between sample RS3_S_252 from the DArT dataset and BR_5_121_S125 from the WGS dataset indicating that this sample was mislabelled at some point after DArT sequencing, and its true origin was the Rowley Shoals.

Variant calling and filtering

Our variant calling pipeline was implemented using snakemake version 5.5.4 (Köster and Rahmann 2012) and is available online at (https://github.com/bakeronit/snakemake-gatk4-non-model).

The initial variant call set was filtered with the objective of minimising bias while maintaining quality biallelic SNPs suitable for the population genomic analysis. Filtering steps were performed sequentially as follows;

1. Sites within 5bp of InDels were removed using BCFtools version (1.10.2) (Danecek et al. 2021)
2. Hard-filtering thresholds were applied using the GATK VariantFiltration tool based on recommended parameters as follows (QD < 10, QUAL < 30, SOR > 3, FS > 60, MQ < 40, MQRankSum < -12.5, ReadPosRankSum < -8). Abbreviated parameters are QD=QualByDepth, QUAL=Quality, SOR=StrandOddsRatio, FS=FisherStrand, MQ=RMSMappingQuality.
3. Sites located in simple repeat regions identified by mdust version 2006.10.17 were removed (Li 2014).
4. Sites were removed if they had more than 10% missing or low quality genotype calls under the thresholds GQ > 20 and DP >3. (GQ=Genotype Quality; DP = sample read depth). This was performed using VCFtools v0.1.16 (Danecek et al. 2011)
5. Sites were removed if their read coverage fell outside expected bounds (mean per-sample depth less than 8 or greater than 31) because this could indicate collapsed repeats or regions with low mappability.

After all filtering steps, we obtained 9,656,554 high-quality biallelic SNPs from 75 samples. A summary of the number of missing genotypes in all samples after filtering is provided in Supplementary Fig 3.1B.

Haplotype phasing

To resolve haplotype information, we used the software SHAPEIT v2 (Delaneau et al. 2011) which can phase segregating sites in a sample of unrelated individuals. To improve phasing accuracy we also incorporated information from phase informative reads in mapping files

(bam format). Phase informative read information was first extracted from bam files using the tool, extractPIRs. Next we used the SHAPEIT assemble command to run the standard population-based phasing together with the read aware phasing module. This was performed separately for each scaffold and the results were combined into a single file in VCF format.

Missing genotypes were imputed by SHAPEIT2 during the assembly run. To evaluate the accuracy of imputation, we performed a "masked analysis" (Verma et al. 2014), in which a subset of genotyped SNPs in the samples was randomly pruned and then imputed as missing data. We compared the imputed genotypes to their original genotypes to estimate the concordance which indicates the performance of imputation with respect to that set of SNPs. A summary of this imputation accuracy check is provided in Supplementary Fig 3.2.

Demographic history with fastsimcoal2

To model demographic history while accounting for population structure, we carried out SFS based demographic modelling using fastsimcoal2 (Excoffier et al. 2021). We used all samples except BR_5_121 as per our SMC++ analysis. To minimise the bias from linkage disequilibrium and selection, we used BCFtools to remove sites located in genic regions and performed LD pruning in 1000bp windows with a cut-off of r2>0.3. To utilise the mutation rate in branch length computation, we estimated the monomorphic sites based on the proportional number of mappability sites defined by the SNPable pipeline we used in MSMC analysis. We also filtered out sites with missing genotypes and then used easySFS (https://github.com/isaacovercast/easySFS) to generate a joint three-dimensional folded SFS with 257,314 SNPs.

All the demographic models tested with fastsimcoal2 assume that gene flow (if modelled) is constant across the genome. To check that this assumption was appropriate we plotted the pairwise SFS for each population pair (Supplementary Fig 3.18) and checked that there was not a large excess of strongly segregating alleles (high frequency in one population, low in another). This type of SFS pattern (see for example (Tine et al. 2014)) suggests ancient divergence followed by secondary contact with genomic islands resulting from barriers to gene flow. None of the pairwise SFS's for our data exhibited this pattern. We also used the PopGenome (Pfeifer et al. 2014) package in R to calculate relative pairwise divergence (F_{ST}) and absolute divergence D_{XY} in 50kb windows across the longest 20 genomic scaffolds. If divergence is occurring under high gene flow via barrier loci (eg resulting from local selection) then regions with very high F_{ST} should also be associated with high D_{XY}. The absence of a

strong association (Supplementary Fig 3.20) is consistent with our broad finding that divergence occurred under low gene flow.

