Nature or nurture?

Testing the correlation between stress tolerance and genotype in the coral *Acropora millepora* on the Great Barrier Reef

Thesis submitted by Young Koo Jin

July 2015

For the degree of Doctor of Philosophy in the College of Marine and Environmental Sciences

James Cook University

Acknowledgements

I am gratefully indebted to my supervisors, Bette Willis, Madeleine van Oppen and Petra Lundgren for giving me the opportunity to explore a field of intriguing marine life, corals, and accomplish the PhD. I would like to thank you for your patience, words of encouragement and wise guidance which helped me push through a difficult time in my life. Your advice on my research has been invaluable for allowing me to grow as a scientist.

I would also like to express thanks to researchers at AIMS, especially Lesa Peplow for her help with technical assistance to take my first steps in a lab.

Special thanks to Stuart Kininmonth for helping me with Bayesian statistics. He patiently guided me, who had no experience and knowledge about Bayesian statistics and software. Additionally, he provided extremely generous support to greatly improve Chapter 5 with his competent advice.

I would like to thank my friends in Townsville, Kero, Aomi, Mona, Jeroen and all the other people for making my time in Townsville special.

Finally, my sincere thanks to my daughter, Sana. Despite the hard time we have been going through, she gave me smiles and loved me just as she did before. This gave me strength to continue and finish my thesis.

Statement on the contribution of others

The research within this thesis was supported by the Australian Institute of Marine Science (AIMS), funding provided through the Australian Research Council (ARC) Centre of Excellence for Coral Reef Studies, and a James Cook University (JCU) Graduate Research Scheme grant.

A supervisory team consisting of Professor Bette Willis (JCU), Dr Madeleine van Oppen (AIMS), and Dr Petra Lundgren (AIMS/Monash University) provided intellectual and editorial support throughout the thesis.

Dr Stuart Kininmonth (the Stockholm Resilience Centre) provided statistical help with environmental correlation network analysis in Chapter 3 and statistical, intellectual and editorial support for Chapter 5.

Lesa Peplow (AIMS) and Roxana Capper (University of Texas) contributed to the development of the curves method and peaks method, respectively in Chapter 2.

Lesa Peplow (AIMS), Jason Doyle (AIMS), Andrew Muirhead (AIMS) provided DNA samples for gene by environment association (Chapter 3).

Dr Allison Paley (JCU) and Dr Emily Howells (JCU/AIMS/NY U Abu Dhabi) provided DNA samples and bleaching data (Chapter 4).

Dr Jean-Baptiste Raina (JCU/AIMS/UT Sydney) and Dr Adrian Lutz (AIMS/JCU) provided samples and phenotypic data from their heat stress experiment (Chapter 4)

Abstract

Corals are known to have both high phenotypic plasticity and significant capacity for acclimatisation, either through physiological, biochemical and behavioural responses to environmental stochasticity, or through contributions of symbiotic *Symbiodinium* to the physiological tolerance of the coral holobiont. Local adaptation is suggested by the fact that populations along environmental gradients maintain differences in environmental tolerance after acclimatisation to the same condition, as well as by the existence of correlations between the genetic diversity of coral hosts and environmental gradients in natural coral populations. However, no study has yet identified genetic markers underlying stress tolerance variation in corals.

The research presented here aims to improve knowledge of the adaptive importance of genetic diversity within coral populations by: (1) developing a cost- and time-efficient single nucleotide polymorphism (SNP) genotyping method that accurately estimates allele frequencies from pooled samples (Chapter 2); (2) investigating correlations between allele frequencies at candidate loci and temperature and water quality gradients along the Great Barrier Reef (Chapter 3); (3) validating genotype-environment correlations at candidate loci for phenotypes associated with bleaching tolerance in both field and laboratory settings (Chapter 4); and, (4) predicting the spatial distribution of allele frequencies at the anti-oxidant capacity loci identified in Chapter 4 through Bayesian belief network modelling (Chapter 5).

SNPs are the most common type of genetic variation and are present at both functional and neutral loci, making them ideal markers to study adaptive molecular variation. SNP allele frequency distributions can be used to identify coral populations that show signatures of adaptation to local environmental conditions, because the frequency of genetic variants (i.e., alleles) will vary with selection for or against these

mutations and any other linked loci that affect fitness of the organism. Examination of genomic signatures of environmental change at the population level provides information about the resilience of coral populations at broad geographical scales. Genotyping of pooled DNA for estimating allelic proportions provides a time- and cost-effective method for large sample sets. In Chapter 2, I introduce two quantitative High Resolution Melting (qHRM) methods for measuring allele frequencies at known SNP loci in pooled DNA samples: the “peaks” method, which can be applied to large numbers of SNPs, and the “curves” method, which is more labour intensive but also more accurate.

In Chapter 3, the distributions of SNP allele frequencies at five functional loci were investigated in 25 populations of *Acropora millepora* along the Great Barrier Reef using the curves genotyping method. Significant correlations between allele frequencies and gradients in both water quality and temperature were detected at two functional loci.

In Chapter 4, I validated the two genetic markers as true quantitative trait loci for anti-oxidant capacity in the common coral, *A. millepora*. Field surveys revealed that colonies carrying G and T alleles, respectively, at the two loci were significantly less affected by two bleaching events. These results were corroborated in an aquarium-based experiment showing constitutively higher anti-oxidant capacity of colonies with these alleles, particularly in the homozygous state. These genetic markers are located on different chromosomes and closely linked to a range of genes involved in bleaching-associated processes, such as thermal, hypoxic and oxidative stress, suggesting they are indicative of tolerance to a multitude of environmental perturbations.

Markers such as the ones identified in this study provide an opportunity to use environmental mapping to identify populations that are resistant to a range of environmental stressors. In Chapter 5, a spatial Bayesian belief network (BBN) model

was developed to predict GBR-wide spatial distributions of allele frequencies at the anti-oxidant capacity locus C29226S281. A combination of long-term environmental monitoring data, allele frequency data, expert input and statistical evaluation was used to build the model, with the goal of refining prior beliefs and examining dependencies among environmental variables that are proxy indicators of selective forces on allele frequencies at locus C29226S281. The model predicts that a high allele frequency is correlated with poor water quality and large fluctuations in sea temperatures. The Bayesian simulation approach demonstrates that the synergism between highly fluctuating sea temperatures and high nitrate concentrations may be the primary driver in selecting alleles at this locus. High temperature fluctuation in isolation does not increase the probability of finding high allele frequencies. Consistent with this, spatial mapping of predicted allele frequencies reveals that increased anti-oxidant capacity alleles are most likely to be concentrated near the mouths of the Burdekin and Fitzroy Rivers, where nutrient levels are chronically high. Coupling GIS and BBN enables visualisation of the model output, specifically qualitative measurements of allele frequencies, enabling non-experts to explore and interpret complex interactions between environmental and genetic factors.

This thesis provides compelling evidence that the two loci identified are true QTLs associated with responses to bleaching-associated stressors. Correlations between the spatial distribution of allele frequencies and genotypic variation associated with bleaching tolerance at the loci indicate the existence of adaptive genetic diversity in current coral populations. These findings open up opportunities for novel coral reef management approaches, such as spatial mapping of stress tolerance for use in Marine Protected Area design, the identification of stress tolerant colonies for assisted migration and marker-assisted selective breeding to create more tolerant genotypes for restoration of denuded reefs.

Contents

Abstract i

List of Tables viii

List of Figures ix

Chapter 1.0 Background and General Introduction 1

1.1 Responses of coral reefs to climate change on the Great Barrier Reef 1

1.2 Role of gene flow in adaptation 5

1.3 Identification of genetic markers for gene by environment associations 7

1.4 Thesis objectives 9

Chapter 2.0 Quantitative high resolution melting: two methods to determine SNP allele frequencies from pooled samples 12

2.1 Introduction 13

2.2 Materials and Methods 16

2.2.1 Coral collection 16

2.2.2 DNA extraction 17

2.2.3 Peaks method 18

2.2.3.1 Pooling 19

2.2.3.2 Internal melting calibrators 20

2.2.3.3 Asymmetric template amplification and melt stage 20

2.2.3.4 Analysis of denaturation profiles 21

2.2.3.5 DNA titrations 22

2.2.3.6 Analysis of 384 loci in six individuals 24

2.2.3.7 Analysis of 384 loci in 98 individuals collected from two locations 24

2.2.3.8 Analysis of genetic subdivision between coral populations 25

2.2.4 Curves method 25

2.2.4.1 Pooling 26

2.2.4.2 Asymmetric template amplification and melt stage 26

2.2.4.3 Analysis of denaturation profiles 26

2.2.4.4 DNA titrations and method validation 28

2.2.4.5 Application to large pools of DNA 29

2.2.4.6 qHRM reproducibility 29

2.3 Results 30
2.3.1 Peaks method 30
2.3.1.1 Allele titrations 30
2.3.1.2 Analysis of 384 loci in six individuals 32
2.3.1.3 Analysis of 384 loci in 98 individuals collected from two locations 32
2.3.1.4 Detection of novel alleles 34
2.3.2 Curves method 35
2.3.2.1 Allele titrations 36
2.3.2.2 Application to large pools of DNA 36
2.3.2.3 Reproducibility of qHRM 37
2.4 Discussion 38
2.4.1 Comparisons with other genotyping methods 38
2.4.2 Pooled genotyping by HRM and unlabelled probes 40
2.4.2.1 Peaks method 40
2.4.2.2 Curves method 41
2.5 Conclusion 43

Chapter 3.0 Allele frequency difference along water quality and temperature gradients at SNP loci in panmictic coral populations 44
3.1 Introduction 46
3.2. Methods and Materials 48
3.2.1 Study sites 48
3.2.2 Environmental variables 49
3.2.3 Target SNPs 50
3.2.4 HRM genotyping 51
3.2.5 Statistics 51
3.2.5.1 Detection of divergent populations 51
3.2.5.2 Gene by environment association 52
3.2.5.3 Environmental correlation network 52
3.2.5.4 Spatial clustering of water quality and temperature factors 53
3.2.5.5 Combined effects of NO_3 and temperature range 53
3.3 Results 53
3.3.1 Outlier detection 53

3.3.2 Gene by environment association 55
3.3.3 Environmental and biotic correlation network 56
3.3.4 Spatial clustering of water quality and temperature factors 58
3.3.5 Combined effects of NO_3 and temperature range 58
3.4 Discussion 59
3.4.1 Gene by Environment association 59
3.4.2 Detection of outlier populations 63
3.5 Conclusion 64

Chapter 4.0 Genotype-phenotype association: quantitative trait loci for coral antioxidant capacity 66
4.1 Introduction 67
4.2 Materials and Methods 70
4.2.1 Mendelian inheritance 70
4.2.2 Natural bleaching experiment 70
4.2.3 Heat stress experiment 73
4.2.4 SNP genotyping 74
4.2.5 Statistics 75
4.2.6 Functions of genes in linkage groups 75
4.3 Results 75
4.3.1 Mendelian inheritance 75
4.3.2 Natural bleaching experiment 76
4.3.3 Heat stress experiment 77
4.3.4 Linkage study 79
4.4 Discussion 80

Chapter 5.0 Predicting the spatial distribution of allele frequencies for a bleaching tolerance gene in *Acropora millepora* using Bayesian belief network modelling 86
5.1 Introduction 88
5.2 Methods and Materials 90
5.2.1 Genetic and environmental data 90
5.2.2 Environmental correlation network 92
5.2.3 Bayesian belief network 94

5.2.4 Predictability of BBN model 97
5.2.5 Sensitivity analysis 98
5.2.6 Simulation analysis 98
5.2.7 Spatial mapping 98
5.3 Results 99
5.3.1 Prediction of likelihood of allele frequencies through BBN 99
5.3.2 Sensitivity 102
5.3.3 Predictability (error rates) 103
5.4 Discussion 104

Chapter 6.0 General discussion 109
6.1 Major findings of this thesis and their implications 109
6.2 Further considerations and recommendations arising from thesis results 111
6.3 Future research directions 114

References 118
Appendices 132

List of Tables

Table 2.1. Collection year, sample size, latitudes and longitudes of 28 populations

Table 2.2. Known frequency mixtures and their qHRM-estimated frequencies of a low-melting allele for three SNP loci: C22161S248; C45133S676; C70S236. Sample batches that were pooled at gradient allelic ratios.

Table 2.3 Replicate errors associated with 25 populations and 5 SNP markers.

Table 3.1. SNP markers used in this study

Table 3.2. Spearman's rank correlation coefficients (rho) between environmental gradients and allele frequencies. *P value = 0.05-0.1 **P value < 0.05

Table 3.3. Spearman's rank correlation coefficients, testing for pairwise correlations among environmental variables. Correlation coefficients >0.6 are underlined.

Table 3.4. Summary of multiple linear regression model for testing correlations between allele frequencies and two environmental variables, SST range and NO_3.

Table 4.1. Mendelian inheritance of C29226S281 and C70S236 markers. Genotypes of female and male parents were AG and GG at C29226S281 and TC and CC at C70S236, respectively.

Table 4.2 Putative functions of linked genes. A) Genes linked to C29226S281. B) Genes linked to C70S236

Table. 5.1 Discretisation for nodes used in the BBN model. Discretised states of PN, NO_3, PP, SS, chlorophyll and secchi depth nodes are based on environmental maps obtained from the e-atlas website (see http://eatlas.org.au/data/uuid/df7012eb-a23f-4cf2-b92d-6fff77718987).

Table 5.2 Likelihoods of finding high allele frequency when associated with good (top) or poor (bottom) water quality and low (top) or high (bottom) temperature range states.

Supplementary S.1. R scripts for peak height finder

Supplementary Table S.2. 384 SNP loci from the coral linkage map selected from Wang et al. 2009.

Supplementary table S.3. Data sources for environmental variables used in this study.

Supplementary Table S.4. The numbers of samples (fragments) and colonies that were used to measure three phenotypes (Fv/Fm; chlorophyll; CoQH2).

List of Figures

Fig. 2.1 Flow chart of qHRM, including benchmarking experiments performed for each method.

Fig. 2.2 Schematic diagram of qHRM.

Fig. 2.3 Standard deviations of differences between estimated and expected allele frequencies across melting temperature range.

Fig. 2.4 DNA titrations using the peaks method.

Fig. 2.5 Analysis of 384 loci in six individuals (Peaks method).

Fig. 2.6 Analysis of 384 loci in 98 individuals collected from Orpheus Island and Magnetic Island.

Fig. 2.7 qHRM detection of novel alleles and possible gene duplication event.

Fig. 2.8 DNA titrations using the curves method.

Fig. 2.9 Outlier sample identified from Wilkie population using the SNP marker C70S236.

Fig. 2.10 Melt curves of three reference samples.

Fig. 3.1 Map showing temperature and nitrate gradients and locations of 25 populations of *Acropora millepora* sampled along the length of the Great Barrier Reef.

Fig. 3.2 Map showing the locations of genetically connected and divergent populations detected in this study.

Fig. 3.3 Outlier populations detected by 1,000 iterations of Mahalanobis distance calculations.

Fig. 3.4 Environmental correlation network for variables used in genotype by environment analyses.

Fig. 3.5 Principal component analysis reveals two clusters of reefs based on NO_3, chlorophyll and SST range data for 18 reefs.

Fig. 3.6 Relationships between allele frequencies at the two loci (C29226S281 and C70S236) and both NO_3 and SST range in 18 populations of *Acropora millepora* along the Great Barrier Reef.

Fig. 4.1 Environmental conditions at sampling sites in the Palm Island Group preceding a thermal stress event in 2006 and a low salinity event in 2009.

Fig. 4.2 Map showing sampling sites in the Palm Island Group for natural bleaching experiment.

Fig. 4.3 Genotype-phenotype associations.

Fig. 4.4 The relationship between %$CoQH_2$ in tissues of *Acropora millepora* following a heat stress experiment and genotype for SNP loci C29226S281 (top panel) and C70S236 (bottom panel).

Fig. 4.5 Genotype-phenotype association.

Fig. 4.6 Linakge maps of C29226S281 and C70S236 markers showing the genetic distance between the markers and linked genes, and corresponding KEGG-based functional groups.

Fig. 5.1 Map showing the 18 sampling sites for genetic data (red circles) on the Great Barrier Reef, environmental data (green polygon), the region excluded from analyses (red polygon) and the Burdekin and Fitzroy River catchments (blue and purple lines, respectively).

Fig. 5.2 Environmental correlation network showing high and medium correlation coefficients (dark blue=>0.7; light blue=0.6-0.7; dashed line=0.449-0.6, respectively).

Fig. 5.3 Model structure of a Bayesian belief network predicting how water quality- and temperature-related variables affect allele frequencies of the bleaching tolerance locus C29226S281 in *Acropora millepora*.

Fig. 5.4 Marginal probabilities of the environmental variables given the likelihood of finding high or low allele frequencies.

Fig. 5.5 BBN simulation analysis on synergistic effect of SST range and NO_3 on probabilities of finding high allele frequencies.

Fig. 5.6 The BBN-based prediction of bleaching tolerance allele frequency.

Fig. 5.7 Sensitivity of allele frequency at C29226S281 locus to environmental variables with no evidence (left) and high range SST entered (right).

Fig. 5.8 Error rates (predictability) of the BBN models.

Fig. 6.1 Scatterplot showing hypothetical patterns of allele frequencies at a bleaching tolerance locus in relation to a gradient of future environmental stress.

Supplementary Fig. S.1. Strip maps (left to right) showing spatial variation in: chlorophyll, PN, PP, SS and secchi depth on the Great Barrier Reef.

Chapter 1.0 Background and General Introduction

Extensive coral bleaching observed on the Great Barrier Reef and elsewhere in the world over the last few decades (e.g. Hoegh-Guldberg 1999; Berkelmans et al. 2004) has prompted researchers to examine adaptive mechanisms of corals in response to rapid climate change. The capacity of corals to respond to bleaching-associated stressors is known to vary among and within populations as a consequence of variation in stress tolerance conferred by different clades of their algal symbiont *Symbiodinium* and because of acclimatisation to historical environmental regimes or stressors previously experienced. Although a genetic map of the coral *Acropora millepora* has revealed fundamental information about the molecular structure of the genome and provides crucial information for evolutionary studies (Wang et al. 2009), our understanding of adaptive genetic traits in corals is still in its infancy (Baums 2008). In this chapter, I will provide an overview of the impacts of climate change on ocean ecosystems and corals, and the importance of population genetic connectivity for adaptation to rapid environmental change at meta-population level.

1.1 Responses of coral populations to climate change on the Great Barrier Reef

The Great Barrier Reef (GBR) is the largest existing aggregation of coral reefs in the world, stretching over a distance of 2900 km along the east coast of Australia. Due to its vast geographical size, GBR reefs are exposed to a wide range of environmental parameters that fluctuate in time and space (De'ath 2007). Sea surface temperature varies with latitude, and annual averages range from 24.8°C in the southern GBR to 28 °C in the far northern GBR (http://eatlas.org.au/data/uuid/df7012eb-a23f-4cf2-b92d-6fff77718987). Water quality differs greatly between offshore and inshore regions,

particularly near the mouths of large river catchments because of the high loads of nutrients and sediments transported in river discharges (Devlin et al. 2012). In addition to seasonal fluctuations, the GBR also experiences large inter-annual variability in temperature, rainfall and cyclone occurrence that is linked to El Niño-Southern Oscillation (ENSO) events (Lough 1994).

Variation in environmental conditions is a major driver of distribution patterns of species and biological community types along the GBR (Fabricius et al. 2012). An ecological system with redundant species and diverse functional communities enhances its capacity to return to ecological equilibrium after environmental perturbations (Peterson et al. 1998; Folke et al. 2004). Furthermore, the structural complexity provided by coral reefs strongly influences associated invertebrate and fish species richness, abundance and biomass (Friedlander & Parrish 1998; Gratwicke & Speight 2005). However, the resilience of coral reefs and other marine ecosystems in many regions worldwide has been compromised by a combination of direct and indirect pressures, including climate change (Hughes et al. 2003; Hoegh-Guldberg et al. 2007; Mumby et al. 2007; Graham et al. 2008).

Coral reefs are currently experiencing 2 to 3 times greater shifts in environmental conditions than they have experienced in the past 420,000 years during which marine organisms have evolved and established the current complex ecosystem (Hoegh-Guldberg et al. 2007). Temperature increases in the ocean and atmosphere affect the strength and direction of ocean circulation (Toggweiler & Russell 2008) and subsequently ENSO variability (Timmermann et al. 1999), which further influences coral reefs due to associated changes in meteorological conditions (Hoegh-Guldberg et al. 2007; Loeb et al. 2009). Increases in the heat content of the upper layer of the ocean also create greater thermal stratification in the water column, reducing vertical mixing

and influencing primary production caused by changes in nutrient availability (Polovina & Howell 2008). The warming of the oceans also impacts the metabolism of organisms, thus influencing interactions between prey and predators. In response to temperature increase, consumers have increased energy demands that reduce plant biomass due to a constraint of resource availability, leading to knock-on effects higher up the food chain (O'Connor et al. 2009).

The dramatic increase in concentrations of carbon dioxide during the 20th century has driven increases in the global average sea surface temperature by 0.85 °C and in sea level by 19 cm, and a decrease in acidity by 0.1 pH unit (Pachauri et al. 2014). This significant environmental alteration has had, and will continue to have serious impacts on marine ecosystem, ocean productivity and food web dynamics. These rapid changes in temperature and acidity are particularly concerning for corals, which live close to their upper thermal tolerance limits (Berkelmans and Willis 1999) and rely heavily on carbonate chemistry for their structure (Hoegh-Guldberg et al. 2007).

The GBR experienced intensive mass coral bleaching events in 1998 and 2002 due to multiple environmental factors, including elevated sea surface temperature and solar radiation (Berkelmans et al. 2004). Additionally, nutrient levels have a strong influence on algal cell densities (Muscatine et al. 1989) and act synergistically with thermal stress to cause bleaching when high cell densities produce increased concentrations of harmful reactive oxygen species (ROS) under photoinhibition (Wooldridge 2009; Cunning & Baker 2012; Wiedenmann et al. 2013). Bleaching is a paling or whitening of coral tissues due to the loss of endosymbiotic algae, *Symbiodinium* spp., or the loss of photosynthetic pigments (Lesser 2011).

Symbiodinium is a key component of coral fitness, as the symbiotic algae translocate a major proportion of their photosynthates to coral hosts (Trench 1979).

Many bleached corals suffer mortality if *Symbiodinium* cell densities do not recover, and those that regain their *Symbiodinium* cell densities often show reductions in tissue biomass, reproductive output and disease susceptibility (Szmant & Gassman 1990; Harvell et al. 2001; Baird & Marshall 2002; Miller et al. 2009), highlighting the strong influence bleaching has on coral fitness.

Bleaching is preceded by photoinhibition in *Symbiodinium,* which is caused by damage to photosynthetic machinery at photosystem II (PSII) (Niyogi 1999) and leads to increased cellular concentrations of ROS (Lesser 2011). Both the coral host and *Symbiodinium* symbionts produce antioxidant enzymes, such as superoxide dismutase (SOD) and coenzyme Q, to quench the ROS (Shick et al. 1995; Lutz et al. 2014). When such antioxidant defences are overwhelmed under stressful conditions, *Symbiodinium* cells are lost from coral tissues, leading to bleaching (Lesser 2006).

Variation in tolerance to bleaching-associated stressors has been observed in both field and laboratory experiments and can be attributed to symbiont shuffling (Berkelmans & van Oppen 2006; Stat et al. 2006), environmental history (Guest et al. 2012) and host genotypic diversity (Barshis et al. 2010). Polymorphism in coral phenotypes conferred by genotypic diversity provides hope that genetic markers associated with tolerant phenotypes may help to identify and subsequently protect resilient and resistant coral populations. Studies analysing gene expression of coral larvae produced from controlled crosses have reported family-specific variation that indicates heritable genetic variation for traits related to the thermal stress response within coral host populations (Meyer et al. 2009; Meyer et al. 2011; Polato et al. 2013). Recent genomic studies also demonstrate that corals from two different temperature regimes show genotypic-related differences in expression levels of bleaching-associated genes (Barshis et al. 2013; Bay & Palumbi 2014; Dixon et al. 2015). The existence of

genetic variation underpinning phenotypic diversity provides the substrate upon which selection can operate.

1.2 Role of gene flow in adaptation

Current understanding of the importance of gene flow for adaptation in rapidly changing environments is complex, but critical to the selection of suitable species for studies of gene by environment associations. Moderate to high gene flow that mixes genetic information among populations can enhance their genetic diversity, providing greater adaptive potential than genetically isolated populations in the face of rapid climate change. However, because genomic sequences are rearranged in sexual reproduction to produce new combinations of alleles (i.e. genetic recombination), rearrangements of alleles at multiple loci that independently contribute to a single fitness-related trait can potentially retard genetic divergence at a single locus and ultimately retard directional selection. A study by Johnson et al. (2011) evaluating the constraining effects of genetic recombination on diversification in *Oenothera* sp. (evening primrose), which is capable of both sexual and asexual reproduction like many coral species, demonstrates this. They found lower diversification rates in sexually-reproducing populations than in clonal populations, which they attributed to the homogenising effect of genetic recombination in sexual populations that, although producing higher genotypic diversity, slowed down divergence at loci investigated (Johnson et al. 2011). Adding to the complexity, many studies suggest that the homogenising effect of gene flow can dilute local adaptation by introducing maladapted alleles into sink populations, whereas limited gene flow facilitates divergence in response to directional selection leading to enhanced local adaptation (Langerhans et al. 2003; Hendry et al. 2004; reviewed in Räsänen & Hendry 2008). In support of this, a quantitative genetics simulation study demonstrated that both limited gene flow and asexual reproduction result in rapid fixation of alleles that underpin local adaptation (Le

Corre & Kremer 2003). In particular, in populations with low genotypic diversity, fast allelic fixation can result from chance fluctuations in allelic proportions (i.e. genetic drift).

Although low gene flow may facilitate divergence among populations, there are important population genetic consequences associated with low gene flow. For example, although higher fitness may be associated with adaptive diversification in an isolated small population (Hunter 1993), limited gene flow can lead to a large genetic load in an isolated gene pool, such as increased frequency of homozygous recessive deleterious loci (i.e. inbreeding depression) (Lynch et al. 1995). Despite the accumulation of deleterious alleles, individuals with higher fitness conferred by alleles at loci with large effects may increase in number if effects of advantageous alleles overcome combined effects of disadvantageous alleles. However, even if a single genotype is dominant in an environment, high susceptibility to environmental change is expected due to low phenotypic variation in a population with limited gene flow.

A recent study highlights the likelihood that, on balance, adaptation to rapid environmental change is facilitated in populations with high gene flow and provides experimental evidence contradicting the widely held view that limited gene flow facilitates adaptive divergence. In a microcosm experiment with yeast, Gray and Goddard (2012) revealed that, when gene flow between different niches was tightly controlled in both sexually-reproducing and clonal populations, gene flow did not constrain local adaptation but rather conferred higher fitness in sexuals, enabling simultaneous and more rapid local adaptation to different niches. This superior ability to respond to environmental change by sexually-reproducing populations is explained by high genetic diversity generated through sexual reproduction and gene flow.

In summary, for studies of gene by environment associations, it is important to avoid isolated populations, which for historical factors (i.e. drift in isolation) rather than

selection may have resulted in allele frequency differences. Therefore, identification of genetic loci that are associated with adaptation to rapid environmental change may require investigation of populations with high gene flow, like those of coral species with a broadcast spawning mode of reproduction.

1.3 Identification of genetic markers for gene by environment associations

Rapid adaptation may be possible through the exchange of existing allele variants at a locus that confers a variety of advantageous phenotypic effects. Gagnaire et al. (2012) demonstrated that genetic variants underlying local adaptation may be best captured in gene by environment (G x E) analyses that compare the frequency of alleles among populations inhabiting different environmental conditions. There are many observations of clinal patterns in both phenotypes and genotypes along latitudinal gradients or along environmental gradients (Schneider et al. 1999; Sezgin et al. 2004; Zhen & Ungerer 2007; Mullen & Hoekstra 2008; Paaby et al. 2010). Traits that show geographic patterns are often under strong selection (Schemske & Bierzychudek 2007; Mullen & Hoekstra 2008), and hence alleles that are correlated with traits are thought to be important for adaptation.

G x E association studies take advantage of differences in environmental conditions and allele frequencies among study sites. Precautions need to be taken to avoid spurious associations that may be identified in such analyses. In the field, a clinal pattern of one environmental parameter is likely to encompass changes in multiple variables. For example, Lundgren et al. 2013 demonstrated that coral populations on reefs associated with both water clarity and temperature gradients that differed in geographic direction were characterised by patterns in allele frequencies that corresponded to the two environmental gradients but at different loci. Furthermore,

multiple stressors, such as water quality and temperature stresses may act synergistically to cause bleaching in corals (Wooldridge 2009; Cunning & Baker 2012). This may affect frequencies of alleles at loci associated with bleaching tolerance.

It should be noted that without consideration of genetic connectivity, gene by environment analysis may reveal false positives. If no information on genetic connectivity is available, disentangling the effects of isolation by distance, genetic drift and recent admixture from environmental effects on genomic patterns of diversity may prove difficult (Schmidt et al. 2008). Another complication that might arise in independent or loosely connected populations is convergent evolution, where different genetic pathways or mutations are responsible for the same phenotypic endpoint (Tishkoff et al. 2007; Stanek et al. 2009; Steiner et al. 2009). Therefore, information on genetic connectivity is essential for gene by environment analyses.

Many studies have found divergence at putatively adaptive loci in association with environmental gradients despite the existence of high levels of gene flow (Smith et al. 1997; Schneider et al. 1999; Schemske & Bierzychudek 2007; Gagnaire et al. 2012). Although such environmental pressures may be strong enough to shape clinal patterns in allele frequencies, they are often not stable but change over time. A classic study of allele frequency distribution of the *Lap* locus in mussels showed a temporal change in the distribution of alleles in some populations related to changes in food supply, temperature and salinity (Hilbish 1985; Hilbish & Koehn 1985; Hilbish & Koehn 1987). In theory, if generation times are fast enough relative to the rate of environmental change, genetic connectivity enables populations in different locations to exchange adaptive alleles and quickly adapt to environmental change. Thus, in organisms with short generation times and high gene flow across a wide range of environmental conditions, such as the reef-building coral, *Acropora millepora*, allele frequency analysis may be able

to capture adaptive loci that are shaped by current (contemporaneous) or recent selection.

Genetic markers showing G x E associations can be further validated by examining genotypic effects on phenotypes. A single nucleotide polymorphism (SNP) is a point mutation of a single nucleotide and is the most abundant variation in a genome. SNPs are revealed by comparing genomic DNA among individuals and have been widely used as a genotypic marker in the field of medical genetics. SNP-trait (e.g. disease) association studies have successfully identified many robustly linked loci (Feero et al. 2010). The SNP-phenotype association approach has also been applied to ecological studies and proven successful in identifying linked loci important for ecological adaptation (Paaby et al. 2010; Renaut et al. 2011). Some studies on corals have also attempted to show genotypic associations with phenotypes, but were unable to exclude epigenetic effects (acclimatisation) (Bay & Palumbi 2014; Dixon et al. 2015) given that environmental experiences of both mothers and fathers can be transmitted over the next generations (Youngson & Whitelaw 2008).

Genetic markers associated with coral fitness variations that are independent of acclimatisation effects can be used for a wide range of applications to coral reef conservation and management, including spatial mapping of tolerant corals and identification of tolerant corals for selective breeding and assisted migration.

1.4 Thesis objectives

In this thesis, I aim to identify genetic markers that are associated with tolerance to bleaching-associated stressors, particularly water quality and temperature, in the reef-building coral *Acropora millepora*. First, I develop a single nucleotide polymorphism genotyping method that determines allele frequencies from pooled samples to improve the cost and time efficiency of analysing data from large-scale studies (Chapter 2).

Secondly, using the novel method developed in Chapter 2, I investigate correlations between allele frequencies of genetic markers and environmental gradients along the GBR, focusing on temperature and water quality (Chapter 3). Thirdly, I validate the genetic markers that showed correlations between allele frequencies and environmental gradients by examining genotype-phenotype associations in both field and laboratory experiments (Chapter 4). Finally, I develop a prediction model to explore complex effects of multiple environmental variables on spatial distributions of bleaching tolerance allele frequencies in populations of *A. millepora* throughout the GBR (Chapter 5).

- In Chapter 2, I demonstrate the high accuracy of the SNP genotyping method developed, which enables cost and time efficient genotyping of pooled samples using a quantitative high resolution melting (qHRM) technique. Two qHRM methods are presented and compared in terms of their accuracy in allele frequency measurement and advantages over previously published methods.

- In Chapter 3, allele frequencies of 25 populations along the GBR are estimated using the method developed in Chapter 2. With the exclusion of genetically isolated populations, I examined correlations between allele frequencies of 5 candidate loci and gradients in 21 environmental parameters related to either water quality or temperature.

- In Chapter 4, I validate the two genetic markers identified in Chapter 3 by showing genotypic associations with bleaching-associated phenotypes in both field and laboratory settings. Additionally, to account for functional associations between the phenotypes and genotypes, putative functions of genes linked to the loci are investigated.

- In Chapter 5, I develop a spatial Bayesian belief network model to assess the spatial distribution of allele frequencies at locus C29226S281 in GBR-wide populations of *A. millepora,* by using a combination of long-term monitoring data for nine environmental variables, allele frequency data for 18 genetically connected populations, an expert knowledge-based causal web, and statistical evaluation.

Chapter 2.0 Quantitative high resolution melting: two methods to determine SNP allele frequencies from pooled samples

This chapter has been published in the journal BMC Genetics (2015) doi:10.1186/s12863-015-0222-z: Roxana L. Capper, Young K. Jin, Petra B. Lundgren, Lesa M. Peplow, Mikhail V. Matz and Madeleine J.H. van Oppen

YKJ and RLC are co-first author and co-wrote the manuscript. YKJ designed, conducted and analyzed the curves method. RLC designed, conducted and analyzed the peaks method. LMP contributed to designing the curves method. PBL, MVM and MJHvO supervised the project and contributed to experimental design and editorial support.

Abstract

The advent of next-generation sequencing has brought about an explosion of single nucleotide polymorphism (SNP) data in non-model organisms; however, profiling these SNPs across multiple natural populations still requires substantial time and resources. Here, I introduce two cost-efficient quantitative High Resolution Melting (qHRM) methods for measuring allele frequencies at known SNP loci in pooled DNA samples: the “peaks” method, which can be applied to large numbers of SNPs, and the “curves” method, which is more labor intensive but also slightly more accurate. Using the reef-building coral *Acropora millepora*, both qHRM methods can recover the allele proportions from mixtures prepared using two or more individuals of known genotype. Furthermore, advantages of each method over previously published methods are demonstrated; specifically, the “peaks” method can be rapidly scaled to screen several hundred SNPs at once, whereas the “curves” method is better suited for smaller numbers of SNPs. Compared to genotyping individual samples, these methods can save considerable effort and genotyping costs when relatively few candidate SNPs must be profiled across a large number of populations. One of the main applications of this method could be validation of SNPs of interest identified in population genomic studies.

2.1 Introduction

Population genomics seeks to link genome-wide genetic variation to evolutionary processes. By querying a large number of individuals for multiple single nucleotide polymorphisms (SNPs) distributed across the genome, polymorphisms with unexpected patterns of genetic differentiation can be identified. These SNPs may be linked to loci under divergent selection between disparate environments (Stinchcombe & Hoekstra 2008). Recent advances in high-throughput sequencing have made it easy to identify large numbers of SNPs in a limited number of individuals (Davey et al. 2011).

However, profiling a smaller number of SNPs across many populations, for example to study allele frequencies at specific candidate SNP loci, remains time consuming and expensive if based on individual genotyping.

There are many conceptual approaches to determine allele frequencies from pooled samples, including (*i*) direct sequencing (Bansal et al. 2011; Gautier et al. 2013), (*ii*) primer extension using pyrosequencing (Lavebratt & Sengul 2006; Doostzadeh et al. 2008), SNaPShot (Norton et al. 2002) or MALDI-TOF mass spectrometry (Werner et al. 2002), and (*iii*) preferential hybridization or amplification using allele-specific primers, such as TaqMan (Holland et al. 1991), qPCR (Germer et al. 2000), the Illumina GoldenGate platform (Campino et al. 2011), or hybridizing microarrays (Butcher et al. 2004). For a more extensive review of quantitative genotyping technologies, see Sham et al. (2002) and Garvin et al. (2010). Many of these technologies show high correlations between the estimated genotype frequency from the pooled sample and the true allele frequency assessed from genotyping of individual samples, but most require considerable investment of funds and effort for each new SNP assay. While direct sequencing with next-generation sequencing methods is an increasingly popular option, the sequencing of whole genomes, reduced representation libraries (RRL), or even specific amplicons incurs unnecessary costs for research that seeks to investigate the frequencies of a subset of preselected candidate or otherwise potentially informative ("tag" or "proxy") SNPs across populations.

Conventional high resolution melting (HRM) has been established as a very sensitive tool for detecting even low levels of a target allele (Er & Chang 2012). Quantitative HRM methods have been developed in a range of applications Palais et al. 2005), including detection of the adulteration of food or drug products (Mader et al. 2011; Sakaridis et al. 2013a; Sakaridis et al. 2013b; Ganopoulos et al. 2013), methylation

status (Balic et al. 2009; Malentacchi 2009; Stuopelyte et al. 2013) and species composition in samples over time (Lin & Gänzle 2014). One commonality of these approaches is that they make use of relatively large differences among variants, such as multiple methylated sites within an amplicon, multiple SNP differences among species in the locus of interest or even entirely different genes representing different species. These methods are not sufficiently sensitive to detect frequencies of single SNPs among populations of the same species.

The conventional, unlabelled probe version of HRM method of SNP genotyping uses a small oligonucleotide probe to increase the differences in melting temperature between single SNP alleles. It makes use of a dye that fluoresces only when intercalated into duplex DNA structures. First, the SNP-bearing locus of interest is amplified with asymmetric PCR such that an excess of one strand is produced. Next, the short probe that overlays the SNP site itself is added to the reaction. The reaction is subsequently heated, cooled and heated again while the loss of fluorescent signal (i.e. duplex DNA disassociation) is monitored. The melting temperature of the probe-SNP duplex region is dependent on the SNP allele state; a perfect match will have a higher melting temperature than will a mismatch (Zhou et al. 2004). A conventional HRM assay typically seeks to determine whether the probe melting profile contains only the higher-temperature component, the lower-temperature component, or both.

The main objective of this study is to explore whether the unlabelled probe HRM melting profile can provide quantitative information about frequencies of alternative alleles in a pooled DNA sample. Two versions of qHRM were developed using different dye chemistries, brands of HRM machines and analytical techniques to quantify the relative proportions of each allele. Six experiments were performed using adult

Acropora millepora corals, an emerging ecological genomics model, to demonstrate the new method's accuracy.

2.2 Materials and Methods

2.2.1 Coral collection

Adult corals were used in all experiments. For the allele titration and SNP panel validation experiments (peaks method), corals were collected from Trunk Reef (-18.368 S, 146.818 E) and Little Pioneer Bay, Orpheus Island (-18.604 S, 146.486 E), Great Barrier Reef, Queensland, Australia in 2009. Samples used in the Orpheus/Magnetic population comparison experiment (peaks method) were collected from Orpheus Island and Nelly Bay, Magnetic Island (-19.165 S, 146.851 E). Coral samples used to validate the curves method experiments were collected from 28 coral populations, summarized in Table 2.1. All corals were collected under the appropriate Great Barrier Reef Marine Park Authority permits.

Table 2.1. Collection year, sample size, latitudes and longitudes of 28 populations.

Locations	Collection year	Number of samples	Latitude	Longitude	Methods
Wallace Islet Reef	2008	46	-11.454005	143.04433	
Night Reef	2008	56	-13.182107	143.577187	
Wilkie Reef	2008	49	-13.774318	143.639923	
Emily Reef	2005	39	-15.613364	145.662024	
Sudbury 2 Reef	2005	43	-16.99523	146.271164	
Sudbury 1 Reef	2005	42	-17.02292	146.172561	
Myrmidon Reef	2009	43	-18.266005	147.38343	
Kelso Reef	2011	48	-18.424167	146.985556	
Rib Reef	2009	47	-18.47786	146.86781	
SE Pelorus Reef	2004	49	-18.563195	146.499514	
NE Orpheus Reef	2004	49	-18.569751	146.492126	
N Fantome Reef	2004	50	-18.659781	146.512061	
Nelly Bay, Magnetic Island	2007	52	-19.162022	146.852245	Curves
Darley Reef	2005	47	-19.173619	148.116332	
Holbourne Island Reef	2003	50	-19.729914	148.354758	
Ross Reef	2005	35	-19.876433	149.562767	
Boulton Reef	2005	48	-20.474688	150.27273	
Goble Reef	2005	49	-20.766017	150.469284	
Calder Reef	2003	50	-20.771921	149.617692	
21-121 Reef	2005	41	-21.231497	151.383576	
High Peak Reef	2003	36	-21.959152	150.691722	
North Kepple Island Reef	2003	49	-23.08675	150.890433	
Half Tide Reef	2009	47	-23.1597	150.93921	
Halfway Island Reef	2008	48	-23.20046	150.97062	
Humpy Reef	2010	48	-23.214	150.972	
Trunk Reef	2009	3	-18.368449	146.817698	
Little Pioneer Bay, Orpheus Island	2009	50	-18.603944	146.485877	Peaks
Nelly Bay, Magnetic Island	NA	48	-19.165175	146.851042	

2.2.2 DNA extraction

For the coral individuals used in the allele titration and SNP panel validation experiments (peaks method), genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen) and was eluted into Buffer EB. Individuals sampled from Orpheus and Magnetic Islands (peaks method) and from the 28 populations across the Great Barrier Reef (curves method) were extracted according to the method described in Wilson et al. (2002). DNA concentrations were initially determined using the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). The workflow of both methods are illustrated in Fig. 2.1. The descriptions of both methods are given in Fig. 2.2 and the following sections.

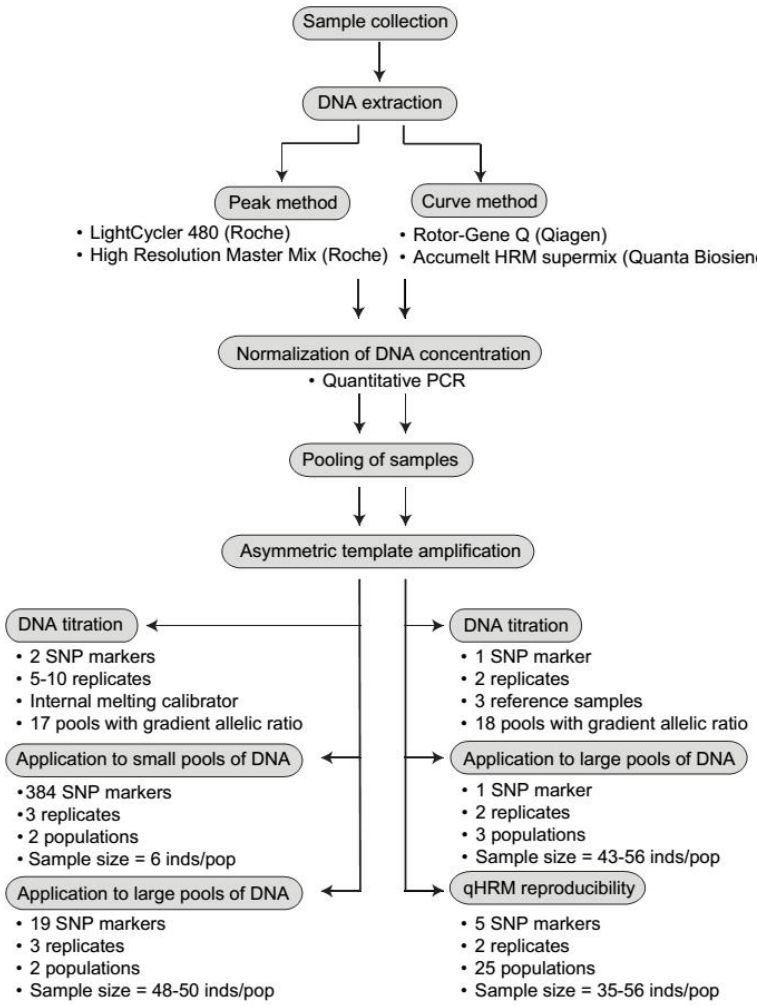

Fig. 2.1 Flow chart of qHRM, including benchmarking experiments performed for each method.

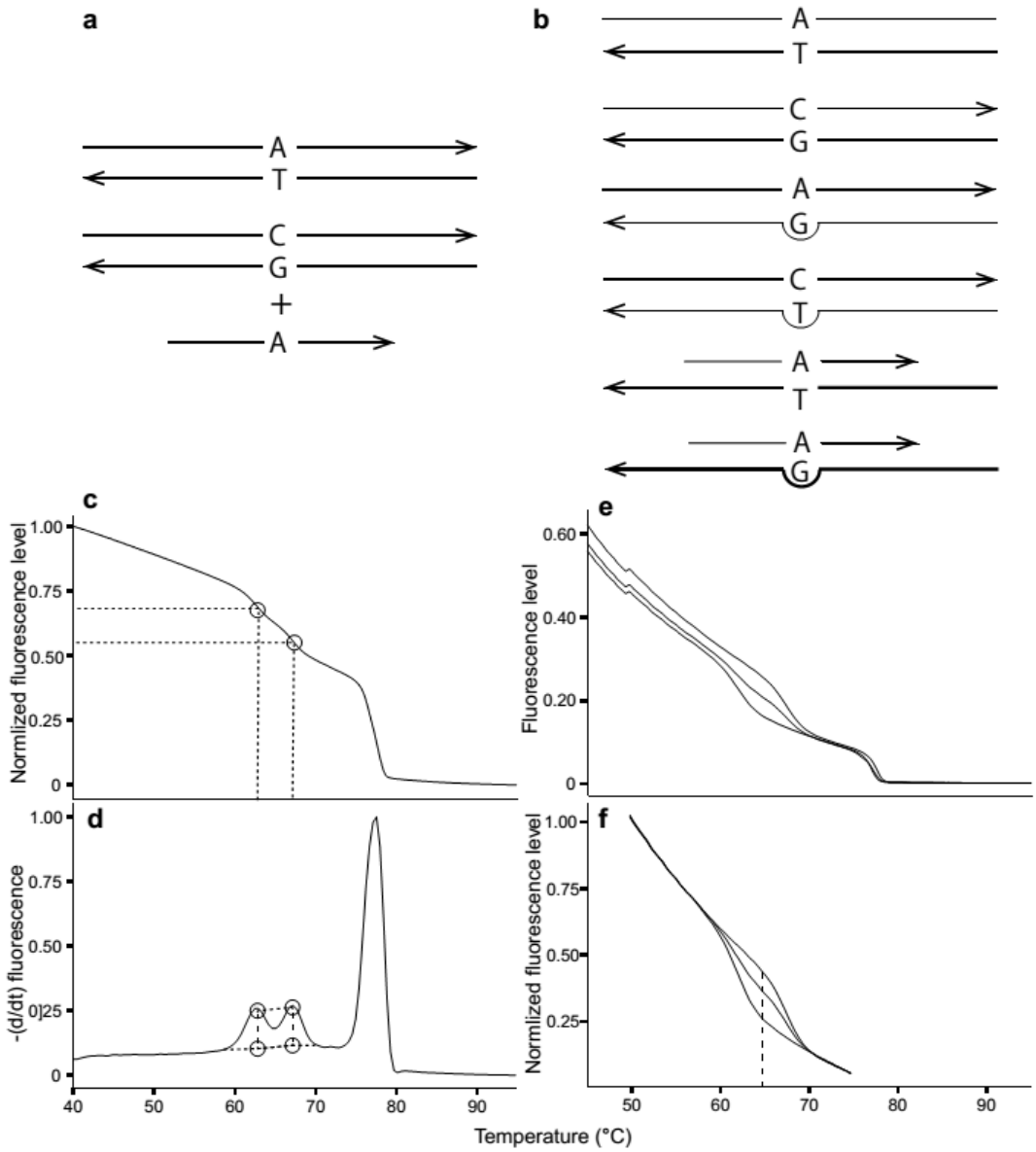

Fig. 2.2 Schematic diagram of qHRM. An individual heterozygous for a particular SNP locus is shown as an example. a) For both methods, the locus is amplified from genomic DNA using asymmetric PCR to generate an excess of the reverse strand. Next, an oligonucleotide probe is added to the reaction. b) The mixture is heated to melt all duplexes apart, and subsequently cooled rapidly to promote unbiased duplex reformation of all possible duplex DNA species. For a biallelic SNP site, this produces six distinct types of duplexes: two perfect-match amplicon duplexes, two mismatch amplicon duplexes, one perfect-match probe duplex, and one mismatch probe duplex. c) For the peaks method, the reaction is heated again to denature all duplexes while the fluorescence of the reaction, which quantifies double stranded DNA, is monitored. d) Next, the negative first derivative of the decreasing fluorescence curve is calculated to transform the curve into a peaks profile. The height of the lowest-melting, mismatched probe duplex peak is divided by the total heights of both probe peaks to yield the frequency of the mismatched allele in the sample. e) For the curves method, the same process was repeated as described for panel c. Three known samples that consist of each genotype (i.e. heterozygote, high-melting homozygote and low-melting homozygote) were used as references. f) Melt curves from the raw channel were normalized by averaging fluorescence value outside a melting phase and forcing the values to be the same for each sample.

2.2.3 Peaks method

The peaks method was developed as a rapid means to quantify SNP allele frequencies across several hundred SNP loci. Briefly, proof of concept was first established using titrated allele frequencies and two SNP markers, then a panel of 384 SNPs was validated using six corals, genotyped individually and as a pool. This panel of SNPs was finally applied to samples of two natural coral populations to detect loci with different frequencies among the reefs.

2.2.3.1 Pooling

In order to convert conventional HRM into a quantitative assay for pooled samples, equal amounts of each individual's DNA must be added to the initial asymmetric PCR reaction. This DNA normalization step is particularly critical for pooling organisms that may have known or unknown assemblages, infections or symbioses as the relative amounts of target and contaminating DNA can vary among samples. To circumvent these problems, quantitative PCR (qPCR) was used to accurately measure the quantity of coral DNA in individual samples. Primers to amplify the SNP locus C23209S177 (Wang et al. 2009) were used in a conventional qPCR reaction performed under the following conditions using a LightCycler 480 machine (Roche): 5 ng holobiont DNA, 0.1 μM each forward and reverse primers, 2 mM $MgCl_2$ and 1x High Resolution Master Mix (Roche) in a 15 μl volume. Reactions were heated to 95°C for 10 minutes, then cycled 55 times as follows: 40 s at 95°C, 40 s at 60°C, 40 s at 72°C, then cooled to 40°C and held for 20s. Concentrations of target DNA were normalized based on differences in quantification cycle (Cq) by assuming that the fold-difference per each Cq unit difference was equal to 1/*E*, where *E* is the amplification factor per PCR cycle for the particular primer set (Pfaffl 2001). The accuracy of this approach was verified by re-amplifying the adjusted DNA concentrations again with the same primer pair.

2.2.3.2 Internal melting calibrators

Though within-plate variation among replicates was low for the peaks method, between-plate variation could shift the melting profile of all duplex DNA species (i.e. for an analysed heterozygote, such species include two perfect-match amplicons, two mismatched amplicons, one perfect-match probe-SNP duplex, and one mismatch probe-SNP duplex; Fig. 2.2a,b) by as much as 2°C. To accommodate this phenomenon, which occurred only in the peaks method, standard internal melting calibrator primers were included with each reaction in order to provide landmarks in the melting profile. Four oligonucleotide calibrators were designed to form two complementary DNA duplexes with known melting temperatures. They were based on sequences from Gundry et al. (2008) but were adjusted to melt at the slightly lower and higher temperatures of 47.5°C and 90°C than the published sequences in order to melt well outside of the range of the probe and amplicon duplexes. The 3' terminal ends of each of the four oligos were modified with an inverted T to block any potential primer extension from occurring. The forward sequence of the low-melting calibrator used is 5'- ATT TTA TAT TTA TAT ATT TAT ATA TTT TT/3InvdT/ -3', while the forward sequence of the high-melting calibrator is 5'- GCG CGG CCG GCA CTG ACC CGA GAC TCT GAG CGG CTG CTG GAG GTG CGG AAG CGG AGG GGC GGG/3InvdT/. The calibrator oligos were included in the initial asymmetric PCR reaction at a concentration of 0.05 μM for each of the high-melting oligos and 0.5 μM for the low-melting oligos; addition of the calibrators did not interfere with amplification of the target products.