We firstly tried to test which population tree topology the SFS data support without considering the population size changes and migrations. In this step we tested four alternative topologies indicating alternative splitting modes among three populations including inshore split first, south offshore split first, north offshore split first, or a polytomy tree of three populations (Supplementary Table 3.9). For each model, fastsimcoal2 (version 2705) was used to fit parameters to the joint SFS with 50 ECM optimization cycles and 200,000 coalescent simulations used to compute the likelihood. This model fitting process was repeated 100 times based on different randomly sampled starting parameter values. This gave clear support for the inshore split first model as it always had the lowest Akaike information criterion (AIC) value across all 100 runs. We report the best AIC and likelihood values for all four models (across the 100 runs) in Supplementary Table 3.9.

Based on the population tree ((NO, SO), IN), we then tested six competing models all with exponential population size change (Supplementary Fig 3.13). These models were primarily designed to test different migration scenarios and comprised; 1) strict isolation (SI), 2) continuous migration between all demes at all times (IM), 3) continuous migration among three populations only after offshore divergence, ie secondary contact for offshore-inshore but isolation with migration for offshore-offshore (IMc), 4) isolation with recent secondary contact (SC), 5) early migration after offshore divergence (EM), 6) ancient migration between inshore and offshore ancestor followed by strict isolation (AM). We specified the search ranges for the current and ancestral effective population sizes between 1,000 and 1,000,000, and the effective population size for the offshore ancestor to between 100 and 10,000, but with an open upper bound that is extended if parameters get close to the boundary during the ECM optimisation. Divergence times were allowed to vary between 100 and 10,000 generations. The range of migration rates was assumed to be between 10-7 to 10-3 with open upper bounds. For the SC and EM models, we allowed the time of changed migration (TMIG) to be between 100 generations and the offshore divergence time (using paramInRange).

Parameters for all six models were initially estimated using the same process as outlined above. After parameter estimation, we observed that the SC and EM models were converging towards the IMc model as TMIG kept being pushed to the lower bound (100 generations) in the EM model while being optimised to be close to offshore divergence time in the SC model (Supplementary Fig 3.14). We thus deprecated these two models in the

following likelihood estimates and model comparison. Next, we compared different models using the model normalised relative likelihood (Excoffier et al. 2013) (Supplementary Fig 3.15; Supplementary Table 3.10), and estimated the parameter ranges (Supplementary Table 3.10). As a result of this process we chose the IMc model as it had a model normalised relative likelihood of close to 1 whereas this was 0 for all other models. We then estimated confidence intervals for the parameters of the best model using 100 non-parametric bootstrapping datasets, each of which was generated by sampling 257,314 SNPs with replacement from the original set of SNPs. This sampling was performed using the sample tool (alexpreynolds.github.io/sample). For each bootstrapping data set, we performed 20 independent runs. Final results shown in Supplementary Table 3.10 show 95% confidence intervals based on the distribution of fitted parameters from these independent runs.

To check the fit of the IMc model to the 2D SFS we plotted the residuals (observed SFS - modelled SFS; Supplementary Fig 3.19). Although this showed a good fit across the majority of the 2D SFS we found that the model alternately under and over estimated allele frequencies in alternating bands along the lower left corner. Since the largest of these bands is also visible in the observed SFS (Supplementary Fig 3.18A) we used two alternative methods of conversion from vcf to SFS to ensure that it was not due to an artifact of our SFS generating pipeline. Instead of using easySFS to generate the SFS we used the scripts, vcf2sfs, foldSFS and SFSTools.R available on the fastsimcoal website. As input to vcf2sfs we used phased genotypes pruned for LD (same settings as for our easySFS pipeline) with the ancestral allele encoded as the reference (see section estimating allele age with GEVA). Without any additional filtering this resulted in the SFS labelled (vcf2sfs - A) in Supplementary Fig 3.18. An alternative SFS (labelled, vcf2sfs - B) was also generated by first removing all sites within 500kb of the top 1% of windows identified as selective sweeps by either iHS, XP-EHH and XP-nSL. Since the same band is visible in all versions of the SFS we concluded that it is not an artefact of our SFS generating pipeline.