2.2.3.3 Asymmetric template amplification and melt stage

In HRM, the asymmetric PCR stage serves to generate an excess of the amplicon strand that is complementary to the probe. This excess strand can then form a duplex with the unlabelled probe, whereas if it was not in excess it would preferentially bind to

the longer, complementary amplicon strand instead. Though some versions of HRM use symmetric PCR, their applications typically involve whole-amplicon melting analysis. In contrast, unlabelled probe HRM, which yields the increased sensitivity necessary for SNP discrimination, requires asymmetric PCR in order to outcompete the amplicon. The asymmetric PCR reaction prior to each assay's melt stage was performed as follows on the LightCycler 480 (Roche): 5 ng total pooled DNA, 0.3 μM reverse primer, 0.067 μM forward primer, 0.05 μM each high-melting calibrator forward and reverse oligos, 0.5 μM each low-melting calibrator forward and reverse oligos, 2 mM $MgCl_2$, and 1x High Resolution Master Mix (Roche). Reactions were heated to 95°C and held for 10m, then cycled as follows: 40 s at 95°C, 40 s at 60°C, 40 s at 72°C for 55 cycles, which, for most SNPs, was at least 10 cycles past the beginning of the reaction plateau. The reaction was then finally cooled to 40°C and held for 20 s.

Following the asymmetric target amplification, the reaction plates were unsealed and 0.5 μM of the oligo probe complementary to the excess reverse strand of the amplicon was added. The plates were resealed and then heated to 95°C for one minute to melt all double-stranded DNA species apart. Subsequently, the plates were rapidly cooled to 45°C and held for one minute to allow reannealing of all possible probe and amplicon duplexes. Next, the reactions were heated to 95°C at the maximum rate allowed by the Roche LightCycler 480 (> 0.02° C/s). During this heating period, the LightCycler 480 monitored the amount of fluorescence over time and collected over 1300 data points over a 50°C range, producing a high-resolution graph of the melting profile of each species of duplex DNA in the reaction.

2.2.3.4 Analysis of denaturation profiles

The negative first derivative of the melt stage's decreasing fluorescence curve were calculated to produce a profile with discrete peaks centred at the melting

temperature of each dissociating duplex (Fig. 2.2c,d). Next, a baseline connecting the linear regions bounding the melting peaks was drawn to determine the background rate of dye disassociation. Then, the height of the probe melting peaks above the baseline was measured using a custom [R] script (Supplementary S.1). For pooled reactions, three technical replicates were performed for each locus and sample pair and their relative peak heights averaged. The genotype of a sample was determined by the presence and position of the probe peaks, relative to other samples for the same locus and to the calibrator melting peaks. The frequency of the low-melting allele in a pooled sample was estimated by the contribution of its corresponding melting peak to the total summed height of probe-related melting peaks. Across all SNP assays, a minor allele was called only if it exceeded a frequency of 2%.

2.2.3.5 DNA titrations

To demonstrate a linear relationship between known frequencies of a given allele and the qHRM-estimated frequency estimated from a pooled sample, two individuals homozygous for different alleles of the same SNP were first identified. This experiment was performed for two SNP loci mined from the *A. millepora* transcriptome, C22162S248 and C45133S676 (Wang et al. 2009). The concentrations of DNA were normalized with qPCR as described above, then the two homozygotes were mixed in different proportions to obtain mixtures that incremented the frequency of the low-melting allele from 0-100% in 5% or 10% steps (Table 2.2). qHRM was then performed on each mixture with five replicates for each mixture for two independent normalization and mixing trials, for both SNPs assayed. The estimated allele frequencies for all replicates of each mixture were then averaged and compared to the known allele frequencies.

Table 2.2. Known frequency mixtures and their qHRM-estimated frequencies of a low-melting allele for three SNP loci: C22161S248; C45133S676; C70S236. Sample batches that were pooled at gradient allelic ratios.

SNP C22161S248

# Total replicates	Expected freq. of low-melting allele	Average qHRM-estimated freq.	Standard deviation
10	1	0.9567	0.093
10	0.95	0.8265	0.0481
10	0.9	0.7997	0.057
10	0.85	0.7655	0.0345
10	0.8	0.7148	0.0447
10	0.75	0.6822	0.0541
10	0.7	0.658	0.0396
10	0.6	0.5393	0.1178
10	0.5	0.5164	0.044
10	0.4	0.4035	0.0878
10	0.3	0.3594	0.0505
10	0.25	0.3085	0.0604
10	0.2	0.2925	0.0473
10	0.15	0.2241	0.0493
10	0.1	0.1954	0.0393
10	0.05	0.1424	0.0172
5	0	0	0

SNP C45133S676

# Total replicates	Expected freq. of low-melting allele	Average qHRM-estimated freq.	Standard deviation
5	1	1	0
10	0.95	0.8399	0.0525
10	0.9	0.7404	0.0489
10	0.85	0.6012	0.074
10	0.8	0.5057	0.0621
10	0.75	0.5719	0.0544
9	0.7	0.5316	0.0832
10	0.6	0.4658	0.1022
10	0.5	0.3594	0.0508
10	0.4	0.2882	0.0498
10	0.3	0.2363	0.0618
10	0.25	0.1906	0.0303
10	0.2	0.1634	0.0307
10	0.15	0.1303	0.0233
10	0.1	0.0956	0.0132
9	0.05	0.0613	0.018
9	0	0.0082	0.0106

SNP C70S236

# Total replicates	Expected freq. of low-melting allele	Average qHRM-estimated freq.	Standard deviation
2	1	1	0
2	0.9	0.9138	0.009
2	0.85	0.861	0.0098
2	0.8	0.7738	0.0113
2	0.75	0.7429	0.0212
2	0.7	0.7102	0.016
2	0.65	0.6405	0.0142
2	0.6	0.5858	0.0175
2	0.5	0.5109	0.0176
2	0.5	0.5	0
2	0.4	0.4064	0.0075
2	0.35	0.3608	0.0001
2	0.3	0.3076	0.0029
2	0.25	0.2689	0.0028
2	0.2	0.1983	0.0061
2	0.15	0.1695	0.0024
2	0.1	0.1084	0.0024
2	0	0	0

2.2.3.6 Analysis of 384 loci in six individuals

DNA samples from three adult coral individuals sampled from Trunk Reef and three from Orpheus Island were genotyped individually by conventional HRM for 384 loci selected from SNPs previously developed for a previously published coral linkage mapping project (Wang et al. 2009) (Supplementary Table S.2). Following this, the samples were normalized for coral DNA concentration as explained above and pooled to generate a DNA sample with known allele frequencies for each SNP locus. qHRM was performed in triplicate on the pooled sample for the same 384 SNPs. To fully validate qHRM of each SNP in this panel, a validation analysis was done for only data from SNPs for which all six individuals were both successfully amplified and had easily distinguishable peaks when analysed using conventional HRM, and which had clear qHRM peaks greater than 2% of the total fluorescence signal in the reaction (usually relative to the amplicon peak's height). At least two successful replicates were required for each SNP in the qHRM reactions for analysis to proceed.

2.2.3.7 Analysis of 384 loci in 98 individuals collected from two locations

The peaks method was used to compare allele frequencies in two populations of corals collected from Magnetic Island (n = 48) and Orpheus Island (n = 50). DNA was normalized, pooled according to reef of origin, and qHRM-genotyped in triplicate using a panel of 384 SNPs. Nineteen of these SNPs were further validated by individual HRM genotyping of each individual contributing to the pools. The set of SNPs to validate was selected to represent loci where the low-melting allele was major (8 SNPs) or minor (11 SNPs), loci with small or large minor allele frequencies (11 SNPs with MAF < 0.25; 8 SNPs with MAF > 0.25), SNPs with more than two alleles (7 SNPs with third alleles in one or

both populations), and loci selected to span the range of estimated F_{ST} values (7 SNPs with F_{ST} less than 0.01; 4 SNPs with F_{ST} between 0.01 and 0.02; 8 SNPs with F_{ST} greater than 0.02, including four of the top seven SNPs with the greatest F_{ST}).

2.2.3.8 Analysis of genetic subdivision between coral populations

SNP allele frequencies were compared between Orpheus and Magnetic populations by plotting the averaged qHRM-estimated frequencies against each reef.

F_{ST} for each SNP was calculated according to Equ. 1, where *p* is the frequency of the high-melting allele, *q* is equal to (1-*p*), $\overline{p}$ is the average frequency of *p* allele between the two populations, and $\overline{q}$ is equal to (1- $\overline{p}$).

Equ. 1: $F_{ST} = \frac{(p^1 - p^2)^2}{4\overline{p}\overline{q}}$

Global F_{ST} was calculated by averaging all SNP F_{ST} values.

2.2.4 Curves method

The curves method was developed to extend qHRM and demonstrate functionality across another dye chemistry and analytical approach. This method is best suited for highly accurate determination of allele proportions across small numbers of SNPs. It requires the usage of three reference sequences (two homozygous samples and a heterozygous sample) for comparison with an unknown, pooled sample. The curves method demonstrates that reactions are extremely reproducible and robust among different markers and pooled samples. The workflow of the curves method is illustrated in Fig. 2.1. Briefly, the curves method was validated through an initial proof of concept experiment using different known proportions of alleles for a single SNP marker. This

marker was additionally validated on three DNA pools made from 43-56 individuals representing distinct coral populations by comparing the true allelic frequency of the pool (revealed by summing the genotypes of each individual) to the qHRM-estimated frequency. Thirdly, reproducibility of this method was investigated by using five SNP markers for pooled samples representing 25 coral populations.

2.2.4.1 Pooling

Each individual sample was amplified individually in an asymmetric qPCR reaction using primers for the locus C70S236 (Wang et al. 2009) under the following conditions using a Rotor-Gene Q machine (Qiagen): 0.1 μM forward primer, 1 μM reverse primer, 10 ng holobiont DNA, 1x 7.5 μl of the Accumelt HRM supermix (Quanta Biosciences) in a 15 μl volume. The reaction mixture was heated at 95°C for 10 min, then cycled 40 times with the following thermoprofile: 95°C for 40 s, 58°C for 40 s, and 72°C for 40 s. DNA concentrations were normalized according to Pfaffl (2001) as summarized in the Peaks - pooling section. Experiments also contained tubes without DNA template (negative controls).

2.2.4.2 Asymmetric template amplification and melt stage

After normalization of individual samples, each individual was combined into a single pooled sample. The pooled sample was then amplified asymmetrically with the Rotor-Gene Q machine (Qiagen) with the locus C70S236. Next, 1 μM of probe complementary to the excess strand was added to each reaction. Then, the reactions were heated to 95°C and held for 60 s, rapidly cooled to 45°C and held for 150 s, heated again to 95°C at a rate of 0.1°C/s with a 2 s hold each step, collecting 500 data points in total.

2.2.4.3 Analysis of denaturation profiles

Because many factors such as pipetting errors causing variations in quantity of fluorescent dye and DNA can affect the relative fluorescence levels among normalized DNA samples, it is important to standardize the fluorescence levels to exclude the noise for accurate allele frequency measurements. Fluorescence values of the SNP-specific melt curves were standardized by selecting regions before and after the probe melting phase where nucleotide differences do not lead to variations in fluorescence level, then averaging fluorescence levels of the selected regions among all samples using the Rotor-Gene Q Series Software 2.0.2.4 (Corbett) in order to mitigate the effects of sample- and SNP-specific variation (Fig. 2.2e). The exclusion of noise through this normalization process allows the direct comparison of melting curves from different reactions.

The inclusion of three standard reference samples (positive controls) of known genotype (two different homozygotes and a heterozygote) allows for even small differences in allele frequency to be resolved. These standards set a reference fluorescence level to which unknown samples can be compared for absolute allele frequency quantification (Fig. 2.2f) and also serve to reduce variations between runs.

The y-axis point at which the greatest difference in fluorescence between the two homozygotes occurs is used to determine the relative proportion of alleles in an unknown sample as it gives the highest accuracy in estimating allele frequencies (Fig. 2.3). This point should also be near where the inflection point of the heterozygote's curve falls. But, because the heterozygote can suffer from amplification biases, the position of the inflection point can vary. The accuracy increased by calculating the allele frequency of the unknown sample with respect to the heterozygote rather than simply using the midpoint of the two homozygotes. For example, if an unknown sample has a higher frequency of the high-melting allele (i.e. the unknown's curve falls above of the heterozygote's curve and/or the point of the greatest distance between homozygotes),

then Equ. 2 can be used to estimate the proportion of the high-melting allele. If the unknown has a higher frequency of the low-melting allele (i.e. the unknown's curve falls below the heterozygote's curve and/or the point of the greatest distance between homozygotes), then Equ. 3 should be used to calculate the frequency of the high-melting allele. In both equations, the proportion of fluorescence at the inflection point of the unknown sample (*x*) relative to the heterozygote (f_{het}) was calculated. After adjusting for empirical differences between the known heterozygote and one of the known homozygotes (f_{high} or f_{low}), the value was added or subtracted from 0.5 to determine the allele frequency of the unknown sample relative to the low-melting homozygote.

Equ. 2: frequency of high-melting allele = $0.5 + 0.5\left(\frac{x - f_{het}}{f_{high} - f_{het}}\right)$

Equ. 3: frequency of high-melting allele = $0.5 - 0.5\left(\frac{f_{het} - x}{f_{het} - f_{low}}\right)$

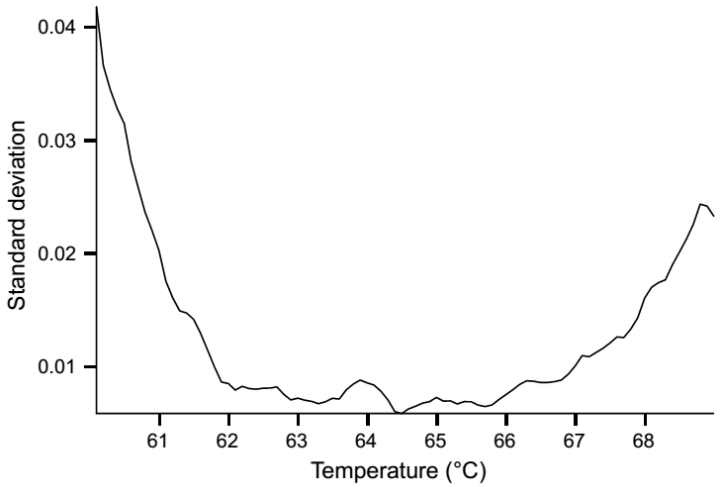

Fig. 2.3 Standard deviations of differences between estimated and expected allele frequencies across melting temperature range. The standard deviation (i.e. inaccuracy) increased towards either end of the melting temperatures. The smallest standard deviation was observed at the temperatures between 64.4-64.5°C, indicating the highest accuracy in estimating allele frequencies at this temperature range. Consistent with this, the largest difference in fluorescence level between two homozygotes was also observed within the same temperature range for C70S236 marker.

2.2.4.4 DNA titrations and method validation

The use of three reference genotypes per SNP assay and the fluorescence standardization step should theoretically allow for great sensitivity to detect allele frequency differences among pooled samples. To assess this, multiple individuals of each of the three possible genotypes for the locus C70S236 were identified and DNA concentrations of each sample were normalized. DNA samples were then combined into one pool per genotype and mixed in varying proportions to generate a gradient of alleles spanning 0-100% of the low-melting allele by 5-10% increments. The estimated allele frequencies for two technical replicates of each mixture were averaged and compared to the expected allele frequencies (Table 2.2).

2.2.4.5 Application to large pools of DNA

DNA sampled from individual corals sourced from Myrmidon (n = 43), Night (n = 56) and Wilkie (n= 49) Reefs was normalized and individually genotyped using SNP C70S236. Normalized samples were then pooled by population and quantitatively genotyped in duplicate in order to examine the accuracy of the curves method when used with many individuals per pool.

2.2.4.6 qHRM reproducibility

Next, the reproducibility of qHRM on 25 populations sampled along the Great Barrier Reef was evaluated. DNA from 35-56 individuals for each population was pooled, and then curves qHRM was performed on each pool in duplicate for five SNPs (C29226S281, C70S236, C11461S560, C16774S791 and C20479S292), using samples normalized as per above.

Consistency between the two replicates was measured by estimating the half-confidence interval of the difference as follows:

$$\frac{z \times SD}{mean}$$

Standard deviations (SD) and mean values were calculated using fluorescence levels in the temperature range where the largest difference in fluorescence levels between the two homozygotes were observed (Fig. 2.2f). A z value of 1.96 was used so the measure of precision represents half of a 95% confidence interval.

Furthermore, the qHRM reproducibility was examined for the 5 markers and 25 reefs by comparing mean errors. Errors in percentage were calculated as follows:

Percentage of error between replicates = $\left(\frac{f_1 - f_2}{(f_1 + f_2)/2}\right)$

where f_1 and f_2 are two replicates of fluorescence data. Geometric means of errors were used as population- and marker-specific errors.

2.3. Results

2.3.1 Peaks method

2.3.1.1 Allele titrations

As an initial proof of concept, a validation analysis was done on DNA from two adult individuals determined to be homozygous for different alleles of the same SNP, mixed in varying proportions to represent a spectrum of allele frequencies. The probe peaks for three allele frequency examples (50%, 25%, 15%) are presented in Fig. 2.4a. For the two SNPs analysed in this way, there was a strong linear correlation between the qHRM estimations and the true proportion (Pearson *r* = 0.97-0.99). However, the slope of the regression for both SNPs was 0.80 (Fig. 2.4b,c).

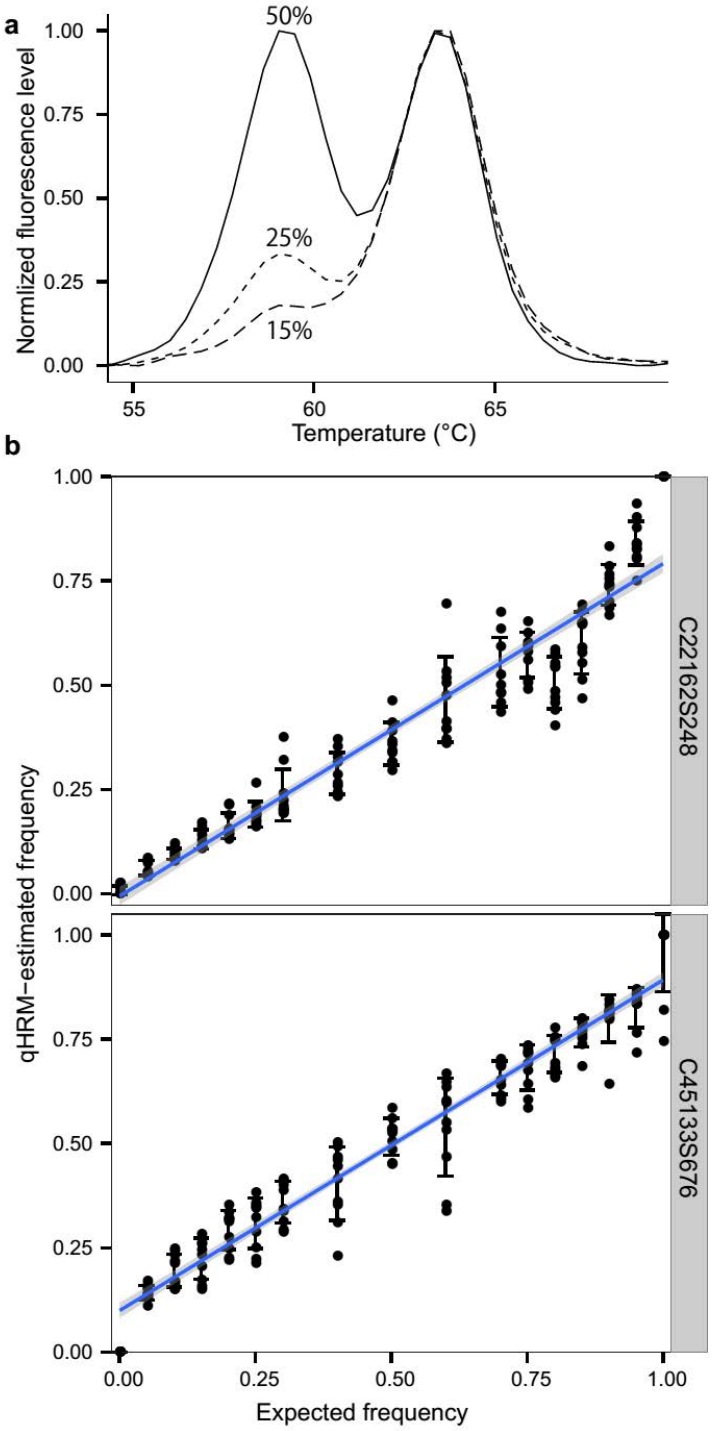

Fig. 2.4 DNA titrations using the peaks method. a) Probe melting peaks for different allele titrations demonstrating clear resolution among allele frequencies (15%, 25%, 50%). Two homozygotes were mixed in varying proportions to test qHRM using b) SNP C45133S676 (Pearson's *r* = 0.97, regression slope = 0.82) and c) SNP C22162S248 (Pearson's *r* = 0.99, regression slope = 0.80). The qHRM estimates were tightly correlated to expected frequencies; however, qHRM appears to underestimate the true allele frequency for the low-melting allele for both SNPs considered.

2.3.1.2 Analysis of 384 loci in six individuals

To further validate the method, DNA samples from six adult coral colonies from two reefs were individually genotyped for 384 SNPs selected from the coral linkage map using conventional HRM. DNA from these six individuals was pooled and the allele frequency at each SNP was estimated using qHRM. Of the 291 SNPs that passed quality filtering, 54 were homozygous, all of which were correctly identified by qHRM. For 237 polymorphic SNPs, the Pearson correlation between the expected and estimated allele frequencies was 0.89 with a regression slope of 0.94 (Fig. 2.5), qHRM failed to detect polymorphism for eight of the polymorphic SNPs, all of which had minor allele frequencies of 8%, the lowest possible allele frequency in this experiment.

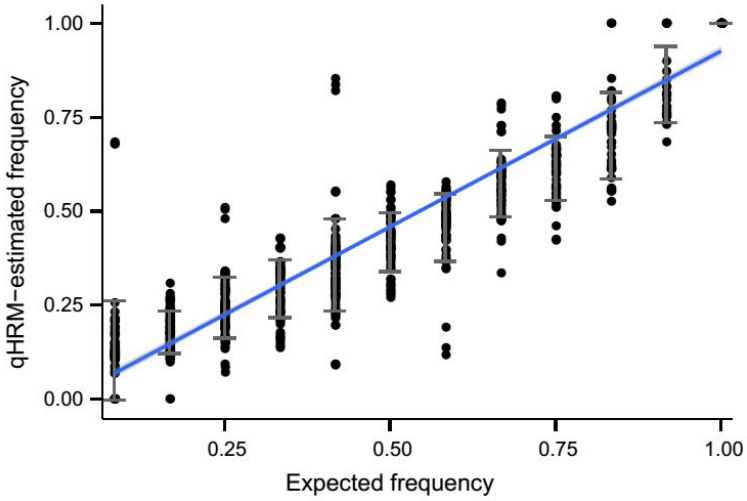

Fig. 2.5 Analysis of 384 loci in six individuals (Peaks method). Six adult corals were HRM-genotyped individually for 384 SNPs, then qHRM-genotyped as a pool in triplicate. The expected allele frequency was calculated by summing the genotypes for each individual (12 alleles possible), and the qHRM-estimated frequencies were calculated as described in the Materials and Methods section. The Pearson correlation between the expected and the qHRM-estimated values is 0.89. The slope of the linear regression is 0.94.

2.3.1.3 Analysis of 384 loci in 98 individuals collected from two locations

To demonstrate that qHRM can identify SNPs with different allele frequencies among real populations, the same 384 SNPs in pooled samples representing the Magnetic Island and Orpheus Island populations of *A. millepora* were surveyed. These populations are geographically close together but have very different environmental parameters and some degree of genetic separation (F_{ST} = 0.041) (van Oppen et al. 2011).

Of the 384 SNPs assayed, 261 were represented by at least two technical replicates in each population, including 34 monomorphic loci. Notably, the qHRM-estimated allele frequencies for the 261 loci were very similar among populations (Pearson's *r* = 0.93, slope = 0.97) (Fig. 2.6a). Global F_{ST} between the Magnetic and Orpheus populations using 261 SNPs was calculated as 0.006.

The qHRM allele frequency calls for 19 SNPs were further validated by individually genotyping members of each population (total *n* = 98) using conventional HRM. The Pearson correlation between expected allele frequencies and the qHRM-estimated frequencies for these 19 SNPs was 0.93 with a regression slope of 0.77 (Fig. 2.6b). Three of the 19 SNPs from Orpheus Island and two from Magnetic Island populations failed to generate reliable HRM data for one to six individuals, suggestive of a mutation in the HRM primer biding site(s), and therefore were not possible to fully validate. However, it is likely that these missing individuals also failed to contribute to the pooled sample's allele frequency estimate.

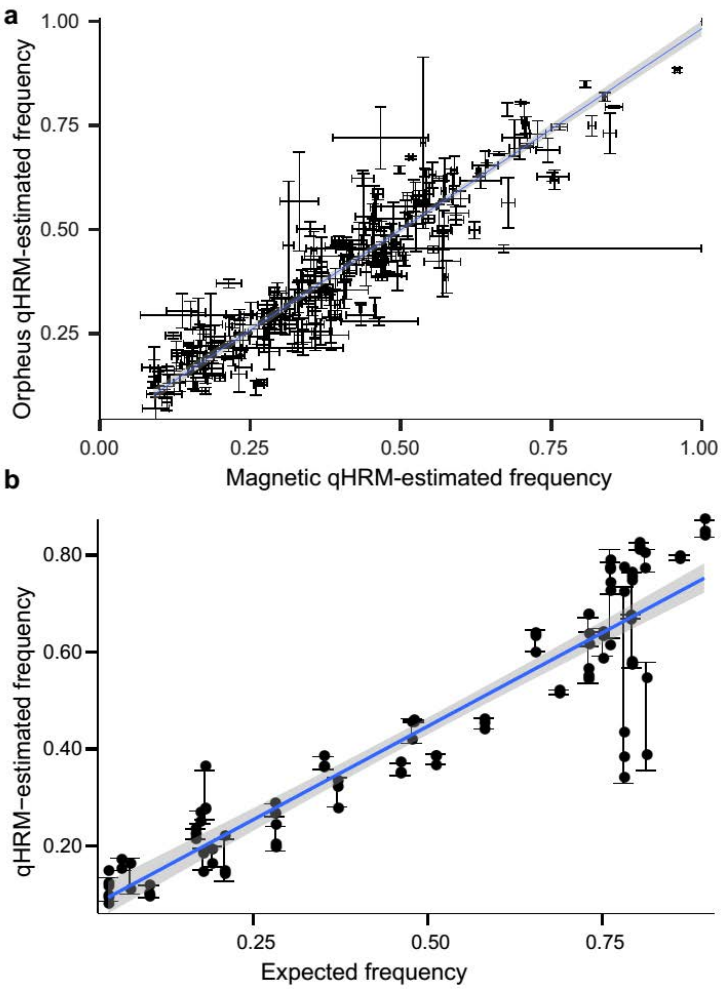

Fig. 2.6 Analysis of 384 loci in 98 individuals collected from Orpheus Island and Magnetic Island. a) 48 and 50 individuals from the respective reefs were pooled and quantitatively genotyped using qHRM. The allele frequencies between the reefs were tightly conserved between the two populations, with a Pearson's *r* of 0.93 and a slope of 0.97. b) 19 SNPs were validated by genotyping individual members of each pool. The Pearson correlation between known and pooled genotypes for these SNPs is 0.93 with a regression slope of 0.77.

2.3.1.4 Detection of novel alleles

Unlike other technologies that are only able to detect the alleles sought and may therefore suffer from ascertainment bias, HRM is able to detect novel sequence variants within the probe-SNP duplex, which manifest as additional melting peaks. During the course of this study, qHRM suggested the presence of a third SNP within the interrogated probe region in 34 (9%) of SNP assays tested for the pool of six individuals collected from Trunk Reef and Orpheus Island. In another assay, four distinct melting

peaks were found across six separately HRM-genotyped individuals with a maximum of two alleles for each sample, though a separate SNP assay revealed three alleles within a single sample. This could indicate a high level of polymorphism (multiple SNP sites within the 20-base probe region), a potential genomic duplication event, a somatic mutation or a chimeric individual (Fig. 2.7). Each unusual case was validated by individual HRM genotyping.

Additionally, for the analysis of the Orpheus Island and Magnetic Island populations, third alleles were identified in qHRM for 9 SNP assays in the Magnetic Island pooled sample and for 10 SNP assays in the Orpheus Island pooled sample. Of the 19 assays included in the individual validation experiment, 7 revealed three alleles in one or both populations. For several of these assays, the third allele was not detected by qHRM because its frequency was too low or was otherwise undetectable in the pooled sample.

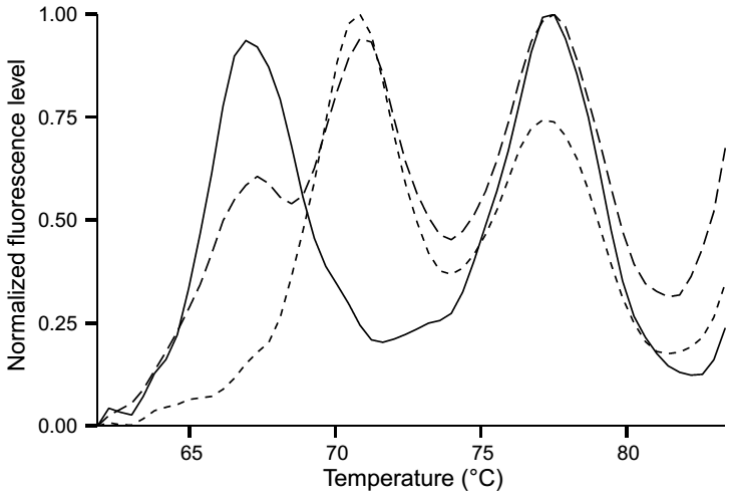

Fig. 2.7 qHRM detection of novel alleles and possible gene duplication event. The probe peaks for SNP C1023S218 show two individuals heterozygous for different alleles and the presence of a third allele within a single individual. Three individuals of the six genotyped for this locus showed the same pattern of three alleles each.

2.3.2 Curves method

2.3.2.1 Allele titrations

As with the peaks method, the curves method showed a high correlation between true and estimated allele frequencies with only two replicates using the SNP marker C70S236 (Pearson's r = 1 and slope = 0.982) (Fig. 2.8). Furthermore, error rates ranged between 0.0017 and 0.026 with a standard deviation of 0.006, which verifies the high accuracy of the curves method for absolute quantification of allele frequencies with this marker (Fig. 2.8a,b).

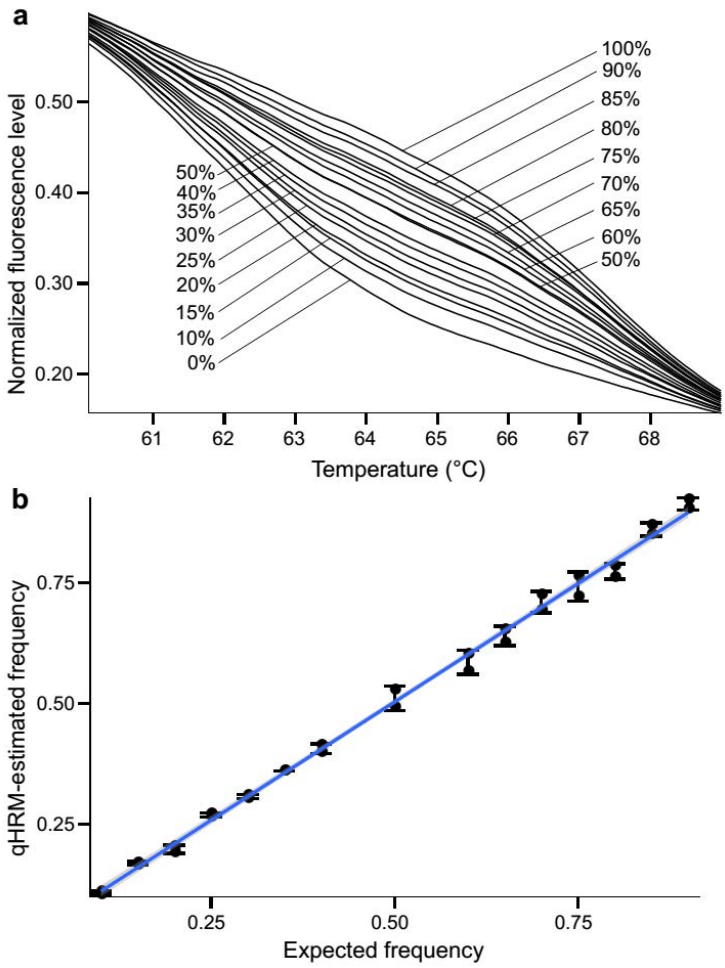

Fig. 2.8 DNA titrations using the curves method. a) HRM normalized curves produced from sample batches that were mixed at gradient allelic ratios. b) A linear regression showing a relationship between estimated and expected low-melting allele frequencies of the C70S236 locus. Pearson's r = 1. Slope=0.982.

2.3.2.2 Application to large pools of DNA

The curves qHRM method proved highly accurate on pooled DNA samples. The allele frequencies of Myrmidon, Night and Wilkie populations were estimated by both individual HRM and pooled qHRM genotyping using the marker C70S236. There was very close agreement between the true allele frequency, as calculated by summing the allele frequencies for each individual contributing to each pool, and the qHRM-estimated frequency. The differences between estimated allele frequency and the true values were 0.9 % for Night, 1.1 % for Myrmidon, and 3.6 % for Wilkie Reef. An outlier sample was identified in Wilkie population (Fig. 2.9).

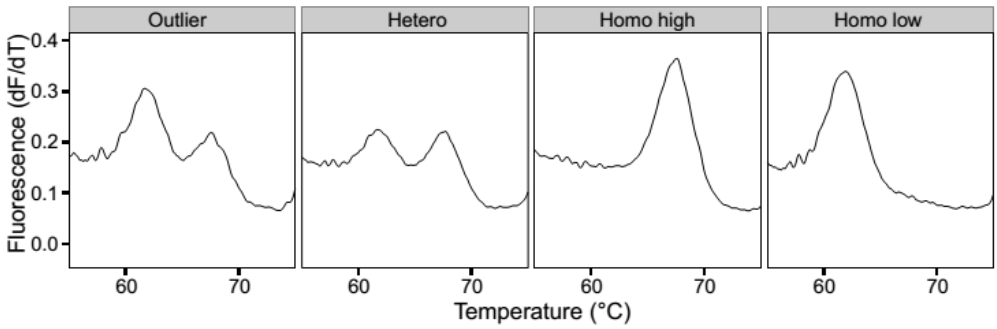

Fig. 2.9 Outlier sample identified from Wilkie population using the SNP marker C70S236. The outlier sample has two peaks with low-melting peak being higher than the other. A heterozygous (Hetero) and two homozygous (Homo high and Homo low) were clearly different from the outlier, indicating that the outlier sample may cause an error in allele frequency estimates for Wilkie population.

2.3.2.3 Reproducibility of qHRM

The curves method demonstrated high precision in measuring fluorescence levels between replicates (<2% of Half-CI with mean errors of <0.4%) for the five markers examined with this method (Table 2.3). Population-specific mean errors for all populations ranged between 0.1 and 0.75%. A simple calculation using average fluorescence values of the two homozygote reference samples shows that 1% of fluorescence error is equivalent to about 0.14% allele frequency error for C70S236.

Table 2.3 Replicate errors associated with 25 populations and 5 SNP markers.

Location	C70S236	C29226S281	C11461S560	C20479S292	C16774S791	Mean error
North Keppel Island Reef	0.27	0.63	0.36	2.32	1	0.68
Calder Reef	0.27	1.32	0.71	0.05	0.26	0.32
Sudbury 2 Reef	0.22	0.49	0.25	0.28	0.39	0.31
High Peak Reef	0.69	0.12	0.22	0.32	0.48	0.31
NE Orpheus Reef	1.24	0.71	0.9	0.84	0.16	0.64
Wallace Islet Reef	1.25	0.01	0.66	0.69	0.01	0.13
Sudbury 1 Reef	0.55	0.25	1.29	0.36	1.12	0.59
Boulton Reef	1.59	0.92	0.41	0.31	0.95	0.71
21-121 Reef	0.35	0.02	0.05	0.36	0.08	0.1
Goble Reef	0.49	0.32	0.99	1.35	0.13	0.49
Holbourne Island Reef	0.46	0.64	0.28	1.22	0.04	0.33
N Fantome Reef	0.77	0.44	0.35	0.17	0.28	0.36
SE Pelorus Reef	0.29	0.09	0.33	0.16	0.4	0.22
Emily Reef	3.78	0.87	0.12	1.17	0.32	0.69
Half Tide Reef	0.55	1.12	1.21	0.05	0.49	0.45
Magnetic Island Nelly Bay	0.68	0.23	0.25	0.53	0.16	0.32
Humpy Reef	0.06	0.57	0.05	0.61	0.54	0.22
Halfway Island Reef	0.54	1.06	0.58	0.28	0.2	0.45
Kelso Reef	0.58	0.22	0.84	0.15	0.94	0.43
Rib Reef	1.56	0.77	0.57	0.08	0.06	0.32
Ross Reef	1.46	0.45	0.95	0.35	1.07	0.75
Darley Reef	0.6	0.85	0.14	1.16	0.83	0.58
Myrmidon Reef	0	0.14	0.2	0.45	1.3	0.11
Wilkie Reef	0.77	0.13	0.59	0.6	0.8	0.49
Night Reef	0.01	0.24	0.5	0.9	1.2	0.23
Mean_error	0.42	0.36	0.42	0.43	0.36	
Half-CI	1.99	1.19	1.35	1.66	1.11	

2.4 Discussion

2.4.1 Comparisons with other genotyping methods

Genotyping using next-generation sequencing, although no doubt powerful and versatile, is neither feasible nor necessary for applications that must maximize the number of genotyped populations rather than the number of genotyped loci. For example, genetic connectivity studies might not require more than a hundred SNP markers (Morin et al. 2009), but would instead benefit from careful selection of the markers to ensure that they are physically unlinked, adaptively neutral and polymorphic

across the geographical range of interest (Helyar et al. 2011). Determining their allele frequencies across a large number of distinct populations, on the other hand, can be uniquely beneficial for identifying physical and ecological barriers to migration and disentangling the effects of geographic and ecological isolation on genetic differentiation (Bradburd et al. 2013). Profiling allele frequencies at a few SNPs across numerous populations can be an efficient approach to validate potential targets of natural selection ("outlier" SNPs) suggested by genome scans (Foll & Gaggiotti 2008; Helyar et al. 2011).

The main hurdle of the approach for population or ecological genomic studies in non-model organisms, is overcoming the difficulty and expense of genotyping a sufficient number of individuals to accurately estimate allele frequencies of a population. Some technologies are able to overcome this barrier by using DNA pooling methods in which many individuals can be combined into a single reaction for each group, drastically improving the cost and time efficiency of a study (Sham et al. 2002). However, not all genotyping technologies are quantitative and therefore may not be appropriate for use with pooled samples.

Other quantitative applications of HRM have been described (Palais et al. 2005; Balic et al. 2009; Malentacchi et al. 2009; Mader et al. 2011; Ganopoulos et al. 2013; Sakaridis et al. 2013a; Sakaridis et al. 2013b; Stuopelyte et al. 2013; Lin & Gänzle 2014), but to date these methods have been designed to detect relatively dramatic differences in sequence between alleles. These methods have not been successfully applied to the relatively cheap unlabelled probe method of qHRM, where differences between single SNPs are magnified through the use of a small oligo probe. In addition, unlike other qHRM methods that use duplex PCR to amplify different loci or different alleles (Wittwer et al. 2003; Liew et al. 2004; Dimitrakopoulos et al. 2012), the method presented in this

study should not suffer from biases due to different efficiencies of PCR primers' binding to alternative alleles because the interrogated SNP is in the middle of the amplicon.

Various quantitative comparison methods that analyse the relative contribution of each variant in a pooled sample have been designed, including determining the amount of fluorescence at an allele's melting curve inflection point (Ganopoulos et al. 2013), subtracting the unknown sample's melting curve from a homozygous reference's melting curve and comparing the heights of differential curve's peaks (Malentacchi et al. 2009), or deriving the melting curve into a melting peaks graph and analysing the areas under the resulting peaks (Lin & Gänzle 2014). However, these methods are not directly applicable to the cost-efficient unlabelled probe qHRM method presented in this study. Because of error introduced when normalizing the probe region subset of a melting curve due to the baseline disassociation of the dye, as well as the necessity to temperature-shift assays across different plates using the internal calibrator sequences, comparing the fluorescence level at an inflection point is difficult. Additionally, both this inflection-point comparison method and the subtractive method require the usage of reference samples, which scales poorly when using more than a few SNP markers. The melting peaks' area analysis presented in Lin and Gänzle (2014) is sophisticated, but it requires the use of proprietary software, deconvolution of peak curves and manual peak modelling. Though accurate, the expensive software and the time involved in calculating peak areas may limit the application of this type of analysis for medium-throughput projects.

2.4.2 Pooled genotyping by HRM and unlabelled probes

2.4.2.1 Peaks method

In this study, Pearson's correlations between expected allele frequencies and estimated allele frequencies were high for all experiments. The regression slopes of the experiments validated via individual genotyping (allele titration, two SNPs; six individuals, 291 SNPs; and reef comparison, 19 SNPs validated for members of each population) ranged from 0.77 to 0.98. The peaks qHRM method slightly underestimates the true allele frequency of the low-melting allele (probe-mismatch allele). Ostensibly, this could be due to a bias during probe-DNA duplex formation during the initial cooling phase of HRM that favours the perfectly-matching probe. Regardless of its cause, this minor bias is unlikely to affect the conclusions of genetic studies outlined in the beginning of this Discussion since such studies rely upon variation in allele frequency among populations rather than its mean. Individual genotyping showed that two populations had very similar allele frequencies and the qHRM method similarly recovered these allele frequencies indicating that the peaks method is reproducible. An alternative to counteracting the effects of the rare alleles and DNA cross-contamination on allele frequency estimation is to increase sample size and thereby reduce errors caused by a minority of samples. Johnston et al. (2013) showed improved mean adjusted R^2 and mean difference between estimated and true allele frequencies with increasing numbers of samples (also see Vaillant et al. 2002).

2.4.2.2 Curves method

The high reliability of the curves method (Pearson's r = 1 and slope = 0.982) is attributed to the inclusion of the heterozygote as reference sample and identification of the temperature region that gives the least error in allele frequency estimation. However, since the curves method of qHRM analysis uses reference genotypes to determine the relative contribution of each allele in an unknown sample, an allele that is not represented by the reference genotypes will be missed as a fluorescence produced

by a third allele would be merged and difficult to detect by fluorescence of other alleles when pooled. Additionally, it should be mentioned that pooling of samples leads to loss of linkage information as a proportion of an allele rather than genotype is estimated in a pooled sample.

The low level of sample-specific errors in the curves method proves that the replicate accuracy was not significantly affected by differences in pipetting errors and DNA quality among samples. The accurate replication among samples of different DNA qualities can be attributed to a short length of PCR product (about 100 base pairs) (Fernandez-Jimenez et al. 2011).

Gruber et al. (2002) found the accuracy of pyrosequencing was compromised in SNPs that were flanked by identical bases to the SNP alleles. In contrast, the qHRM method (the curves method) showed a clear difference in fluorescence level among genotypes for three SNPs (C11461S560, C16774S791 and C20479S292) despite the presence of identical flanking sequences (Fig. 2.10). This simplifies primer design and eliminates the need for sequencing SNP sites for primer optimization. High accuracy and reproducibility of allele frequency calls for all the 5 SNPs tested for the curves method support its applicability to other biallelic markers.

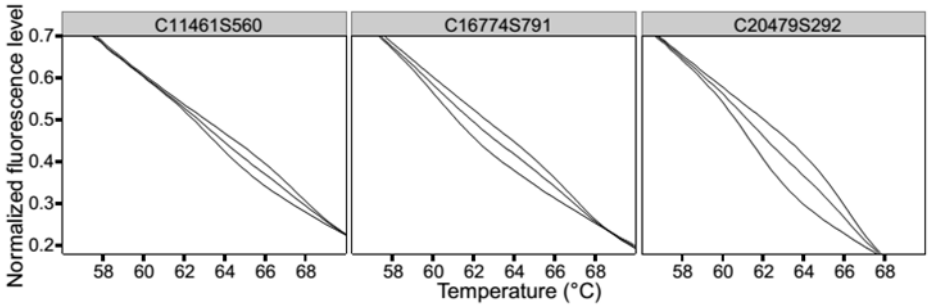

Fig. 2.10 Melt curves of three reference samples. Two homozygotes and one heterozygote show the same pattern of normalized fluorescence level among markers with different combinations of nucleotides at SNP sites.

2.5 Conclusions

The proposed qHRM pooling methods demonstrated high accuracy and reproducibility in several cross-validation experiments. The key advantages of these methods over other targeted allele frequency profiling techniques include the ease of assay design and low overall cost of the analysis. Total costs are reduced as samples are not tested individually, but as a single pooled sample. All of these factors make qHRM a method of choice for cases requiring genotyping a few hundred SNP markers across a large number of populations.

Chapter 3.0 Allele frequency differences along water quality and temperature gradients at SNP loci in panmictic coral populations

This chapter and Chapter 4 were combined and submitted as a paper to Science Advances.

Young K. Jin, Petra B. Lundgren, Bette L. Willis and Madeleine J.H. van Oppen.

YKJ, PL, BLW and MJHvO conceived the study. YKJ conducted and collected data from gene by environment association experiment.YKJ wrote the chapter with intellectual contributions from all co-authors.

Abstract

Environmental gradients in parameters such as temperature and water quality provide important opportunities for detecting adaptive signatures in natural populations. The reef-building coral *Acropora millepora* is exposed to a wide range of environmental conditions that vary throughout its distribution. Field and laboratory experiments show that variation in adaptively important phenotypic traits exists among ecologically dissimilar regions and within populations. Identification of the genomic basis for within-population variation will help to decouple acclimatisation (phenotypic plasticity) from adaptation and increase understanding of the adaptive dynamics of genetic changes. In this study, the distributions of SNP allele frequencies at 5 functional loci were investigated in 18 populations of *A. millepora* along the Great Barrier Reef. Significant correlations between allele frequencies and gradients in both water quality and temperature were detected at two functional loci. Environmental correlation network analysis distinguished two large clusters corresponding to temperature-related and water quality-related variables. Strong correlations among variables within each cluster suggest independent links between individual metrics within each environmental category and allele frequencies. A multiple linear regression model examining the combined effects of NO_3 and sea surface temperature (SST) range explained 48% and 58% of variability of allele frequency patterns for the two loci. Gene by environment association analysis identified loci that may represent quantitative trait loci for tolerance to temperature and water quality stressors. The identification of quantitative trait loci for these major coral stressors represents a significant step forward in the understanding of host genetic factors controlling stress tolerance, although further field and aquarium testing of the validity of these QTLs is required (see Chapter 4).

3.1 Introduction

Ecologically important traits in many organisms are known to be correlated with environmental variation along latitudinal gradients (Hallas et al. 2002; Heibo et al. 2005; Zhen & Ungerer 2008). These associations between traits and environmental variation provide strong evidence for local adaptation. Divergent selection pressures exerted by contemporaneous environmental gradients provide an excellent opportunity to detect carriers of potentially adaptive traits. In particular, phenotypic traits that are adapted to local environments are sometimes mirrored at the level of genetic polymorphisms (Umina et al. 2005; Mitchell-Olds et al. 2007; Schmidt et al. 2008). One of the best examples for rapid genetic adaptation to current environmental conditions is found in the American eel (Gagnaire et al. 2012), for which genetic signatures of post-dispersal local selection along environmental gradients were demonstrated in spite of the existence of panmixia. As a result, investigating genotypic variation along environmental gradients is a means to discover genotypes that have a higher tolerance to factors such as higher temperatures. This in turn can be used to predict the capacity for populations to adapt to environmental change through the dispersal of more tolerant genotypes from populations that currently exist further along the gradient.

It is now widely accepted that rapid ocean warming negatively affects a wide range of coral reef species, including reef-building corals. Corals are the foundational species of the coral reef ecosystem and form complex structures that provide niches for a range of reef organisms that further increases biodiversity within coral communities. Predicted decreases in associated reef fishes and other biota following the loss of coral structure (Wilson et al. 2006; Pratchett et al. 2008) highlight the importance of identifying stress markers for corals that would alert managers and stakeholders about conditions affecting reef corals.

In addition to global warming, the Great Barrier Reef (GBR) has been experiencing rapid environmental changes due to land modifications adjacent to coral reefs that lead to discharges of sediment, chemicals and nutrients into nearshore waters (Devlin & Brodie 2005). It is well documented that nutrients and sediment loads discharged into the GBR lagoon have increased dramatically since European settlement (McCulloch et al. 2003; McKergow et al. 2005; Wooldridge et al. 2006; Hughes et al. 2009). Bleaching resistance has been shown to be compromised in corals that are exposed to high nutrients (Nordemar et al. 2003; Schloder & Croz 2004; Wooldridge 2009; Fabricius et al. 2013). This cumulative, human-induced environmental alteration raises grave concerns regarding the ability of corals to cope with changing environments (Hughes et al. 2003; Pandolfi et al. 2003).

A number of species of plants and animals show rapid adaptive responses in the face of rapidly changing environments caused by human activities (Stockwell et al. 2003; Berteaux et al. 2004; Jump & Penuelas 2005; Carroll et al. 2007; Bradshaw & Holzapfel 2008), such as industrial melanism in the peppered moth (Cook & Saccheri 2013), increased resistance to heavy metals in plants (Wu & Kruckeberg 1985), and changes in morphology, plumage and song in birds in response to anthropogenic habitat changes (Smith et al. 2008). Rapid adaptations associated with strong selection may also be possible in corals, particularly in cases of high gene flow (Smith et al. 2008; Gagnaire et al. 2012). High gene flow results in high genetic diversity due to increased genetic recombination rates as more unique genotypes are intermingled (Stumpf& McVean 2003). At the population level, rapid adaptation would involve selecting tolerant genotypes from existing variants through strong selective forces (Barrett & Schluter 2008; Baums 2008;Maynard et al. 2008; Guest et al. 2012; Pratchett et al. 2013).Gradients in environmental variables represent an important opportunity to test

whether such selection has occurred in coral populations; however, until now, the lack of appropriate stress markers has limited opportunities for such studies in corals.

Acropora millepora is a widespread and abundant scleractinian coral that is distributed along the entire length of the GBR. Population genetic studies have shown high levels of gene flow, especially in the far northern and central GBR regions (van Oppen et al 2011), making this species an ideal candidate for studies of genotype-environment associations along latitudinal and environmental gradients. In this chapter, I investigate gene by environment associations in *A.millepora* by examining spatial patterns of allele frequencies along gradients in both temperature and water quality.

3.2 Methods and Materials

3.2.1 Study sites

Samples of the coral *Acropora millepora,* collected by scientists at the Australian Institute of Marine Science (AIMS) for a separate study of large-scale gene flow (van Oppen et al. 2011), were genotyped as described in Chapter 2. A total of 25 populations, spanning ~1,550km (~12 degrees of latitude) of the GBR, were sampled (Fig. 3.1). At each location, one branch was collected from each of ~50 colonies. Coral branches were collected in an area of approximately 300m x 300m at a depth between 1 m and 11 m at each site and then fixed in absolute ethanol for DNA analysis. Out of these 25 populations, 20 formed part of the large-scale population genetic study (van Oppen et al 2011).

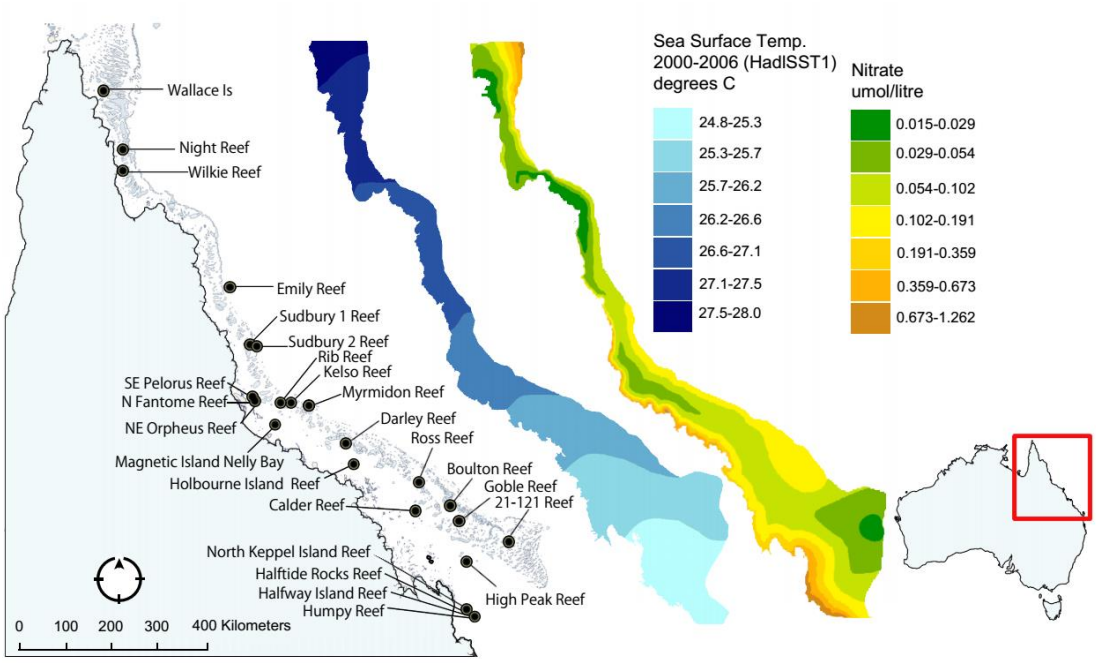

Fig. 3.1 Map showing temperature and nitrate gradients and locations of 25 populations of *Acropora millepora* sampled along the length of the Great Barrier Reef.