Empirical false discovery rate for signatures of selection based on population branch statistics

We used simulated data under our best-fitting demographic model with fastsimcoal2 to calculate the distribution of population branch statistics (PBS) for each population arising under neutrality. Simulations were performed 50 times using randomly selected values across the bootstrap-estimated 90% confidence intervals for model parameters. Since this generated a much larger number of PBS values to our real dataset, and includes many sites in LD, we randomly selected 100k values from this simulated data and from our real data.

The resulting 200k were then ranked by PBS value (0 the highest) and the false positive rate for the ith ranked value was calculated by counting the number of false (ie simulated) values from ranks 0 through to i and dividing this value by 0.5i. We then calculated the threshold value, above which this empirically calculated error dropped below 0.01 (1%) and used this as our criteria for significance. This was done separately for each population.

Mapping to pseudo-chromosomes

We used ragtag v.1.1.1 (Alonge et al. 2019) to align the *Acropora digitifera* genome to the *Acropora millepora* chromosome-level genome assembly (Fuller et al. 2020) with default settings. This placed 735 of the 955 *A. digitifera* scaffolds in pseudo-chromosomes, comprising 97% of assembled bases. We used this mapping to translate between scaffold level and pseudo-chromosome coordinates for the purpose of visualization only. Specifically, it was used to create the Manhattan plot (Figure 3.3A).

Gene annotations

Gene models for the *Acropora digitifera* version 2 assembly were obtained from the authors of its original publication(Shinzato et al. 2020) in gene feature annotation (GFF3) format. As these gene models are based on scaffolds from the original assembly (available at https://marinegenomics.oist.jp/adig/viewer/info?project_id=87) that have not undergone the RefSeq curation process their coordinates needed to be updated to match the ncbi assembly (GCA_014634065.1) that we used for our analyses. To do this we first aligned the two genomes with Cactus (Armstrong et al. 2020) and then used the ucsc chain and liftOver utilities(Kuhn et al. 2013) to generate updated gene model coordinates. The resulting updated gene models and full details of the lift-over process are available via the online code repository https://github.com/bakeronit/acropora_digitifera_wgs for this paper.

Starting from these updated gene models we first selected the longest transcript per gene using cgat toolkit(Sims et al. 2014) and then extracted nucleotide and protein sequences for each coding sequence using gffread(Pertea and Pertea 2020). Functional annotations for these genes were then obtained by performing blastp and blastx searches on protein and nucleotide sequences respectively against the Swissprot database (downloaded 2021 May 9)(Bairoch and Apweiler 2000), filtering to include hits at e-value < $1e^{-5}$ only. We then selected the best available blast[xp] hit for each gene and assigned this as its closest putative homolog. In addition, we used the Uniprot ID mapping service to look up detailed functional information (including GO terms) for these homologs.

Our initial gene ontology enrichment analysis was performed based on these GO terms assigned based on blast hits to Swissprot, however, we found that this often resulted in enrichment of highly specific gene ontology terms that were clearly spurious as they involved functions that are not present in Cnidarians. To resolve this issue we decided to use GO terms assigned using Interproscan version 5.53-87(Jones et al. 2014), which uses functional information assigned to conserved domains rather than to specific genes.

GO enrichment analysis

Formal statistical analysis for enrichment of GO terms is challenging because the terms themselves are not independent, and because genes are not randomly distributed across the genome. The R package topGO v2.42 (Alexa et al. 2006) attempts to deal with the first of these issues (non-independence of GO terms) by weighting the assignment of genes to terms in a way that increases the significance of more specific terms at the expense of more biologically general parent terms. We, therefore, used topGO with the default “weight01” algorithm for all enrichment tests. To deal with the second issue (non random distribution of genes across the genome) we calculated enrichment statistics at two levels. First we evaluated enrichment at the gene level. In this analysis all genes overlapping with putative selective sweeps were assigned to the target set and the complete set of all annotated genes was assigned as the background set. Since this analysis ignores the fact that multiple genes from the same GO term might be present in the same sweep region we also performed an enrichment test based on sweeps rather than genes. As this test was used as a complement to the first we performed it only for GO terms that were significant at the gene level. To perform this second test we first assigned GO terms to all 50kb regions in the genome based on the GO terms assigned to overlapping genes. This analysis included both sweep regions and non sweep regions. A p-value based on Fisher's exact test was then calculated by counting the number of sweep regions (a subset of all 50kb regions) with a given term and comparing this to the background count across all regions.