3.2.2 Environmental variables

Rising oceanic temperatures and water quality degradation have been identified as major stressors for reef-building corals, therefore 18 variables related to either temperature or water quality were selected for gene-environment association analysis (see Table 3.2). Additionally, three variables that are measures of biodiversity were also included in the analysis, including richness of soft coral zooxanthellate and azooxanthellate taxa and of fish communities (see Table 3.2). As biotic components may be affected by multiple environmental factors, as well as other biotic factors, they are likely to represent an integrated response to the collective effects of the various abiotic factors. Data for the three temperature variables were obtained from two different sources for environmental correlation network analysis and gene by environment association analysis; satellite temperature data were obtained from National Oceanic and Atmospheric Administration (NOAA) (Wooldridge & Done 2004), and logger temperature data from AIMS, respectively (Supplementary table S.3). The remainder of

environmental data for other variables were obtained from spatial prediction maps based on long-term monitoring data gathered by the Australian Institute of Marine Science (http://eatlas.org.au/data/uuid/df7012eb-a23f-4cf2-b92d-6fff77718987) (Supplementary Fig. S.1).

3.2.3 Target SNPs

I used 19 single nucleotide polymorphism (SNP) markers, 9 of which were selected based on a study that revealed frequency variations in SNP alleles in populations of *A.millepora* originating from two environmentally distinct reefs on the GBR (Capper et al. unpublished data.). In summary, they chose Palm and Magnetic Islands (~75km apart) as study sites because water quality characteristics, such as chlorophyll, particulate nitrogen and phosphorus, suspended solids and turbidity, differed significantly between the two sites. Levels of turbidity and levels of variability in all parameters measured, including water temperature, were higher at the Magnetic Island site (Schaffelke et al. 2010). In addition to the 9 SNPs selected from the study by Capper et al. (unpublished data), another 9 SNPs were selected from a gene by environment study (Lundgren et al. 2013) and 1 SNP based on its putative function in ribosomal protein synthesis were tested (Table 3.1).

Table 3.1. SNP markers used in this study

Gene	Marker name*	Reference	Linkage group**
40S_ribosomal_protein_S25_(RPS25)	C6267S266	NA	12
Circadian_locomoter_output_cycles_protein_kaput_(CLOCK)	C20479S292	Capper et al. unpublished data	7
Vacuolar_protein-sorting-associated_protein_36_(vps36)	C29226S281		4
F-box and WD-40 domain protein 5 (FBXW5)	C70S236		5
Bardet-Biedl_syndrome_7_protein_homolog_(Bbs7)	C11461S560		4
C-myc_promoter-binding_protein_(DENND4A)	C28447S501		8
Protein_XRP2_(rp2)	C2435S173		2
Zinc_finger_CCCH_domain-containing_protein_14_(Zc3h14)	C49448S110		11
Zinc_finger_CCHC_domain-containing_protein_9_(ZCCHC9)	C6250S141		8
Unknown 2	C188S318		3
Galaxin	C20421S239	Lundgren et al. 2013	NA
βγ-crystallin (BGC)	C35180S523		NA
Ubiquitin_like_protein (ubiquitine)	isotig06954		NA
Ligand of numbX2 (LNX2)	C63538S709		9
Thioredoxin 1 (Trx1)	C16728S211		NA
Coatomer	C60613S230		3
Arginine kinase (AK)	C50281S478		7
Hsp60	C52394S280		6
Mn superoxide dismutase (Mn_SOD)	C16774S791		6

*Marker name from http://www.bio.utexas.edu/research/matz_lab/matzlab/Data.html
**Linkage group from Wang et al. 2009

3.2.4 HRM genotyping

Procedures for DNA extraction and HRM genotyping of pooled samples were followed as described in chapter 2. In this chapter, frequencies of the high-melting allele (i.e. one allele at each locus) were calculated for each population and used for statistical analyses.

3.2.5 Statistics

3.2.5.1 Detection of divergent populations

Outlier populations were excluded from gene by environmental association analysis to minimise the likelihood of falsely identifying SNPs that correlate with environmental gradients due to historical factors or limits to gene flow. The R package, rrcovHD (Filzmoser & Todorov 2013) was used to calculate Mahalanobis distance, which

is a multivariate outlier detection analysis that can be used for identifying populations that deviate in allele frequency distribution. The calculation was iterated 1,000 times using standardised allele frequencies of five selected loci from five different linkage groups (C20479S292, C70S236, C60613S230, C16774S791 and C29226S281) and the number of times populations were identified as outliers was counted. Cut-off values for determining outliers were set to 500 (i.e. > 50 % chance of being identified as an outlier).

3.2.5.2 Gene by environment association

Allele frequency data for populations to the exclusion of outliers were used to test correlations with environmental data that were obtained at corresponding locations. Spearman's rank correlations between environmental variables and allele frequencies were calculated using the R package, stats, version 3.0.2. Categorical values were used for environmental data.

3.2.5.3 Environmental correlation network

To investigate correlations among environmental variables, pairwise Pearson correlation coefficients of the 21 environmental variables introduced in section 3.2.2 (see Table 3.2) were calculated using the rcorr function from the R package, Hmisc, version 3.14-5 (http://www.R-project.org). Using the correlation values, environmental correlation networks were created with tools from the igraph package, version 0.6.5-2 for R (Csardi & Nepusz 2006). Environmental variables are represented by vertices that are connected by lines showing interactions among the variables. Negative signs that denote negative correlations were removed to simplify the display of interactions among variables. Correlation values below 0.6 were also removed from the network graph to show the strongest drivers of the interactions (Fig. 3.4). Categorical values were used.

3.2.5.4 Spatial clustering of water quality and temperature factors

A principal component analysis (PCA) plot was constructed based on NO_3, chlorophyll and the range in sea surface temperature (SST) data to explore clustering patterns for both water quality and temperature metrics among 18 genetically connected reefs. The three environmental variables were selected based on their respective roles in coral ecosystem health. Chlorophyll is an indicator of ecosystem status of the GBR (De'ath & Fabricius 2008); NO_3 has adverse effects on coral physiology (Marubini & Davies 1996; Nordemar et al. 2003; Schloder & D'Croz 2004); and temperature as this thesis has shown that stress tolerant alleles are more abundant in thermally unstable environments (see Ch. 4 and 5).The R package, ggbiplot was used to generate a graph. Categorical values were used.

3.2.5.5 Combined effects of NO_3 and temperature range

As allele frequencies of C70S236 and C29226S281 showed strong correlations with NO_3 and range in SST, the combined effects of the two environmental factors were investigated, using multiple linear regression analysis. Environmental factors that showed high collinearity in environmental correlation network analyses were not included in this analysis. Continuous values were used for all the variables. The R package, stats was used to measure the correlations.

3.3 Results

3.3.1 Outlier detection

Of the 19 loci examined, five loci (C20479S292, C29226S281, C70S236, C60613S230 and C16774S791) showed good DNA amplifications and had clear calling of two alleles and were therefore used in further analyses of allele frequencies. Fourteen

loci showed multiallelic genotypes, poor genotyping results or little variation among populations, and therefore were excluded from further analyses.

Seven divergent populations were identified based on Mahalanobis distance measures. Five populations among the divergent populations detected in this study were from offshore reefs in the southern GBR region (Ross Reef, Darley Reef, Boulton Reef, Goble Reef, 21-121 Reef) and two populations were from the southernmost central reefs (High Peak Reef and North Keppel Island Reef) (Figs. 3.2 and3.3).

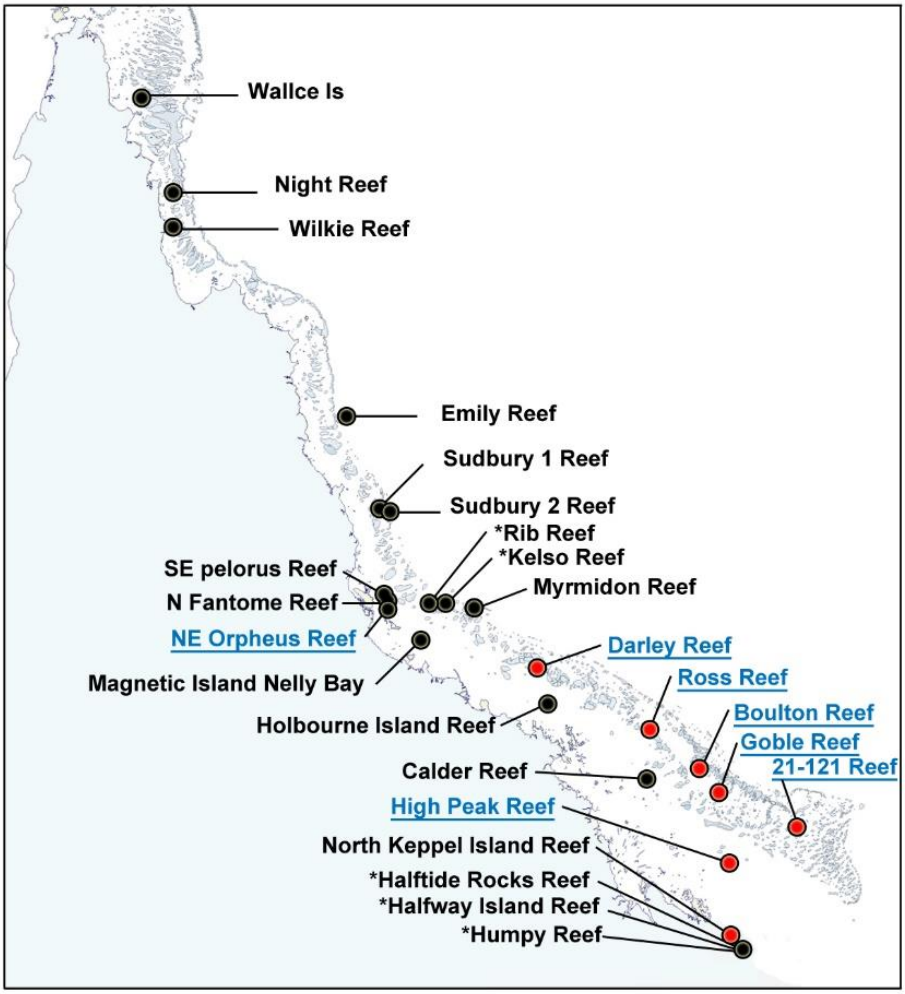

Fig. 3.2 Map showing the locations of genetically connected and divergent populations detected in this study. Black circles: genetically connected populations. Red circles: divergent populations. Blue text and underlining: denote divergent populations detected in van Oppen et al. (2011). Asterisks: denote populations used in this study additional to those used in van Oppen et al. (2011).

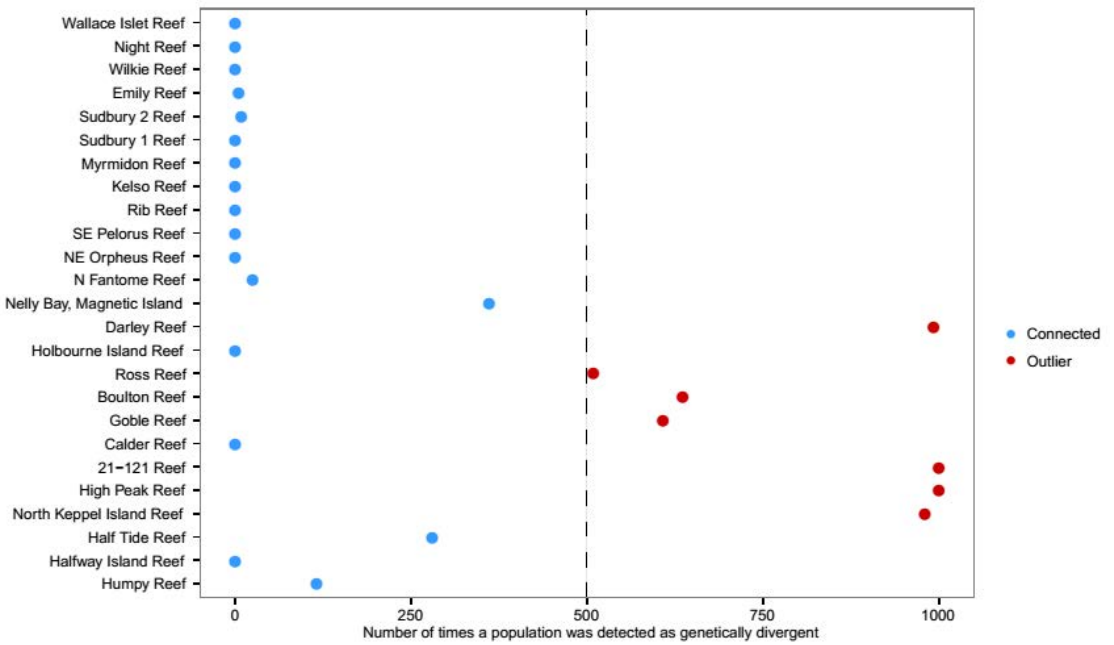

Fig. 3.3Outlier populations detected by 1,000 iterations of Mahalanobis distance calculations. Outlier (red dots) and non-outlier (blue dots) populations, corresponding to values above and below the cut-off value of 500, respectively.

3.3.2 Gene by environment association

Analyses revealed that the loci C70S236 and C29226S281 were significantly correlated with NO_3, DIP, latitude, mean SST and richness of zooxanthellate and azooxanthellate soft coral taxa (see Table 3.2 for values of Spearman's rank correlation tests). Additionally, C70S236 showed significant correlations with minimum temperatures and the range in temperatures (Table 3.2). In the next chapter, I examined putative functional roles of genes linked to the markers, as well as associations between genotypes and phenotypes to ensure that the observed correlations were unlikely to be false positives.

Table 3.2. Spearman's rank correlation coefficients (rho) between environmental gradients and allele frequencies. *P value = 0.05-0.1 **P value < 0.05

Category	Environmental variable	C20479S292	C70S236	C60613S230	C16774S791	C29226S281
Water quality	Shore distance (dis.shore)	-0.152	0.3292	0.1187	-0.0609	-0.1827
	Chlorophyll (chl)	0.1331	**-0.4632***	-0.0788	0.098	**0.4483***
	Suspended sediment (SS)	0.237	**-0.4534***	-0.0022	0.1479	0.1805
	Particulate nitrogen (PN)	0.2166	**-0.4428***	-0.069	0.0269	0.3049
	Particulate phosphorus (PP)	0.2508	-0.397	-0.1537	0.0939	0.031
	Nitrate (NO_3)	0.2594	**-0.634****	**0.4012***	0.0454	**0.6639****
	Nitrite (NO_2)	0.2376	**-0.45***	0.1369	-0.1423	0.0482
	Ammonia (NH_4)	0.1006	0.08	-0.2034	0.1936	**-0.4219***
	Dissolved inorganic phosphorus (DIP)	0.2344	**-0.517****	0.1063	**0.4634***	**0.5225****
	Total dissolved nitrogen (TDN)	0.3695	-0.0181	0.0799	0.2929	-0.2705
	Total dissolved phosphorus (TDP)	0.3404	-0.1532	0.0766	0.2744	0.2564
	Silicate (Sil)	0.2792	-0.0477	-0.1771	0.2202	-0.2179
	Secchi depth (sd)	-0.276	0.1153	0.2483	-0.1509	0.1142
Temperature	Latitude (lat)	0.2447	**-0.4812****	0.2303	0.0145	**0.5266****
	Mean SST (mean)	-0.163	**0.4829****	-0.2286	0.063	**-0.5492****
	MIN_SST (min)	-0.0155	**0.5248****	0.0558	0.0857	-0.3853
	MAX_SST (max)	0.061	0.1219	-0.1921	-0.0052	-0.3564
	RANGE_SST (range)	0.0548	**-0.5289****	-0.1446	0.0382	0.3151
Biotic	Soft Corals: Richness of zooxanthellate taxa (sc.z)	-0.089	**0.5244****	0.0282	-0.0455	**-0.5775****
	Soft Corals: Richness of azooxanthellate taxa (sc.az)	0.017	**0.4705****	-0.2793	0.1051	**-0.6128****
	Species richness of interreefal fish communities (div.)	-0.059	**0.4351***	-0.0989	0.3329	-0.3724

3.3.3 Environmental and biotic correlation network

Two large clusters of SST-related and WQ-related variables were identified based on network analysis of correlation coefficients among the GBR-wide environmental data sets, with strong correlations among individual variables within each cluster (Fig. 3.4). With the exception of range in SST, SST-related variables showed strong correlations with latitude. Both range in SST and fish species richness were correlated to variables within both clusters. Patterns in environmental correlations for richness of zooxanthellate and azooxanthellate soft coral taxa differed; richness of zooxanthellate soft corals was correlated with environmental parameters in the WQ cluster, and richness of azooxanthellate soft corals was correlated with environmental parameters in the SST cluster. The water quality parameter NH_4 did not show a correlation of more than 0.6 to either group (See Table 3.3).

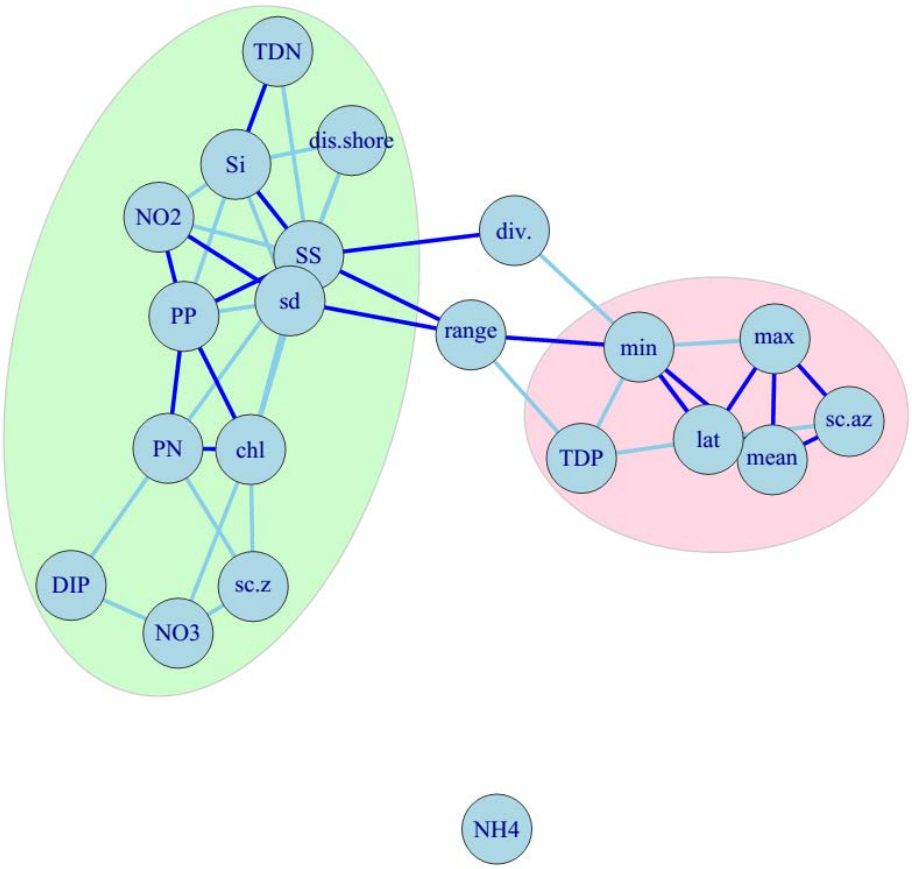

Fig. 3.4 Environmental correlation network for variables used in genotype by environment analyses. Different coloured lines connect variables that are highly correlated (dark blue lines; correlation coefficient>0.7) or moderately correlated (light blue lines; correlation coefficient=0.6-0.7). Coloured ovals highlight that water quality variables (light green oval) and temperature variables (pink oval) tend to group in separate clusters. See Table 3.2 for explanations of the abbreviations.

Table 3.3. Spearman's rank correlation coefficients, testing for pairwise correlations among environmental variables. Correlation coefficients >0.6 are underlined.

	mean	Si	PN	NH_4	NO_2	NO_3	PP	DIP	SS	TDP	TDN	chl	sd	div.	sc.az	sc.z	dis.shore	min	max	range	lat
mean	1	0.04	-0.21	0.39	0.06	-0.37	0.03	-0.44	-0.22	-0.59	0.21	-0.17	0.14	0.46	**0.72**	0.43	-0.22	**0.89**	**0.8**	-0.44	**0.97**
Si	0.04	1	0.34	0.36	**0.62**	0.23	**0.65**	0.15	**0.75**	0.44	**0.72**	0.4	**-0.66**	-0.56	0.29	-0.26	**-0.63**	-0.14	0.47	0.6	-0.07
PN	-0.21	0.34	1	0.44	0.41	0.56	**0.72**	**0.61**	**0.65**	0.32	0.39	**0.74**	-0.44	-0.38	-0.1	**-0.69**	-0.21	-0.38	-0.02	0.47	-0.18
NH_4	0.39	0.36	0.44	1	0.28	0.13	0.51	0.09	0.3	-0.11	0.49	0.33	-0.1	-0.03	0.34	0.15	-0.23	0.31	0.48	0.04	0.34
NO_2	0.06	**0.62**	0.41	0.28	1	0.49	**0.72**	0.38	**0.66**	0.2	0.58	0.56	**-0.73**	-0.4	0.14	-0.39	-0.53	-0.17	0.35	0.54	-0.01
NO_3	-0.37	0.23	0.56	0.13	0.49	1	0.49	**0.65**	0.42	0.3	0.26	**0.62**	-0.34	-0.25	-0.41	**-0.63**	-0.15	-0.46	-0.25	0.37	-0.37
PP	0.03	**0.65**	**0.72**	0.51	**0.72**	0.49	1	0.5	**0.83**	0.27	0.59	**0.76**	**-0.66**	-0.53	0.2	-0.44	-0.58	-0.22	0.36	0.6	-0.03
DIP	-0.44	0.15	**0.61**	0.09	0.38	**0.65**	0.5	1	0.46	0.47	0.25	0.56	-0.35	-0.28	-0.43	-0.53	-0.03	-0.5	-0.28	0.4	-0.46
SS	-0.22	**0.75**	**0.65**	0.3	**0.66**	0.42	**0.83**	0.46	1	0.6	**0.67**	**0.67**	**-0.79**	**-0.72**	0.16	-0.53	**-0.62**	-0.45	0.23	**0.79**	-0.28
TDP	-0.59	0.44	0.32	-0.1	0.2	0.3	0.27	0.47	0.6	1	0.48	0.3	-0.49	-0.55	-0.18	-0.41	-0.23	**-0.65**	-0.22	**0.64**	**-0.65**
TDN	0.21	**0.72**	0.39	0.49	0.58	0.26	0.59	0.25	**0.67**	0.48	1	0.4	-0.59	-0.27	0.37	-0.18	-0.47	0.04	0.55	0.44	0.13
chl	-0.17	0.4	**0.74**	0.33	0.56	**0.62**	**0.76**	0.56	**0.67**	0.3	0.4	1	**-0.65**	-0.44	0	**-0.63**	-0.44	-0.4	0.05	0.55	-0.15
sd	0.14	**-0.66**	-0.44	-0.1	**-0.73**	-0.34	**-0.66**	-0.35	**-0.79**	-0.49	-0.59	**-0.65**	1	0.55	-0.17	0.54	**0.65**	0.42	-0.23	**-0.75**	0.18
div.	0.46	-0.56	-0.38	-0	-0.4	-0.25	-0.53	-0.28	**-0.72**	-0.55	-0.27	-0.44	0.55	1	0.06	0.4	0.37	**0.61**	0.09	**-0.7**	0.52
sc.az	**0.72**	0.29	-0.1	0.34	0.14	-0.41	0.2	-0.43	0.16	-0.18	0.37	0	-0.17	0.06	1	0.27	-0.46	0.55	**0.74**	-0.05	**0.69**
sc.z	0.43	-0.26	**-0.69**	0.15	-0.39	**-0.63**	-0.44	-0.53	-0.53	-0.41	-0.18	**-0.63**	0.54	0.4	0.27	1	0.19	0.6	0.28	-0.52	0.37
dis.shore	-0.22	**-0.63**	-0.21	-0.2	-0.53	-0.15	-0.58	-0.03	**-0.62**	-0.23	-0.47	-0.44	**0.65**	0.37	-0.46	0.19	1	0.08	-0.52	-0.57	-0.15
min	**0.89**	-0.14	-0.38	0.31	-0.17	-0.46	-0.22	-0.5	-0.45	**-0.65**	0.04	-0.4	0.42	**0.61**	0.55	0.6	0.08	1	**0.64**	**-0.72**	**0.89**
max	**0.8**	0.47	-0.02	0.48	0.35	-0.25	0.36	-0.28	0.23	-0.22	0.55	0.05	-0.23	0.09	**0.74**	0.28	-0.52	**0.64**	1	0.07	**0.73**
range	-0.44	0.6	0.47	0.04	0.54	0.37	0.6	0.4	**0.79**	**0.64**	0.44	0.55	**-0.75**	**-0.7**	-0.05	-0.52	-0.57	**-0.72**	0.07	1	-0.49
lat	**0.97**	-0.07	-0.18	0.34	-0.01	-0.37	-0.03	-0.46	-0.28	**-0.65**	0.13	-0.15	0.18	0.52	**0.69**	0.37	-0.15	**0.89**	**0.73**	-0.49	1

3.3.4 Spatial clustering of water quality and temperature factors

A PCA revealed that most reefs grouped in two clusters, based on two water quality variables (Chlorophyll and NO_3) and one temperature variable (range in SST).The pattern of clustering indicates that environmental conditions at Halftide Rocks, Humpy, Halfway and Magnetic Reefs are different from those at the other locations (Fig. 3.5). Principal component 1 explained 76.2% variation of the three variables.

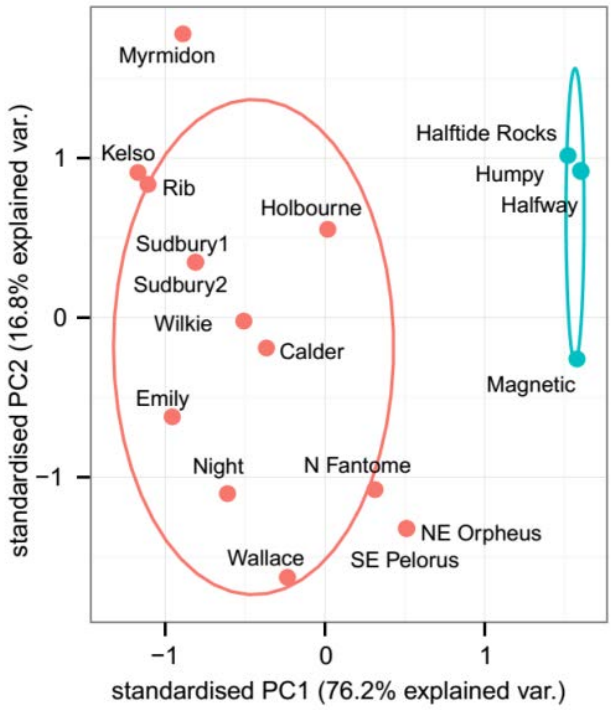

Fig. 3.5 Principal component analysis reveals two clusters of reefs based on NO_3, chlorophyll and SST range data for 18 reefs. The four populations in the smaller cluster of blue dots are characterised by high NO_3 and chlorophyll concentrations, and high annual temperature fluctuations.

3.3.5 Combined effects of NO_3 and temperature range

The multiple regression model that accounted for both the NO_3 and SST range variables explained 48% and 58% of the variability in allele frequency patterns for the C29226S281 and C70S236loci, respectively, with more than a 99% level of confidence (Table 3.4). This was despite the fact that the model predicted that the independent

effect of range in SST on allele frequencies was non-significant and small for both markers.

Table 3.4. Summary of multiple linear regression model for testing correlations between allele frequencies and two environmental variables, SST range and NO_3.

Marker		Value	Std.Error	t-value	Pr(>|t|)
C70S236	(Intercept)	0.5742	0.0663	8.6673	0
	SST range	-0.0122	0.0083	-1.4753	0.1608
	NO_3	-0.1021	0.0543	-1.8783	0.0799
C29226S281	(Intercept)	0.8238	0.0864	9.536	0
	SST range	-0.012	0.0108	-1.1073	0.2856
	NO_3	0.3007	0.0709	4.243	0.0007

Marker	R-squared	Adjusted R-squared	p-value
C70S236	0.4792	0.4097	0.007501
C29226S281	0.5835	0.528	0.001402

3.4 Discussion

3.4.1 Gene by environment association

The frequencies of two SNP loci (C29226S281 and C70S236) were found to be correlated with environmental gradients related to temperature, water quality and biotic variables, based on analyses of 18 populations of the common coral *Acropora millepora* along the length of the GBR. Genetic loci under environmental selection may be quantitative trait loci (QTL) contributing to traits associated with tolerance to environmental stressors in corals. Patterns in SNP-environmental correlations are consistent with three alternative explanations of gene by environment associations. Observed allele frequency patterns may be explained by either: 1) latitude-associated environmental factors (i.e. mean, minimum and maximum SST variables); 2) water quality factors (i.e. NO_3 and DIP); or 3) the combined effects of both types of variables; or 4) false positives.

Significant correlations between allele frequencies and multiple variables within temperature and water quality categories can be attributed to high correlations among the environmental variables, as shown by the environmental correlation network (Fig. 3.4). In addition to temperature-related variables, latitudinal gradients are often associated with other environmental variables, such as length of season, cyclone frequency, rainfall and UVR (see http://www.bom.gov.au/climate/averages/maps.shtml). Therefore, correlations with latitude-associated gradients implicitly suggest a potential involvement of other environmental factors in shaping allele frequency patterns. The parallel changes in these environmental variables along latitudes may have additive or synergistic effects on coral fitness.

A variety of factors contributing to poor water quality, such as enhanced levels of dissolved inorganic nutrients (DIN) and particulate matter, and reduced water clarity, are recognised as important factors influencing coral fitness. For example, high levels of DIN reduce coral calcification rates by 50% (Fabricius 2005). DIN is preferentially taken up by zooxanthellae, leading to an increase in *Symbiodinium* density at the expense of growth of host tissues (Dubinsky & Jokiel 1994). Increased algal populations result in reduction of CO_2 available for coral calcification (Marubini & Atkinson 1999; Marubini & Thake 1999). Furthermore, increased amounts of organic and inorganic particles associated with river run-off can severely reduce light availability. In addition to shading by particulate matter, phytoplankton production enhanced by increased nutrient availability reduces light penetration (Abal& Dennison 1996). Reduced light availability for *Symbiodinium* photosynthesis has negative effects on carbon gain, calcification rates and tissue biomass (Rogers 1979; Telesnicki & Goldberg 1995; Anthony & Hoegh-Guldberg 2003).Sedimentation can also smother corals within a few days (Riegl & Branch 1995; Golbuu et al. 2003). Changes in water quality can also affect abundance of other benthic organisms and induce secondary effects on health and abundance of

corals. When combined with polysaccharides frequently called “marine snow”, possibly produced by bacteria and diatoms, low-level sedimentation may kill coral recruits (Fabricius et al. 2003).Moreover, crustose coralline algae (CCA), important for coral settlement, are negatively affected by sedimentation (Kendrick 1991; Fabricius & De’ath 2001b). Internal bioeroders, such as sponges, polychaetes, ascidians, molluscs and bryozoans that live within coral skeletons bore into or chemically erode the calcium carbonate skeleton. The abundance of bioeroders increases with enhanced nutrient levels (Holmes 2000; Chazottes et al. 2002), negatively affecting the structure of coral reefs.

The significant effects of water quality on species diversity of zooxanthellate soft corals detected in this study indicates that nutrient variation may be an important ecological driver for organisms that harbour symbiotic algae. Previous studies showing that exposure of corals to high NO_3 concentrations results in increases in both chlorophyll concentrations per cell and density of zooxanthellae (Marubini & Davies 1996) provide further support for my conclusion that nutrients affect the richness of zooxanthellate soft corals (also see Fabricius and De’ath 2001). Several studies found exacerbating effects of NO_3 on coral fitness when combined with high temperatures (Nordemar et al. 2003; Schloder & Croz 2004; Fabricius et al. 2013). Excess algal symbionts can render corals susceptible to bleaching. Cunning & Baker (2013) demonstrated that colonies with high densities of *Symbiodinium* were more severely bleached, possibly as a consequence of higher oxidative stress in corals with more algal symbionts.

It is evident that environmental variables related to temperature and water quality have cumulative effects on coral fitness (Nordemar et al. 2003; Schloder & Croz 2004; Wooldridge 2009; Fabricius et al. 2013). However, it should be noted that mean

SST and NO_3 at sampling sites in this study have an inverse relationship (i.e., the locations of high mean SST have low NO_3 concentrations and vice versa; see Table 3.3), suggesting that the populations investigated may not suffer high mean SSTs and high NO_3 concentrations simultaneously.

In contrast, range in SST is positively correlated with NO_3 concentrations across sampling locations, indicating locations with poor water quality (i.e. high nutrient concentrations) are characterised by wider temperature fluctuations. In fact, four populations (Magnetic Island, Humpy, Halftide Rocks and Halfway Island Reefs) that showed allelic differentiation from the remainder of the populations at the two loci are characterised by high temperature fluctuation and poor water quality, as represented by high chlorophyll and NO_3 concentrations (see Figs. 3.5 and 3.6). Furthermore, a multiple regression model that included NO_3 and range in SST showed a significant fit to allele frequency data for both loci (C29226S281 and C70S236). This increase in the goodness of fit of the model, despite a non-significant and small effect of range in SST when analysed independently (Table 3.4), may be partly attributed to the clustering pattern of the two variables, indicating that there may be synergistic effects between NO_3 and range in SST on allele frequencies of the loci.

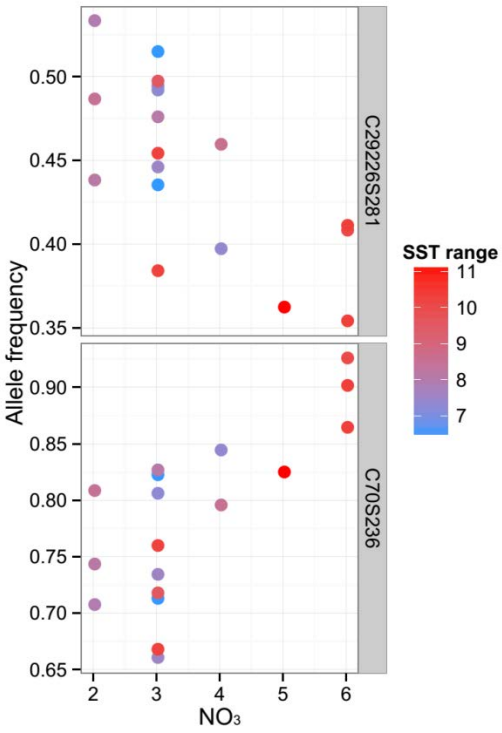

Fig. 3.6 Relationships between frequencies of low-melting alleles at the two loci (C29226S281 and C70S236) and both NO_3 and SST range (°C) in 18 populations of *Acropora millepora* along the Great Barrier Reef. Colour gradient from blue to red indicates an increase in SST range. High NO_3 categorical value indicates high NO_3 concentrations. Categorical values for NO3 were used for a graphical representation.

3.4.2 Detection of outlier populations

Divergent populations detected by Mahalanobis distance of allele frequencies at five SNP loci correspond well to pairwise Fst values and the STRUCTURE analysis of microsatellite divergence by van Oppen et al. (2011). Allele frequencies at the multiple unlinked loci of divergent populations, which together deviate from the average, potentially indicate genome-wide changes within the divergent populations. Potential past allopatric divergence between the Queensland and Marion plateaus of the GBR, as identified by microsatellite markers (van Oppen et al 2011), was apparent in the loci selected for my study, which separated Darley Reef, Boulton Reef, Goble Reef, Ross Reef and 21-121 Reef from all other populations. Significant genetic differentiation of North Keppel Island Reef and High Peak Reef populations from the main group were also

observed in both the previous and current studies. This genetic distinctiveness of the two populations is probably due to major disturbance events in 1991 (Van Woesik et al. 1995), 1998, 2002 (Berkelmans et al. 2004; R. Berkelmans unpublished data), and 2006 (Jones et al. 2008; Diaz-Pulido et al. 2009). However, the three populations added to this study, which are ~13 km south of North Keppel Is Reef, were not detected as outliers despite their close geographical proximity. The 2.2-fold and 1.6-fold higher species richness at Humpy Reef and Halfway Reef, respectively, in comparison to North Keppel Is Reef (Jones & Berkelmans 2010a), may be partly due to mixed depth profiles that may promote habitat diversity and thus resilience (Jones & Berkelmans 2010b).

I detected no significant genetic differentiation of the NE Orpheus population from the main group; a result that contradicts a previous study of these populations (van Oppen et al. 2011). However, after exclusion of microsatellites that contained null alleles from their analysis, divergence levels in that study were no longer as strong. Another population genetics study using a slightly different set of microsatellite markers found significant but small genetic differentiation between Fantome Island population and both NE Orpheus and SE Pelorus populations, all of which are located within the Palm Is group (Fst = 0.01) (Souter et al. 2010). This small scale level of differentiation was not detectable with the markers that I used, as the outlier detection methods used in this study provide information on large-scale patterns of substantial genomic differentiation, hence NE Orpheus is considered panmictic with the other populations retained and was not excluded.

3.5 Conclusion

In this chapter, I have provided evidence of a correlation between two SNP markers (C29226S281 and C70S236) within populations of the coral *Acropora millepora* and both water quality and temperature gradients along the Great Barrier Reef.

Associations between multiple loci in different linkage groups and the same environmental variables suggest that the environmental factors may have a multitude of effects on genetic or physiological functions underlain by these genes or linked genes. The two loci may represent putative QTLs for phenotypic responses to the environmental stressors. In Chapter 4, genotypic differences in bleaching-associated phenotypes and functions of linked genes will be investigated. In Chapter 5, spatial patterns of allele frequencies at these loci will be extrapolated based on water quality and temperature data.

Chapter 4.0 Genotype-phenotype association: quantitative trait loci for coral antioxidant capacity

This chapter and Chapter 3 were combined and submitted as a paper to Science Advances.

Young K. Jin, Petra B. Lundgren, Adrian Lutz, Jean-Baptiste Raina, Emily J. Howells, Allison S. Paley, Bette L. Wllis and Madeleine J.H. van Oppen.

YKJ, PL, BLW and MJHvO conceived the study. AL and J-BR designed, conducted and collected data from a heat stress experiment. EJH and ASP collected samples for a natural bleaching experiment. MJHvO collected larval samples for Mendelian inheritance experiment. YKJ genotyped all the samples and analysed data from all the experiments. YKJ wrote the chapter with intellectual contributions from all co-authors.

Abstract

Coral reefs are declining at an alarming rate because of increasing impacts of global climate change and local anthropogenic disturbances on coral reefs worldwide, however the genetic basis of environmental stress tolerance in reef-building corals is still poorly understood. Transcriptomic studies have revealed genes that are regulated in response to stress, but the allelic or epigenetic basis of phenotypic variation in stress resistance of corals has not been characterized. Here I identify two unlinked single nucleotide polymorphisms (SNPs) as true quantitative trait loci for anti-oxidant capacity in the common coral, *Acropora millepora*. Field surveys revealed that colonies carrying G and T alleles, respectively, at the two loci were significantly less affected by two bleaching events. These results were corroborated by an aquarium-based experiment showing constitutively higher anti-oxidant capacity of colonies with these alleles, particularly in the homozygous state. These genetic markers are located on different chromosomes and closely linked to a range of genes involved in bleaching-associated processes, such as thermal, hypoxic and oxidative stress, suggesting they are indicative of tolerance levels to a multitude of environmental perturbations. These findings open up opportunities for novel coral reef management approaches, such as spatial mapping of stress tolerance for use in Marine Protected Area design, the identification of stress tolerant colonies for assisted migration, and marker-assisted selective breeding to create more tolerant genotypes for restoration of denuded reefs.

4.1 Introduction

Field and laboratory studies of the breakdown in coral-algal symbiosis leading to bleaching (i.e. whitening of coral tissues caused by loss of the endosymbiont *Symbiodinium* from gastrodermal cells of the coral host) have identified a causal link

between bleaching and a variety of environmental stressors (reviewed in Lesser 2011). While thermal stress is often a major cause of bleaching, other stressors also disrupt the symbiosis leading to bleaching. In addition, thermal stress may act synergistically with other stressors by reducing the tolerance of the coral host so that it is more susceptible to environmental stressors such as reduced salinity, and increased sedimentation, light and nutrients (Wooldridge 2009; Lesser 2011). Large-scale, long-term monitoring data for these environmental factors can help to untangle variation in the bleaching response within populations at regional scales (Wooldridge 2009). However, phenotypic diversity may also be present within a population due to non-environmental factors, such as symbiont type and host genotype, making it difficult to separate the role of phenotypic plasticity in the coral response to environmental stressors and population-level variation in host or symbiont genotype (Berkelmans & van Oppen 2006; Bay & Palumbi 2014).

Although investigations into functional links between host genome and phenotypic traits in corals is still in its infancy, the impact of genotype in phenotypic variation is implicated in some studies. A recent transcriptomic study revealed 60 genes that were expressed differentially between thermally susceptible and tolerant corals (Barshis et al. 2013). Reciprocal transplant experiments as well as thermal stress experiment on symbiont-free larvae demonstrated that the genotype of the coral host influences their stress response (Smith et al. 2007; Barshis et al. 2010; Polato et al. 2010). These studies did not identify any effect of symbiont type on the phenotypic variation studied, suggesting that a host-specific mechanism contributes to the resilience of the entire holobiont.

Physiology of symbiotic algae and fitness of the coral holobiont are closely linked because of an indispensable nutritional relationship between the two symbiotic partners. This is further confirmed by reductions in coral reproduction and growth, increased

susceptibility to disease, and high mortality of the host when the obligate symbiosis is disrupted (Harvell et al. 2001; Baird & Marshall 2002). Furthermore, genomic data from *A. digitifera* suggests that there is cooperative metabolic adaptation between host and symbionts (Shinzato et al. 2011; Dunlap et al. 2013). Disruption of this intimate nutritional symbiosis precedes bleaching, thus physiological responses of *Symbiodinium* can be measured as proxies for responses of the coral holobiont to environmental stressors. Typically, performance of *Symbiodinium* is evaluated by measuring chlorophyll concentration per cell and maximum quantum yield as a measure of photochemical efficiency.

Coral bleaching is understood to occur when the coral host's capacity to detoxify reactive oxygen species (ROS) is overwhelmed (Lesser 2011). ROS leak from compromised algal symbionts into the coral host cytosol, and additionally, are produced by host mitochondria (Lesser 2011). As one of the defence systems against ROS, mitochondria contain high levels of the antioxidant, ubiquinol (reduced CoQ) (Turunen & Dallner 2004). Within the coral host, genotypic differentiation in the coenzyme Q (CoQ) pool redox state (%$CoQH_2$; the proportion of reduced to oxidized coenzyme Q) can be measured as an indicator of the anti-oxidant capacity of the coral animal (Lutz et al. 2014).

In this chapter, my overall goal was to determine if the two SNP loci identified in Chapter 3 are true QTLs. My specific objectives were to ascertain if both markers are inherited by standard Mendelian models, and to validate gene by environment associations found in Chapter 3 by examining coral genotype by phenotype associations following bleaching events in natural field studies and in a laboratory heat stress experiment. Because phenotypic traits are not necessarily under the control of genes that show genotypic associations, but can instead be regulated by genes in the same

linkage group (Tishkoff et al. 2007), I also examined functional annotations in the genetic linkage map of *Acropora millepora* (Wang et al. 2009).

4.2 Materials and Methods

4.2.1 Mendelian inheritance

To ensure the target SNP loci (C29226S281 and C70S236) follow standard Mendelian inheritance, gametes from two colonies of *Acropora millepora* from Orpheus Island were crossed *in vitro,* as described in van Oppen et al. 2014. Larvae were reared and fixed in absolute EtOH. DNA extraction was done following methods described in Gloor et al. 1993 and Wilson et al. 2002 for larval and adult coral DNA, respectively. If each locus follows a Mendelian pattern of inheritance, the ratio of offspring with particular genotypes can be predicted according to genotypes of the parents (Table 4.1). This will also verify the presence of prezygotic selection on the loci and whether primers target true biallelic and specific loci.

4.2.2 Natural bleaching experiment

To determine if genotypes at the two marker gene loci corresponded to differences in coral bleaching responses to high temperature and to combined high temperature and low water quality *in situ*, I analysed the relationship between genotype and bleaching history for 364 colonies of *A. millepora* with a known environmental history (Fig. 4.1). Following a summer thermal stress event, 30 coral colonies were randomly sampled from populations of *A. millepora* from each of Cattle Bay, Northeast Orpheus, Southwest Pelorus, Southeast Pelorus and Pioneer Bay on 5 and 6th March 2006 by A. Paley (Fig. 4.2). Following a low salinity event, a variable number of samples of *A. millepora* were randomly collected from Central Pioneer Bay (N = 48), Southern Pioneer Bay (N = 29), Southwest Pelorus (N = 41), Cattle Bay (N = 47) and Fantome Island

(N = 19) on 27, 28, 29th March and 12th April 2009 by E. Howells (Fig. 4.2). Coral nubbins were snap-frozen and stored in liquid nitrogen or directly fixed in 100% ethanol for downstream DNA extractions. In both studies, bleaching states were visually scored in the field (i.e. healthy, pale, and bleached). In genetic analyses described below, healthy and pale conditions were treated as non-bleached.

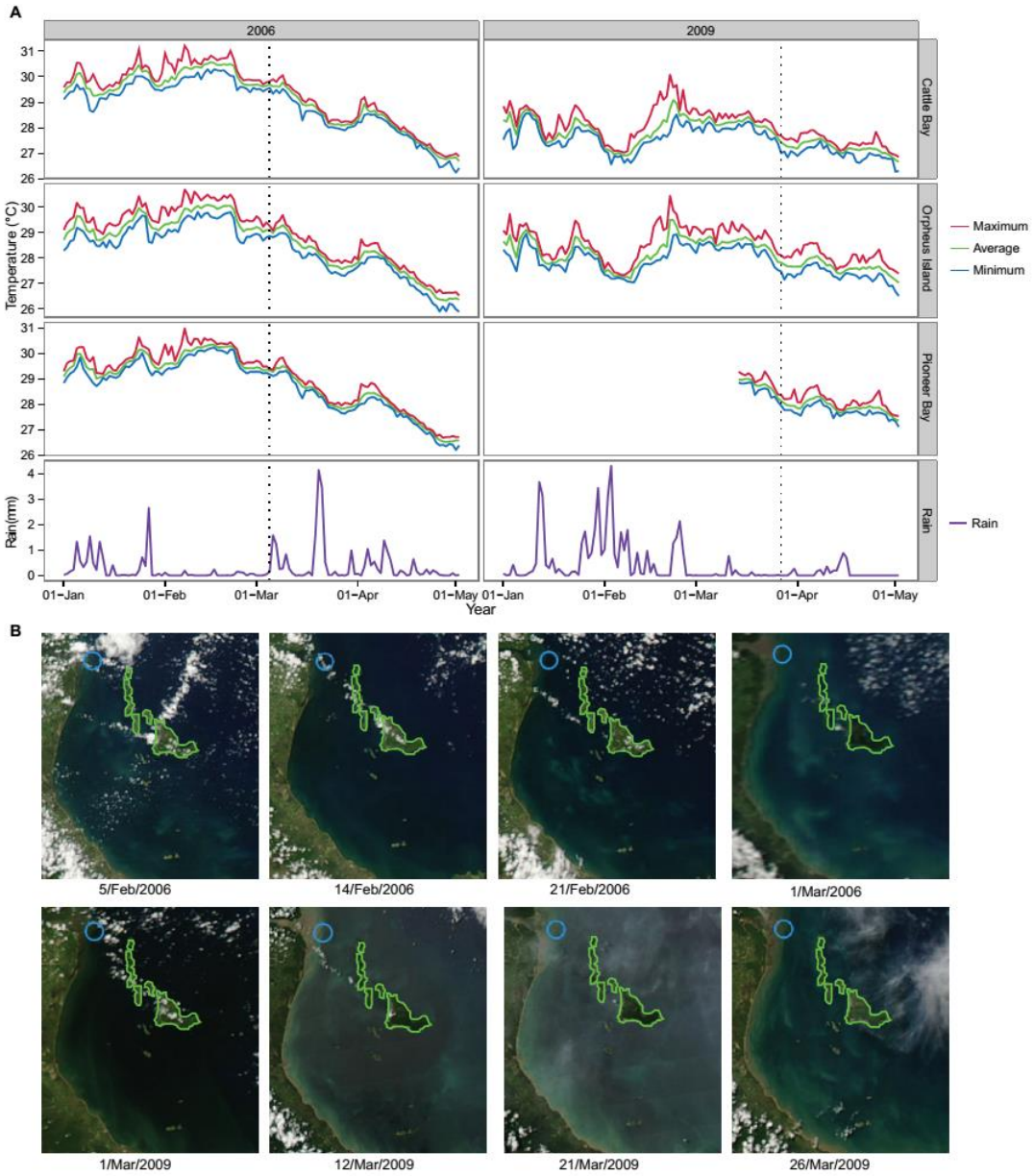

Fig. 4.1 Environmental conditions at sampling sites in the Palm Island Group preceding a thermal stress event in 2006 and a low salinity event in 2009. A) Seawater temperature and rainfall at Orpheus Island. Water temperature data were obtained from data loggers deployed at Cattle Bay and Pioneer Bay, Orpheus Island, and rainfall data were sourced from a weather station at Orpheus Island; both are managed by the Australian Institute of Marine Science. Vertical dotted lines = sampling time point in each study (5th March 2006 and 27th March 2009). **B)** Satellite images showing flood plumes near the Palm Islands in 2006 (top panel) and 2009 (bottom panel). True colour MODIS data 24 to 30 days before observed bleaching were collected from the Aqua satellite (available at https://earthdata.nasa.gov/). Blue circle: indicates mouth of Hinchinbrook channel. The Palm Island group is outlined by green lines.

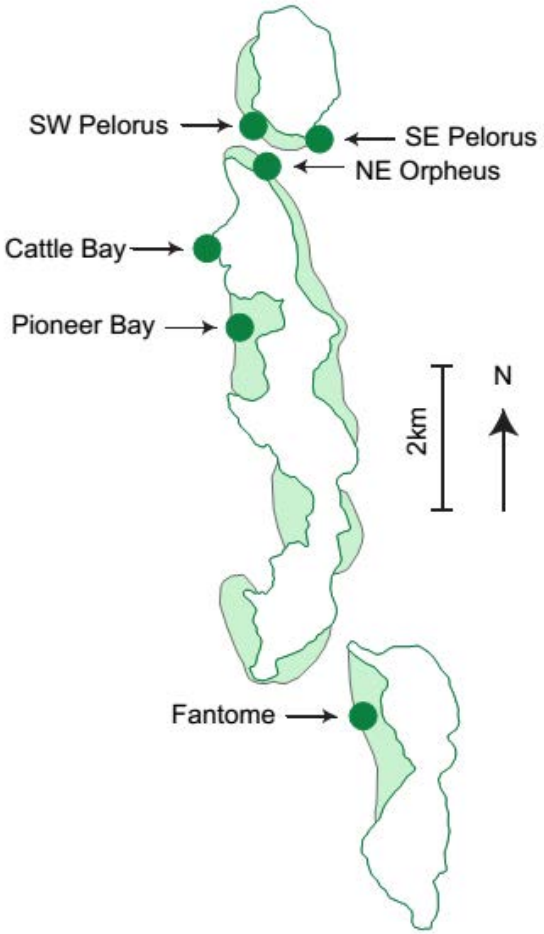

Fig. 4.2 Map showing sampling sites in the Palm Island Group for natural bleaching experiment

4.2.3 Heat stress experiment

To further test the relationship between genotype and thermal bleaching at the two SNP loci, samples of *A. millepora* from a heat stress experiment conducted by Raina et al. (2013) were analysed. Ten colonies of *A.millepora* containing type C2 *Symbiodinium* were collected from a single site at Pelorus Island and transferred to aquaria at the Australian Institute of marine Science (AIMS) by JB Raina and A. Lutz. Colonies were split into a total of 24 fragments, each comprising approximately 25 branches (nubbins). Fragments were placed in 8 indoor tanks in a randomised arrangement. Following 2 weeks of acclimatisation, temperatures in four tanks were ramped up to 32°C over a period of seven days, while the other four control tanks were

maintained at 27°C throughout the experiment. Each tank contained 3 fragments, resulting in 12 fragments in each treatment (Supplementary Table S.4). Corals were sampled at 4 time points: after the acclimation period (t=0), after reaching 32°C in the thermal stress treatment (t=7), and after 5 days (t=12) and 10 days (t=17) at the target temperature. Whole fragments were snap-frozen in liquid nitrogen immediately upon collection. More detailed information on the experimental design is found in Raina et al. 2013. For my genotype-phenotype association studies, I used coral physiological data indicating the oxidative stress state of the coral host, including CoQ pool redox states, which was quantified by Lutz et al. (2014) using LC-MS quantification. I also examined genotypic associations with the chlorophyll a concentrations per zooxanthellae (picogram/ cell) and the maximum quantum yield of photosystem II (Fv/Fm), a measure of photochemical efficiency of *Symbiodinium,* which were quantified using a UV-2550 spectrophotometer and a pulse amplitude modulated fluorimeter (PAM), respectively by Lutz et al. (2014).