Estimating the timing of selection at the peroxinectin locus

To investigate the timing of the selective sweep on the peroxinectin locus we used the R package starTMRCA (commit cf9f021 from github)(Smith et al. 2018) which estimates the time to the common ancestor (TMRCA) of haplotypes bearing a beneficial allele based on the length distribution of ancestral haplotypes and the accumulation of mutations since divergence. Since we did not know the beneficial allele, we instead identified alleles likely to be in complete linkage with the beneficial allele to serve as its proxy. We did this by choosing sites for which the derived allele was nearly fixed (on all but 3 haplotypes) in the inshore population and completely absent offshore. There were 84 such SNPs within the sweep locus, of which 75 were found within the gene s0150.g24 that overlapped with the

strongest statistical indicators of selection (Figure 3.4A). Of these 75 sites we chose 3 spanning the length of the gene (at positions 278594, 281245, 282923)

We then used VCFtools to export a 1Mb region centred on s0150.g24 from our phased vcf. For each of the 3 SNPs chosen as proxies for the beneficial allele we then used the R package REHH(Gautier and Vitalis 2012) to generate a furcation plot, and phytools(Revell 2012) combined with ggtree (Yu et al. 2017) to plot a midpoint rooted neighbour joining phylogenetic tree of the core haplotypes (central 200 sites). These visualisations all produced qualitatively similar results, all showing a clear distinction between selected and background haplotypes in the tree and strong extended haplotype homozygosity in the furcation plot.

We then ran starTMRCA separately for each of the 3 chosen SNPs using the 1Mb phased vcf as input. Other parameters were as follows; mutation rate of $1.2e^{-8}$ per base per generation, a recombination rate of $3.2e^{-8}$ per base per generation, chain length of 10000, proposal standard deviation of 20, initial value of TMRCA drawn from a uniform distribution from 0-10000 generations. Convergence was checked by running 10 independent chains and calculating the Gelman diagnostic using the coda package in R. For each SNP we recorded the median value of the posterior estimates of the TMRCA after discarding the first half as burn-in. Our final estimate for the time of selection on the locus is reported as the range of estimated values across these three SNPs.

The mutation rate used for starTMRCA analyses is the same as used for fastsimcoal2 and SMC++. The recombination rate was estimated based on a linkage map for *Acropora millepora* (Wang et al. 2009; Dixon et al. 2015) which had a length of 1358 centimorgans. The rate used was then calculated by assuming a constant recombination rate and genome size of 430Mb for A. millepora.

Estimating allele age with GEVA

To estimate the time of origin for derived alleles in the peroxinectin locus we used Genealogical Estimation of Variant Age (GEVA)(Albers and McVean 2020). As GEVA requires polarisation of ancestral and derived alleles we performed this task first, using est-sfs(Keightley and Jackson 2018). Inputs to est-sfs were generated by performing a whole genome alignment of the *A. digitifera* genome to the genomes of two related species, *Acropora millepora* (GCF_013753865.1), and *Acropora tenuis* (http://aten.reefgenomics.org/) using progressive cactus v2.0.5(Armstrong et al. 2020). We then updated our phased vcf to encode the ancestral allele as the reference allele and used this vcf as input to GEVA.

GEVA was run assuming an effective population size of 3e4, and used the same mutation rate used throughout ($1.2e^{-8}$ per base per generation), and the same recombination rate ($3.2e^{-8}$ per base per generation) as used for starTMRCA. As GEVA uses a single value of the effective population size (Ne) as a scaling parameter its value cannot properly reflect recent expansions and bottlenecks in demographic history (Figure 3.2 main text). We choose to use the average value across populations estimated by SMC++ between the period 5kya and 200kya as this time period captures the recent bottleneck as well as long-term stable value seen in Figure 3.2. We chose not to include very recent estimates of Ne from SMC++ because these could lead to an inflated value that was not representative of the majority of the time period captured by the phylogenetic tree of haplotypes.