4.2.4 SNP genotyping

The 180 samples of *A. millepora* collected following the natural thermal stress event, and 184 samples collected following the natural low salinity stress event were individually genotyped, and 25 samples from the laboratory heat stress experiment were pooled and genotyped at the two target SNP loci (C70S236, C29226S281). Procedures for DNA extraction and HRM genotyping were followed as described in chapter 2 to individually genotype samples. In this chapter, frequencies of the high-melting allele were calculated for each population and used for statistical analyses.

4.2.5 Statistics

One-way ANOVAs and generalised estimating equation model were used to compare the frequency of T alleles at locus C70S236 and G alleles at locus C29226S281 in samples from each of the natural field experiments and the laboratory heat stress experiment. For each SNP locus, genotypic associations with coral tissue concentrations of $CoQH_2$, chlorophyll a concentrations and photochemical efficiency of associated *Symbiodinium (*Fv/Fm values) were tested by a generalised estimating equation model. For each colony, $CoQH_2$, chlorophyll and Fv/Fm data were partitioned into the same clusters in the model. I used an exchangeable correlation structure as multiple fragments from the same colonies were used for the same treatments (i.e. intra-colony correlations) (Hin & Wang 2009). This procedure takes into account dependence between measurements on the same colonies. Wald tests were used to examine the effect of each term added sequentially to the model. The model was fitted using the R package *geepack* (Halekoh et al. 2006).

4.2.6 Functions of genes in linkage groups

Wang et al. (2009) show that 44 and 34 functional loci are in linkage with C29226S281 and C70S236 gene loci, respectively. A BLAST-X search was performed on the linked genes as well as target loci using ZoophyteBase (Dunlap et al. 2013). Markers that did not return any annotations, annotations with an E-value > 0.001 and a bit score < 40, and microsatellite markers were excluded from gene ontology mapping analysis. As a consequence, the KEGG-based functional classification was conducted with a total of 30 and 25 markers for C29226S281 and C70S236 linkage groups, respectively.

4.3 Results

4.3.1 Mendelian inheritance

Both C29226S281 and C70S236 markers showed strict Mendelian segregation patterns in an analysis involving two parent colonies of *A. millepora* (heterozygote female x homozygote male for both loci) and 50 of their offspring (0.48:0.5 = AG: GG (C29226S281); 0.54:0.44 = TC:CC (C70S236)) (Table 4.1). Chi-squared test shows that expected genotype frequencies are not significantly different from observed frequencies for both markers (p-values > 0.05).

Table 4.1. Mendelian inheritance of C29226S281 and C70S236 markers. Genotypes of female and male parents were AG and GG at C29226S281 and TC and CC at C70S236, respectively.

Marker	Genotype	Expected frequency	Observed frequency	X^2	P-value
C29226S281	AA	0%	2%	0.0204	0.9898
	AG	50%	48%		
	GG	50%	50%		
C70S236	TT	0%	2%	0.0254	0.9874
	TC	50%	54%		
	CC	50%	44%		

4.3.2 Natural bleaching experiment

In the summer of 2009, a short period of high temperature coincided with extensive rainfall, causing increased turbidity and decreased salinity. Of 184 colonies genotyped from five sites in the Palm Islands, non-bleached colonies (n=162) showed ~28% higher frequency of allele G at SNP locus C29226S281 compared to bleached colonies (n=22) (F=8.481, df=1, p=0.0226; Fig. 4.3A). Frequencies of T allele at C70S236 were 74% and 68% for bleached and non-bleached colonies, respectively (F=0.905, p=0.369). Genetic analyses of 150 colonies of *A. millepora* sampled across five sites in the Palm Islands (central GBR) during a temperature-induced bleaching event in the summer of 2006, revealed ~12% higher frequency of allele T at C70S236 in non-bleached (n=119) compared to bleached colonies (n=31) (F=5.799, df=1, p=0.0469; Fig.

4.3A). Frequencies of G allele at C29226S281 were 74% and 63% for bleached and non-bleached colonies, respectively (F=1.077, p=0.334).

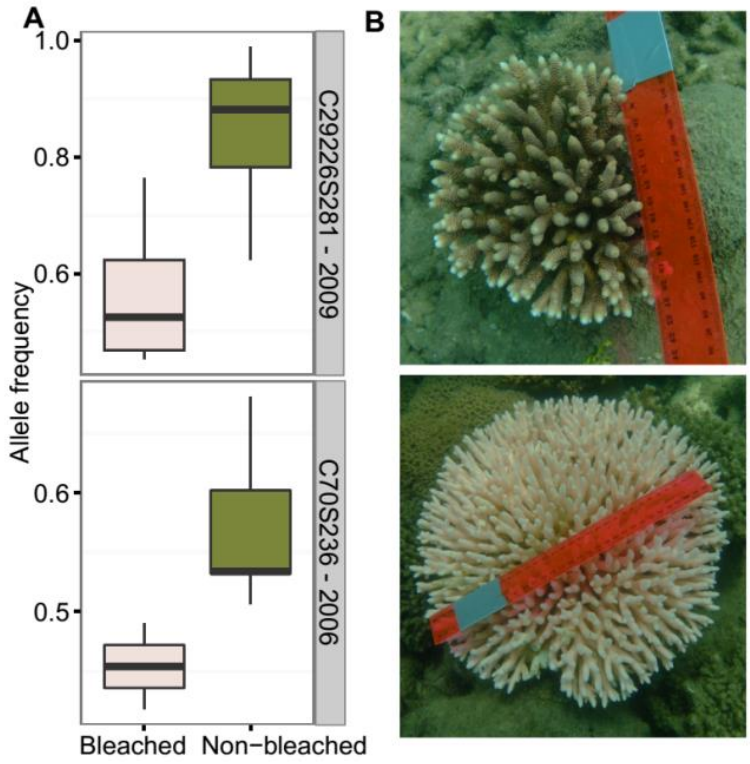

Fig. 4.3 Genotype-phenotype associations. A) The relationships between bleaching responses observed in 2006 and 2009 (N = 180 and 184 coral colonies, respectively) and allele frequencies at C70S236 (p value = 0.0469) and C29226S281 (p value = 0.0226) markers, respectively. Frequencies of G and T are shown for C29226S281 and C70S236 markers, respectively. **B)** Photos showing non-bleached (top) and bleached (bottom) corals.

4.3.3 Heat stress experiment

Colonies of the GG and TT genotypes for C29226S281 and C70S236, respectively, maintained about 3-6% higher $CoQH_2$ levels in both the 27°C and 32°C treatments compared to other genotypes (C29226S281: $p_{27°C}$=0.0096, $p_{32°C}$=0.0013; C70S236: $p_{27°C}$ =0.0014, $p_{32°C}$<0.0001; generalised estimating equation model) (Fig. 4.4). One fragment did not survive 32°C treatment.

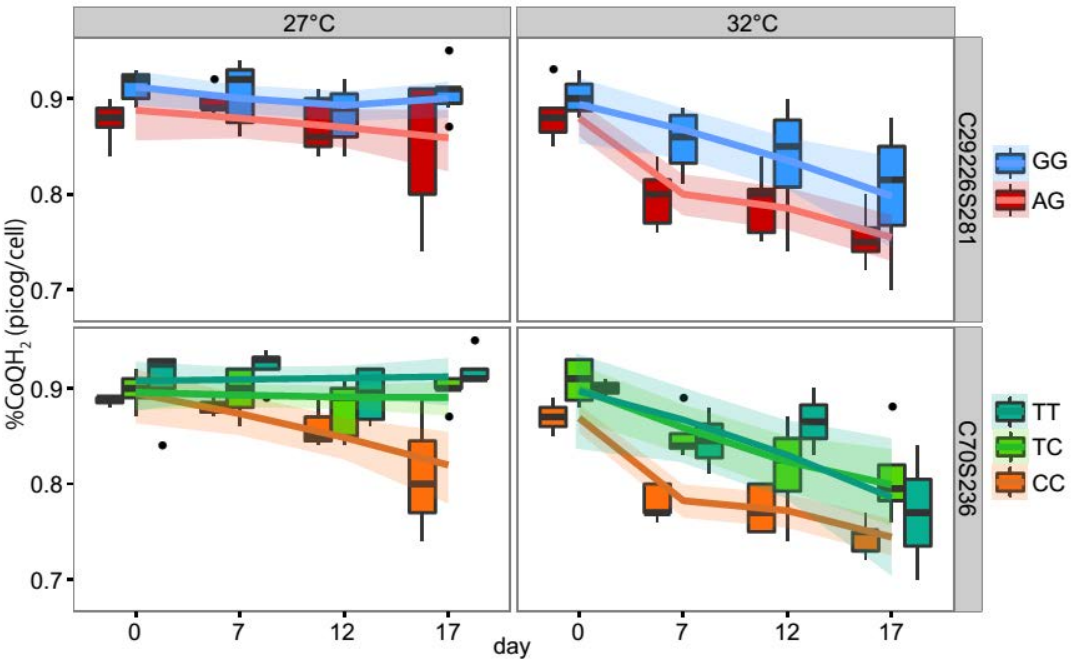

Fig. 4.4 The relationship between %$CoQH_2$ in tissues of *Acropora millepora* following a heat stress experiment and genotype for SNP loci C29226S281 (top panel) and C70S236 (bottom panel). N = 8 coral colonies per temperature treatment. %$CoQH_2$ differed significantly between alleles at both loci in both temperature treatments (C29226S281: $p_{27°C}$ =0.0096, $p_{32°C}$ =0.0013; C70S236: $p_{27°C}$ =0.0014, $p_{32°C}$<0.0001). Shading indicates 95% confidence intervals.

Consistent with a higher percentage $CoQH_2$, I found stable chlorophyll concentrations per *Symbiodinium* cell in GG and TT genotypes of C29226S281 and C70S236, respectively (one-way ANOVA: p=>0.05; Fig. 4.5A), and about 2 and 20% higher Fv/Fm ratios in the C70S236 TT compared to the CC genotype in both the 27°C and 32°C treatments, respectively, over the course of the experiment (generalised estimating equation model: $p_{27°C}$=0.0001; $p_{32°C}$=0.0007; Fig. 4.5B). Two coral fragments did not survive 32°C treatments for both measurements.

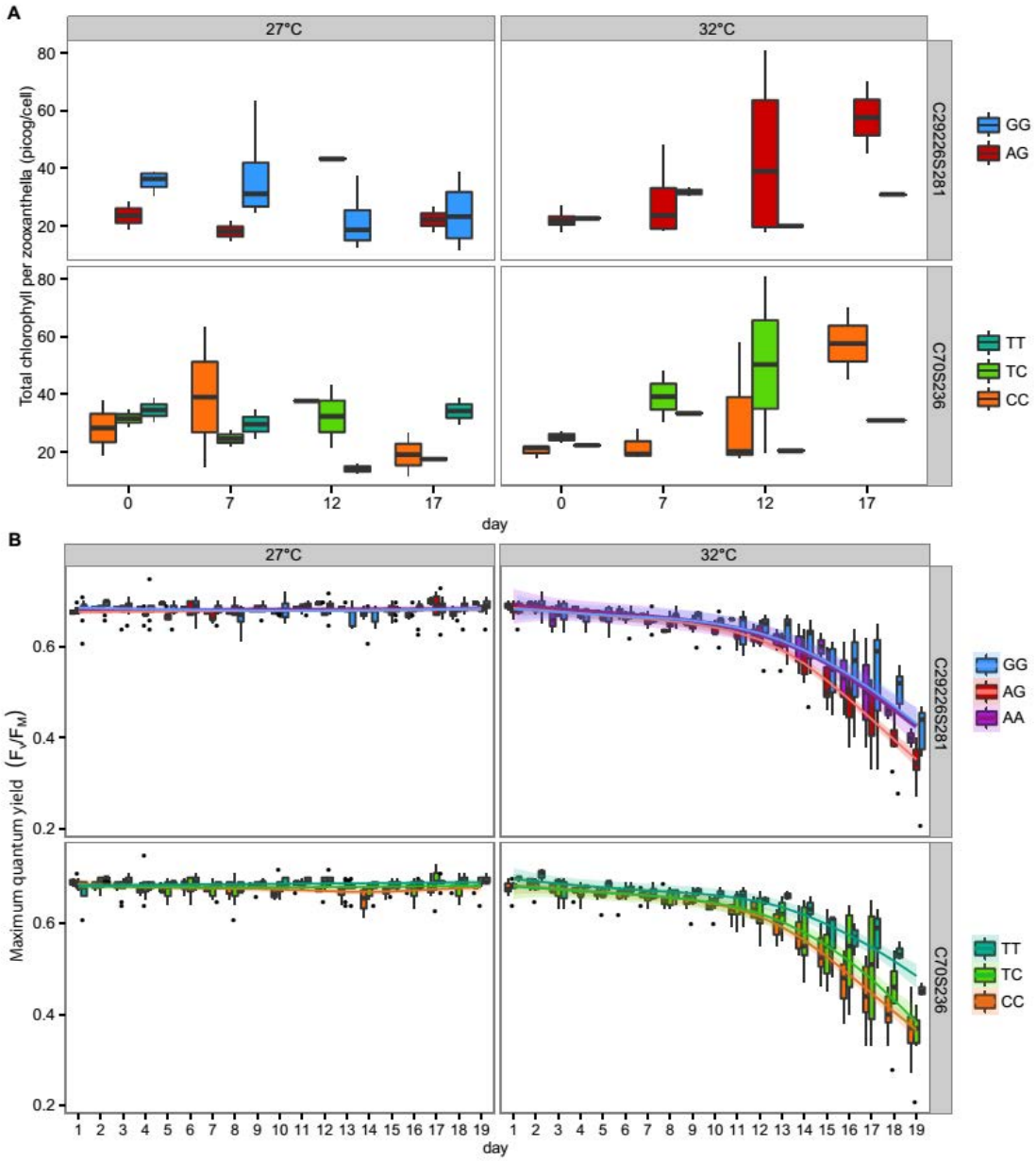

Fig. 4.5 Genotype-phenotype association. A) The relationships between total chlorophyll concentration per *Symbiodinium* cell and coral host genotype at C29226S281 and C70S236 markers in the heat stress experiment (N = 5 samples per temperature treatment). **B)** The relationships between *Symbiodinium* Fv/Fm and coral host genotype at C29226S281 and C70S236 markers in the heat stress experiment (N = 10 samples per temperature treatment). Fv/Fm ratio were significantly higher for the TT genotype at C70S236 in both temperature treatments ($p_{27°C}$=0.0001; $p_{32°C}$=0.0007).

4.3.4 Linkage study

The linkage map constructed by Wang et al. (2009) shows that the C29226S281 and C70S236 markers are found in different linkage groups and are both physically close

(<10 centimorgans) to genes predominantly involved in genetic information processing (Fig. 4.6).

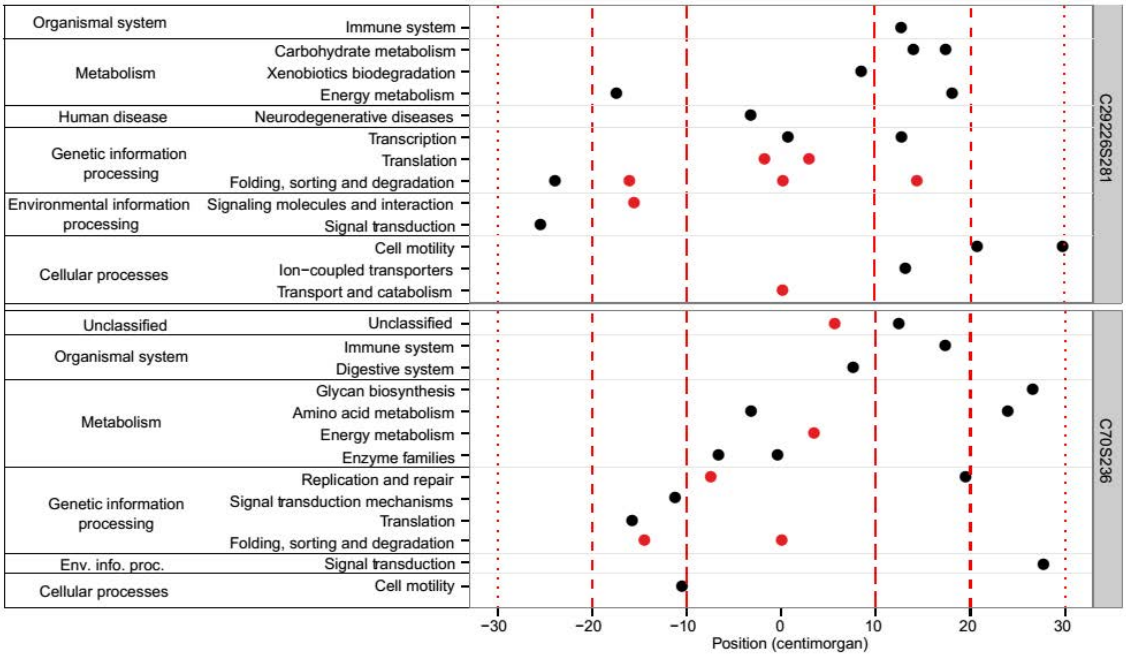

Fig. 4.6 Linakge maps of C29226S281 and C70S236 markers showing the genetic distance between the markers and linked genes, and corresponding KEGG-based functional groups. Three dashed lines delineate genetic locations of linked genes within 10, 20 and 30 centimorgans from the two markers. Red dots indicate the linked genes discussed in the text. Putative functions and genetic positions of genes are found in Table 4.2.

4.4 Discussion

This study provides compelling evidence that the two functional loci (C29226S281 and C70S236) are true quantitative trait loci (QTL) in the Palm Island populations of the common coral *Acropora millepora* on the Great Barrier Reef. The presence of G alleles at C29226S281 and T alleles at C70S236 are indicative of significantly higher anti-oxidant capacity associated with tolerance to thermal and water quality stress.

Colonies with the tolerant alleles showed constitutively higher anti-oxidant capacity, particularly in the homozygous state. A difference in antioxidant expression may be attributed to genotypic variation in constitutive gene expression which

potentially affects response time at the protein level and thus tolerance of corals to frequent environmental insults. This is consistent with evidence of frontloading of thermal tolerance genes, such as heat shock proteins and antioxidant enzymes, in a transcriptomic study on experimentally heat-stressed colonies of the congener, *Acropora hyacinthus*, in American Samoa (Barshis et al. 2013). My conclusion is supported by the close linkage I found between genetic markers and genes putatively involved in genetic information processing (i.e. genes involved in transcription and translation in model organisms) (Fig. 4.5).

The markers are also linked to many genes involved in metabolism (Fig. 4.5). Metabolic coevolution through interacting genomes between coral hosts and symbiotic algae (Shinzato et al. 2011; Dunlap et al. 2013) suggests that a breakdown in the intimate nutritional symbiosis is an integral component of the bleaching mechanism. Cunning & Baker (2012) found that symbiont density increases with temperature prior to bleaching. The high algal density may be partly associated with increased nutrient flux (i.e. metabolism) from coral host to algal symbionts in response to elevated temperature (Coles & Jokiel 1977). In addition to temperature, external nutrients, such as ammonium and phosphate are known to increase algal density and chlorophyll concentration (Hoegh-Guldberg & Smith 1989; Muscatine et al. 1989) leading to high oxygen concentrations and the subsequent formation of abundant ROS (Shick & Dykens 1985). Therefore, differences in metabolism (i.e. nutrient trafficking) among genotypes may result in differentiation in thermal tolerance or ROS production.

The markers are moderately linked to Hsp70 and Hsp90 genes, and tightly linked to RAD7, ERH, PP2C and SEC24 genes, which are responsive to bleaching-associated stressors, particularly high temperature, ultraviolet irradiation, hypoxia and ROS (Table 4.2). These stress factors are intricately associated with mitochondrial and endoplasmic reticulum stress and disruption of Ca^{2+} homeostasis via oxidative stress (Dunn et al.

2012). The two organelles are involved in Ca^{2+} signalling pathways regulating mitochondrial ATP synthase (Jouaville et al. 1999), cytoskeletal networks via calmodulin (Lynn et al. 2001), and apoptosis (Rao et al. 2004). These Ca^{2+}-dependent functions at the organelles are exactly the putative roles of the genes linked to C70S236 (ATP5S), and to C29226S281 (HMMR and calnexin). A transcriptomic study in the Caribbean coral, *Montastraea faveolata*, suggested that thermal stress and bleaching affect Ca^{2+} homeostasis, and subsequently cytoskeletal organization and apoptosis (Desalvo et al. 2008). PP2C, which is tightly linked to the C29226S281 marker identified in this study, is involved in the negative regulation of signalling pathways induced by heat shock, ultraviolet irradiation and oxidative stress, and is implicated in the coordination of cellular responses to these stressors (Table 4.2). Close linkages of the identified markers with genes heavily involved in the stress response are consistent with the genotype-dependent stress tolerance observed in experimental studies here and with G x E associations detected at these loci in field surveys (see Chapter 3).

Table 4.2 Putative functions of linked genes. A) Genes linked to C29226S281. **B)** Genes linked to C70S236

Linkage	Marker Name	Position (centimorgan)	Gene name	Putative function	Reference
Loose	C14364S490	-25.584	DLL	Notch ligand	Munder et al. 2010
	C3724S507	-24.045	FBXL9	Homeostasis by ubiquitin-proteasome proteolysis	Wang et al. 2014
	C14018S197	-17.577	EhaC	Methanogenesis	Lie et al. 2012
Moderate	C19797S331	-16.18	ELMO1	Removal of apoptotic cell	Elliott et al. 2010
	C27153S258	-15.683	HMMR	Associated with calmodulin that is responsive to UV and thermal stress	Assmann et al. 1999; Aranda et al. 2011
	C7134S210	-3.368	DNAH	Movement of cilia and flagella	Mazor et al. 2011
	C13990S341	-1.888	PP2C	Regulation of stress response; oxidative stress response	Gaits et al. 1997; Schweighofer et al. 2004
	C29226S281	0	Vps36	Vacuolar sorting; salt-inducible	Serrano et al. 2003
Tight	C17330S121	0.037	SEC24	Oxidative and UVR stress response	Aranda et al. 2011; Belles-Boix et al. 2000
	C18920S453	0.585	GSH	Development in nervous system	Valerius et al. 1995
	C18185S479	2.81	PP2C	Regulation of stress response; oxidative stress response	Gaits et al. 1997; Schweighofer et al. 2004
	EST149	8.339	PNPPase	Embryo-larva transition in honeybee	Gala et al. 2013
	C7889S263	12.54	Hsp70	Thermal stress response; chaperone for protein folding and cell regulation	Mayer & Bukau 2005
	C18487S1302	12.593	AFF3	Expressed in lymphoid cells	Ma & Staudt 1996
	C11797S545	13.002	TC.GPH	Sugar transporter	Saier 2000
Moderate	C22633S340	13.859	GMPP	An intermediate for a synthesis of L-ascorbic acid which plays an antioxidant role	Hirayama & Suzuki 2014
	C10625S161	14.225	Calnexin	Apoptosis	Guerin et al. 2008
	C13301S439	17.269	GMPP	An intermediate for a synthesis of L-ascorbic acid which plays an antioxidant role	Hirayama & Suzuki 2014
	C1063S181	17.955	Nitrate reductase 2	Nitrate reduction	Smith et al. 2007
Loose	C19928S437	20.56	MYH	Cellular movement	Weiss & Leinwand 1996
	C20443S297	29.637	DIAPH2	Regulation of smooth muscle cell differentiation	Staus et al. 2007

Linkage	Marker Name	Position (centimorgan)	Gene name	Putative function	Reference
Moderate	C15891S454	-15.847	EIF3K	Translation regulation in response to apoptosis	Lasfargues et al. 2013
	C8136S163	-14.543	HSP90A	Thermal stress response; chaperone for protein folding and gene regulation	Zhao & Houry 2005
	C5904S135	-11.509	LTBP (or TGF-β)	Cytostasis and apoptosis	Siegel & Massague 2003
	C25225S451	-10.536	ACTC1	Notch signalling	Noseda et al. 2006
	EST032	-7.551	RAD7	DNA repair	Singh 2004
	C6723S318	-6.663	PSEN2	Notch signalling	Nornes et al. 2009
	C12395S564	-3.338	GCAT	Threonine metabolism	Opperdoes & Coombs 2007
Tight	C15021S282	-0.435	SCY1-like	Protein trafficking between the Golgi and ER	Burman et al. 2008
	C70S236	0	FBXW5	Regulation of centrosome duplication	Puklowski et al. 2011
	C15238S417	3.327	ATP5S	ATP synthesis	Singh 2004
	C23525S293	5.538	ERH	Hypoxia stress response	Ning et al. 2004; Boswell et al. 2009
	C11670S169	7.474	SLC3A1	Amino acid metabolism	Cochat et al. 2010
	C21844S313	12.294	TMEM229B	Carbohydrate and fat metabolism	Tonne et al. 2013
Moderate	C25713S318	17.231	GNG7	Response to cancer	Demokan et al. 2013
	C16442S295	19.381	SMC	Chromosome maintenance	Singh 2004
	C10924S223	23.813	MAAI	Tyrosine metabolism	Fernandez-Canon et al. 2002
Loose	C29432S370	26.478	GANAB	ER stress	Yan et al. 2012
	C24388S705	27.667	NOTCH	Response to UVR; notch signalling	Aranda et al. 2011

The two SNP markers evaluated in this study are representatives of the hard-wired genetic components underpinning phenotypic variation, and can be used in a wide range of applications relevant to coral reef management and restoration. Extrinsic factors, such as temperature and water quality, are critical drivers of the degradation of coral populations. Spatial mapping of the C29226S281 and C70S236 genotypes can provide high-resolution data to predict bleaching susceptibility of corals and enable identification of resilient and susceptible populations. In combination with information on connectivity among populations, my results enable key targets for conservation to be identified, such as highly resilient populations that have the ability to seed surrounding reefs. However, it should be noted that, as seen in this study, genotypes can display phenotypic variation and genotype-phenotype associations may not be strongly linked in some populations for the following reasons: bleaching tolerance may be governed by multiple quantitative trait loci (QTLs) having small effects (see Discussion in Ch. 5); trade-offs exist between tolerance and fitness-related traits (i.e. reproduction and growth) (see Discussion in Ch. 6); other loci that are preferentially selected and expressed in some environmental conditions may counteract or offset effects of the tolerant alleles. These issues can be partly overcome by incorporating multiple genetic markers in a spatial mapping model, thereby increasing the resolution of the method for identifying tolerant corals.