Additional references

Adam AAS, Thomas L, Underwood J, Gilmour J, Richards ZT. 2022. Population connectivity and genetic offset in the spawning coral *Acropora digitifera* in Western Australia. Mol. Ecol. [Internet]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.16498

Albers PK, McVean G. 2020. Dating genomic variants and shared ancestry in population-scale sequencing data. PLoS Biol. 18:e3000586.

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22:1600–1607.

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20:224.

Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, et al. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature 587:246–251.

Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28:45–48.

Cooke I, Ying H, Forêt S, Bongaerts P, Strugnell JM, Simakov O, Zhang J, Field MA, Rodriguez-Lanetty M, Bell SC, et al. 2020. Genomic signatures in the coral holobiont reveal host adaptations driven by Holocene climate change and reef specific symbionts. Sci Adv [Internet] 6. Available from: http://dx.doi.org/10.1126/sciadv.abc6318

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. Gigascience [Internet] 10. Available from: http://dx.doi.org/10.1093/gigascience/giab008

Delaneau O, Marchini J, Zagury J-F. 2011. A linear complexity phasing method for thousands of genomes. Nat. Methods 9:179–181.

Dixon GB, Davies SW, Aglyamova GA, Meyer E, Bay LK, Matz MV. 2015. CORAL REEFS. Genomic determinants of coral heat tolerance across latitudes. Science 348:1460–1462.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9:e1003905.
Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics 37:4882–4885.
Fuller ZL, Mocellin VJL, Morris LA, Cantin N, Shepherd J, Sarre L, Peng J, Liao Y, Pickrell J, Andolfatto P, et al. 2020. Population genetics of the coral *Acropora millepora*: Toward genomic prediction of bleaching. Science [Internet] 369. Available from: http://dx.doi.org/10.1126/science.aba4674
Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv [q-bio.GN] [Internet]. Available from: http://arxiv.org/abs/1207.3907
Gautier M, Vitalis R. 2012. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28:1176–1177.
Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.
Keightley PD, Jackson BC. 2018. Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site. Genetics 209:897–906.
Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28:2520–2522.
Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools. Brief. Bioinform. 14:144–161.
Li H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30:2843–2851.
Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare. F1000Res. [Internet] 9. Available from: http://dx.doi.org/10.12688/f1000research.23297.2
Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. Mol. Biol. Evol. 31:1929–1936.
Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223.
Shinzato C, Khalturin K, Inoue J, Zayasu Y, Kanda M, Kawamitsu M, Yoshioka Y, Yamashita H, Suzuki G, Satoh N. 2020. Eighteen Coral Genomes Reveal the Evolutionary Origin of Acropora Strategies to Accommodate Environmental Changes. Mol. Biol. Evol. 38:16–30.
Shinzato C, Mungpakdee S, Arakaki N, Satoh N. 2015. Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago. Sci. Rep. 5:18211.
Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. 15:121–132.
Smith J, Coop G, Stephens M, Novembre J. 2018. Estimating Time to the Common Ancestor for a Beneficial Allele. Mol. Biol. Evol. 35:1003–1017.
Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F, Belkhir K, Klages S, et al. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5:5770.
Verma SS, de Andrade M, Tromp G, Kuivaniemi H, Pugh E, Namjou-Khales B, Mukherjee S, Jarvik GP, Kottyan LC, Burt A, et al. 2014. Imputation and quality control steps for combining multiple genome-wide datasets. Front. Genet. 5:370.
Wang S, Zhang L, Meyer E, Matz MV. 2009. Construction of a high-resolution genetic

linkage map and comparative genome analysis for the reef-building coral *Acropora millepora*. Genome Biol. 10:R126.
Ying H, Hayward DC, Cooke I, Wang W, Moya A, Siemering KR, Sprungala S, Ball EE, Forêt S, Miller DJ. 2019. The Whole-Genome Sequence of the Coral *Acropora millepora*. Genome Biol. Evol. 11:1374–1379.
Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. Ggtree : An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8:28–36.