Furthermore, active human interference through assisted migration and selective breeding may be necessary in order to facilitate survival of coral reefs into the future (van Oppen et al. 2015). Individuals with double homozygous genotypes at the two genetic markers, such as those identified in this study, can be targeted to select stress tolerant brood stock for translocation and selective breeding to restore highly damaged reefs. A great advantage of the use of genetic markers for identifying resilient corals, rather than relying on historical environmental averages, is the greater resolution

associated with the former (i.e., to the colony rather than reef level). Detection of within-population genetically-determined phenotypic variants minimizes issues often encountered in transplant studies, such as acclimatization-associated trade-offs and environmentally-regulated heritable factors (Brown & Cossins 2011).

In conclusion, this study has revealed the first QTLs for environmental stress tolerance in reef-building corals. The identification of stress-tolerant genotypes will enable exploration of new management and restoration options for the world's rapidly degrading coral reefs.

Chapter 5.0 Predicting the spatial distribution of allele frequencies for a bleaching tolerance gene in *Acropora millepora* using Bayesian belief network modelling

This chapter was prepared for publication but not yet submitted.

Young K. Jin, Stuart Kininmonth, Petra B. Lundgren, Madeleine J.H. van Oppen and Bette L. Willis

YKJ collected allele frequency and environmental data. YKJ developed and analysed a BBN model. SK provided statistical advice. YKJ wrote the chapter with intellectual contributions from all co-authors.

Abstract

As the pace of environmental change on coral reefs increases, the necessity to find structure and predictability in the capacity of corals to respond adaptively in a complex ecological system is becoming increasingly urgent. Recent evidence that coral populations in the Palm Islands with high frequencies of specific alleles are associated with resistance to bleaching (Chapter 4) provides an opportunity to use environmental mapping to identify bleaching-resistant populations. In this study, a spatial Bayesian belief network (BBN) model was developed to predict the spatial distribution of allele frequencies for a specific locus (C29226S281) found to be associated with high antioxidant capacity and bleaching resistance in *Acropora millepora* in the Palm Islands. A combination of long-term environmental monitoring data, allele frequency data, expert input and statistical evaluation was used to build the model, with the goal of refining prior beliefs and examining dependencies among environmental variables that are proxy indicators of selective forces on allele frequencies at locus C29226S281. The model predicts that coral populations with high allele frequencies are found in habitats characterised by poor water quality and a large range of seasonally-fluctuating temperatures. The Bayesian simulation approach demonstrates that the synergism between highly fluctuating temperatures and high nitrate concentrations may be the primary driver of selection for alleles at this locus. However, high temperature ranges in isolation do not increase the uniform probability of finding high allele frequencies across the spatial domain. Consistent with this, spatial mapping of predicted allele frequencies reveals that bleaching tolerance alleles are most likely to be concentrated in populations near the mouths of the Burdekin and Fitzroy Rivers, where nutrient levels are chronically high. Coupling GIS and BBN enables visualisation of the model output, specifically qualitative measurements of allele frequencies, enabling non-experts to explore and

interpret complex interactions between environmental and genetic factors. This approach opens up new opportunities for more efficient and effective coral reef management and conservation through direct intervention to ensure coral populations have the genetic diversity needed to optimise adaptation to rapid climate change.

5.1 Introduction

Corals are facing severe impacts from multiple interacting environmental threats, including rising sea temperatures, ocean acidification, and degrading water quality (Wooldridge 2009; Lesser 2011; Anthony et al. 2015; Wooldridge et al. 2015). The impacts of these stressors are magnified on coral reefs given the narrow environmental niches of corals, which are restricted to narrow depth, substrate and temperature ranges. The key question now is can corals acclimatise, migrate, or evolve with sufficient pace to survive in a rapidly changing environment. Future projections based on biochemical, physiological, and ecological modelling of coral bleaching suggest that the current rate and magnitude of climate change may be beyond the adaptive capacity of the majority of naturally-occurring coral populations (Hoegh-Guldberg 1999; Frieler et al. 2013). Evolutionary changes are difficult to predict given the complexity of stress factors, but the rate of genetic change is unlikely to be a panacea in the face of unprecedented rates of environmental alterations. Given the global spatial configuration of coastal habitats, the potential for corals to move away from the equator to alleviate thermal stress is limited (Guinotte et al. 2003). As a conservation countermeasure against rapid environmental changes, researchers have proposed assisted evolution through transplantation and selective breeding to accelerate adaptive responses to rapidly changing environmental conditions (Thomas 2011; van Oppen et al. 2014). The motivation for such propositions arises from the notion that translocation of tolerant

individuals artificially helps to enhance adaptive potential and resilience of coral populations.

When based solely on observed phenotypes and environmental data, the success of coral transplantation across different environments is notoriously erratic (Howells et al. 2013). This may be primarily attributed to limitations on the acclimatisation potential of corals because acclimatisation capacity may be governed by historical thermal regimes (Howells et al. 2013). To develop alternative approaches to assisted adaptation, information about the heritability of genes associated with stress tolerance is required. A phenotype, such as bleaching response, is likely to be induced by multiple variables (Lesser 2011), thus inclusion of a number of relevant factors in a prediction model potentially increases its resolution in defining stress (bleaching) thresholds. I found significant correlations between the spatial distributions of two key stress factors, temperature and water quality, and spatial patterns in allele frequencies at two genetic loci in *A.millepora* (Chapter 3). In Chapter 4, validation experiments revealed typical genotypic associations with bleaching tolerance and antioxidant capacity in the Palm Island populations. Functional annotations obtained from ZoophyteBase (Dunlap et al. 2013) suggest that the loci identified in Chapter 4 were linked to a range of genes involved in bleaching-associated processes, notably oxidative and thermal stress. Although it has been shown that increases in cellular levels of reactive oxygen species are caused by a multitude of environmental stressors, including elevated temperature, solar radiation and nutrients (Lesser 2011; Cunning & Baker 2012), experimental investigations incorporating complex interactions among environmental variables remain very difficult to execute. Consequently, to provide knowledge for conservation decision science, we need other approaches, such as modelling, to understand spatial relationships between coral genotypes and environmental stressors.

Lack of knowledge about the Great Barrier Reef (GBR)-wide distribution of alleles at locus C29226S281, the locus found to be associated with tolerance to bleaching-associated stressors (Chapter 4), combined with uncertainty about the nature of complex interactions among predictor variables, highlight the need for a modelling approach to develop predictions about the distribution of a tolerant allele at the locus. Therefore, I elected to model correlative relationships between large environmental and genetic gradients within a Bayesian Belief Network (BBN) framework. BBN is a statistical model with a graphical depiction of the probabilistic dependencies between predictor variables and response variables. BBN can maximise the ability to make probabilistic inferences about a system under investigation by merging different types of information, both subjective (e.g. expert opinion) and quantitative (e.g. monitoring data). This enables BBN to develop models representing systems that are inherently complex and uncertain because of incomplete knowledge and understanding of the system under study. BBN models also enable the putative responses of populations investigated to be extrapolated to other 'equivalent' populations that were not surveyed due to constraints of time, cost and logistics.

In this chapter, my over-arching goal is to increase understanding of large-scale patterns of allele frequency distributions of a bleaching tolerance gene along the GBR. More specifically, my objectives are to characterise the distributions of potential environmental drivers of the bleaching tolerance gene, determine if environmental drivers interact synergistically, and visually map spatial patterns of allele frequencies across the GBR onto distributions of environmental drivers.

5.2 Methods and materials

5.2.1 Genetic and environmental data

Allele frequency data for locus C29226S281 were obtained from 18 coral populations distributed along ~1,500 km of the GBR, as described in Chapter 3 (Fig. 5.1). The species targeted was *Acropora millepora*. Genetic data in the southernmost central reefs and southern offshore reefs were excluded because populations in these regions were found to be genetically divergent (Chapter 3). Environmental data in the corresponding regions were also excluded from this study.

Environmental data were obtained for 168 sites along the GBR directly from spatial prediction maps based on long-term monitoring data gathered by the Australian Institute of Marine Science (http://eatlas.org.au/data/uuid/df7012eb-a23f-4cf2-b92d-6fff77718987) (Supplementary Fig. S.1). Sites covered the 18 reefs that were sampled for allele frequency data (Fig. 5.1; Supplementary Table S.3). The study area encompasses a wide range of habitats and environmental conditions, including temperature and water quality, which differ in gradient direction (i.e. a latitudinal gradient in temperature, an inshore-offshore gradient in water quality). Areas frequently exposed to flood plumes from the largest river catchments (Burdekin and Fitzroy Rivers) are also included in the study (Devlin et al. 2012). Latitudes range between 11°S and 23°S and distances to shore from sampling points ranged from less than 1km to more than 130km.

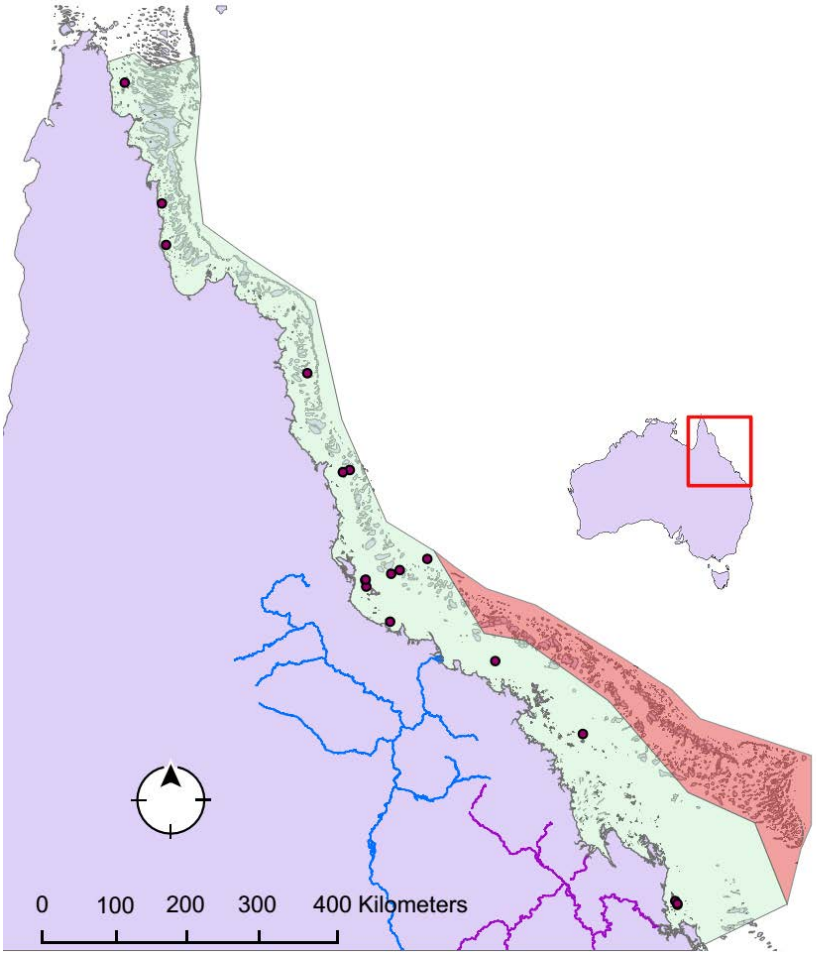

Fig. 5.1 Map showing the 18 sampling sites along the Great Barrier Reef for which genetic data were collected (red circles), the area for which environmental data were available (green polygon), the region excluded from analyses (red polygon), and the Burdekin and Fitzroy River catchments (blue and purple lines, respectively)

5.2.2 Environmental correlation network

Nine environmental variables (i.e. latitude, sea surface temperature range, distance from shore, secchi depth, suspended sediment, particulate phosphorus, particulate nitrogen, chlorophyll and nitrate) (see Fig. 5.1, Supplementary Table S.3 and Table 5.1 for information on spatial and temporal scales and units of the environmental variables, respectively) were selected as potential environmental drivers because of effects of temperature and water quality on a locus C29226S281 identified in Chapter 3 and 4 as well as potential interactions among the variables (see below for a detailed description).

To develop a causal diagram of correlations among the nine environmental variables selected for model development, pairwise Pearson correlation coefficients were calculated using the *rcorr* function from the R package, *Hmisc* (http://www.R-project.org). Using the correlation values, environmental correlation networks were created with tools from the igraph package for R (Csardi & Nepusz 2006). Environmental variables are represented by vertices that are connected by lines showing interactions among the variables (Fig. 5.2). Negative signs, which denote negative correlations, were removed to show a simple degree of interaction among variables. Correlation values below 0.6 were also removed from the network graph to show the dominant interactions, with the exception of a line connected to the latitude and sea surface temperature (SST) variables that did not show higher correlation values (Fig. 5.2). This network formed the basis for the BBN development.

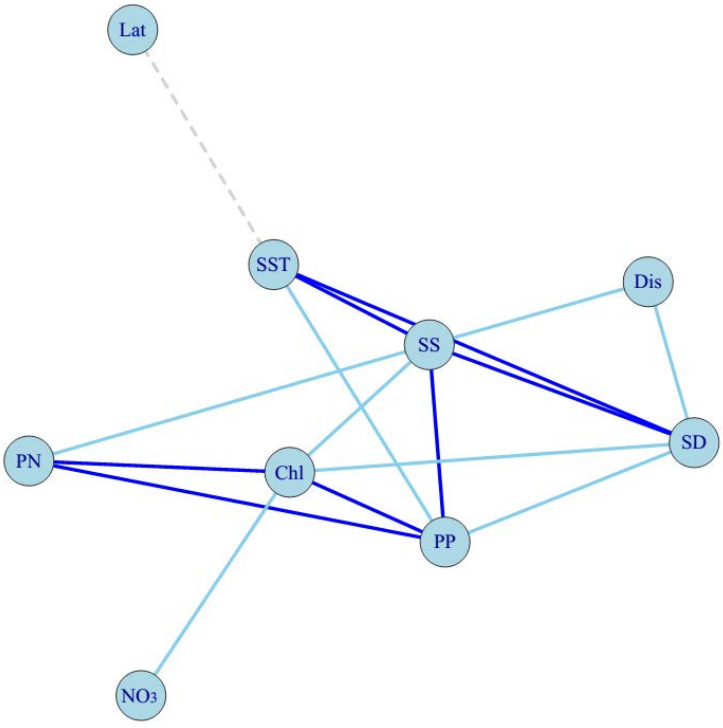

Fig. 5.2 Environmental correlation network showing high and medium correlation coefficients (dark blue lines=>0.7; light blue lines=0.6-0.7; dashed line=0.449-0.6, respectively). Abbreviations: Lat=latitude; SST=sea surface temperature; Chl=chlorophyll; Dis=distance from shore; SD=secchi depth; NO_3=nitrate; SS=suspended solids; PN=particulate nitrogen; PP=particulate phosphorus.

5.2.3 Bayesian belief network

The BBN model was structured to make probabilistic inferences about the influence of environmental variables, particularly those related to water quality and temperature, on allele frequency patterns in the coral *A. millepora* throughout the Great Barrier Reef region. The model was developed based on the environmental correlation network (Fig. 5.2), observations of interactions among the environmental parameters in the literature, and previous results on allele frequency patterns (see Chapter 3) using the Netica software package (Norsys Software, Vancouver, British Columbia, Canada) (Fig. 5.3). The direction of the arc between chlorophyll and NO_3 was purposefully reversed to ensure that joint probabilities were computed with conditional independences among the strongly correlated variables PP, PN and chlorophyll (i.e. NO_3 was not included as a factor affecting chlorophyll when considering the PP-PN-chlorophyll network). NO_3 was excluded from these joint probability calculations because it was only correlated with chlorophyll (Fig. 5.2). Hence, probabilistic associations, as well as causal relations between these variables and chlorophyll, were represented in the graphical structure. Detailed descriptions of the linkages and discretised states for each node are given in Fig. 5.3 and Table 5.1, respectively.

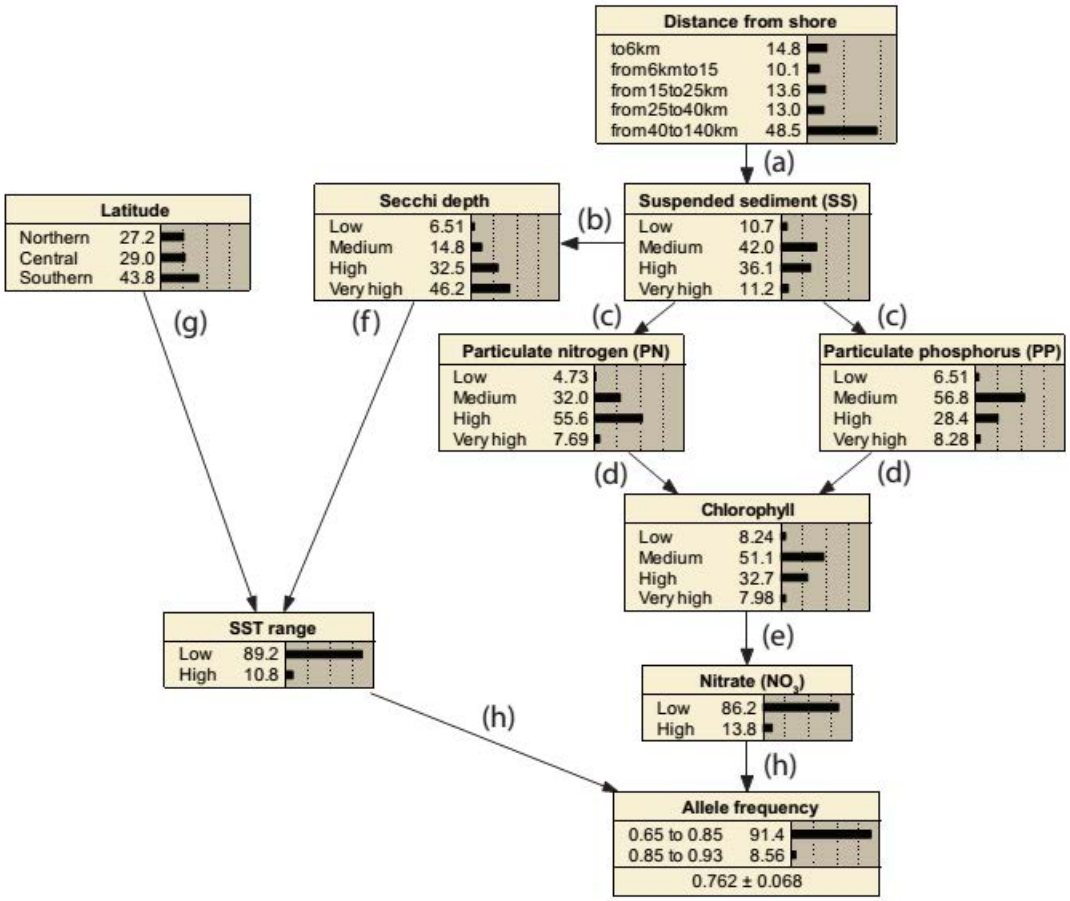

Fig. 5.3 Model structure of a Bayesian belief network predicting how water quality and temperature-related variables affect allele frequencies of the bleaching tolerance locus C29226S281 in *Acropora millepora*. Linkages (lines between the nodes) are defined as conditional probability relationships between child and parent nodes. Descriptions of each linkage are given below.

(a) Suspended sediment is more abundant in shallow inshore regions due to re-suspension caused by wave action, bioturbation (Larcombe et al. 1995) and river discharge (Brodie et al. 2003).

(b) Secchi depth (water clarity) is primarily affected by suspended sediment concentrations (De'ath 2007).

(c) Source areas for PP, PN and SS are basically the same for all Queensland coastal rivers, as particulate matter is typically associated with soil erosion due to grazing and cropping on adjacent land (Brodie et al. 2003; De'ath 2007).

(d) Chlorophyll can be measured as a proxy for phytoplankton biomass. Regions with high primary production are expected to have increased abundance of consumers. Both living and dead plankton contribute to increases in PP and PN concentration. Particulate nutrients are broken down by bacteria and zooplankton and become available for plant growth (Furnas et al. 2011). Additionally, discharge of both particulate and dissolved nutrients to the Great Barrier Reef occurs during river flooding in wet seasons, and coincides with phytoplankton blooms (Packett et al. 2009)

(e) Chlorophyll takes up and assimilates dissolved nitrogen quickly (Furnas et al. 2005). The assimilated nitrogen (N) is subsequently recycled through pelagic food webs via zooplankton, larger predators (e.g. fish and invertebrates), then bacteria, maintaining

high and persistent levels of dissolved inorganic nitrogen (available to plant growth) in the water column (Furnas et al. 2011). Therefore, phytoplankton acts as a moderator and mediator to help incorporate and maintain DIN in the system. Enriched N due to long residence time and increased inputs of N from land is reflected in chlorophyll concentrations in some zones (Brodie et al. 2011; De'ath & Fabricius 2008).

(f) Secchi depth is geographically correlated with temperature fluctuations (Fig. 5.2). Patterns may be attributed to multiple factors. Temperatures are more stable towards the equator and tend to increase in inshore waters due to shallow depths and little exchange with offshore cold waters. Inshore regions are also characterised by low water clarity (secchi depth) because of river run-off during summer wet seasons (Fabricius et al. 2014). Geographical variation in secchi depth may be partly due to large regional differences in annual export of suspended sediments (Brodie et al. 2003). Furthermore, upwelling in offshore waters in the northern GBR brings deep cold water onto the shelf leading to reductions in water temperature (Furnas & Mitchell 1996), and possibly diffusion of sediments and subsequent higher water clarity.

(g) Sea surface temperature is correlated with distance from the equator, as described by latitude values.

(h) Allele frequencies at the C29226S281 locus are strongly correlated with NO_3 (Chapter 3), a variable known to act synergistically with thermal stress to induce bleaching (Cunning & Baker 2012). SST range rather than mean SST was used as a factor affecting allele frequencies in the model because bleaching tolerance alleles are more abundant in thermally unstable environments. The locus is linked to genes related to bleaching-associated stress responses (Chapter 4).

Table. 5.1 Discretisation for nodes used in the BBN model. Discretised states of PN, NO_3, PP, SS, chlorophyll and secchi depth nodes are based on environmental maps obtained from the e-atlas website (see http://eatlas.org.au/data/uuid/df7012eb-a23f-4cf2-b92d-6fff77718987) and also on mean annual values estimated for coastal, inshore and offshore regions of the Great Barrier Reef (De'ath & Fabricius 2008). Discretised states for SST range and NO_3 were based on spatial patterns in correlations with allele frequencies at the C29226S281 locus across the 18 study reefs, as verified by testing the accuracy of model predictions for two versus four states of these environmental variables (see 5.3.3). Guidelines given by Marcot et al. (2006) were also followed.

Variable	States	Values
PN (micromol/litre)	Low	0.53-0.958
	Medium	0.958-1.423
	High	1.423-2.113
	Very high	2.113-3.137
NO_3 (micromol/litre)	Low	0.015-0.359
	High	0.359-1.262
PP (micromol/litre)	Low	0.035-0.053
	Medium	0.053-0.081
	High	0.081-0.125
	Very high	0.125-0.236
SS (mg/litre)	Low	0.256-0.574
	Medium	0.574-1.287
	High	1.287-2.887
	Very high	2.887-9.696
Chlorophyll (micrograms/litre)	Low	0.14-0.25
	Medium	0.25-0.43
	High	0.43-0.74
	Very high	0.74-1.67
Secchi depth (m)	Low	2.8-4.68
	Medium	4.68-7.8
	High	7.8-13
	Very high	13-27.9
Distance from shore (km)	to6km	
	from 6km to 15km	
	from 15 to 25km	
	from 25 to 40km	
	from 40 to 140km	
SST range (°C)	Low	3.77-6.3
	High	6.3-7.59

5.2.4 Predictability of BBN model

To test the accuracy of BBN model predictions, I randomly removed 11% (n=2), 22% (n=4), 33% (n=6), and 44% (n=8) of the data available for the 18 populations with known allele frequencies, a process repeated 5 times for each percentage reduction, and probabilities were recalculated from the respective reduced dataset. Error rates were calculated in a confusion matrix ("test with cases" function in the Netica software) as the proportion of incorrect predictions (type I and II errors) with respect to observed

data. Additionally, error rates were calculated and compared among two models with two or four states at both the NO_3 and SST range nodes.

5.2.5 Sensitivity analysis

Based on entropy reduction, sensitivity analysis was conducted to identify how much the allele frequency node was influenced by a single finding at other nodes in the network. This analysis shows the influential level for each variable on the allele frequency node. I further examined changes in the sensitivity of allele frequencies after high NO_3 or high ranges in SST values were entered as evidence to investigate potential combined effects of NO_3 and SST range on allele frequencies. Complementary to the sensitivity analysis, I also examined ranges in the likelihood of each state of each environmental variable after evidence of allele frequencies was entered.

5.2.6 Simulation analysis

Learning in the BBN was performed based on the web structure and evidence data using the expectation maximisation algorithm (Dempster et al. 1977), and the model was used to simulate cases where states of NO_3 and SST range were known. The simulation test explores the response of the allele frequency node to changes across the NO_3 and SST range nodes to examine potential thresholds or rapid changes that might be present in the response. It should be noted that the simulation test may have been limited due to the small size of the allele frequency data in some parts of the environmental range. However, the strength of BBN lies in its ability to build causal webs based on expert knowledge (i.e. experimental evidence from Ch. 3 and 4 and findings from the literature - see legend for Fig. 5.3.) that maximises inferences when limited information is available.

5.2.7 Spatial mapping

Finally, I visually investigated spatial patterns in the likelihood that populations of *A. millepora* throughout the GBR contained high allele frequencies. This involved mapping the distributions of allele frequency values predicted through BBN using ArcGIS software (ESRI http://www.esri.com).

5.3 Results

5.3.1 Prediction of likelihood of allele frequencies through BBN

GBR-wide patterns in the spatial distribution of allele frequencies associated with bleaching tolerance in *A. millepora* were predicted by the BBN model based on maps of empirically-collected and satellite-derived environmental data related to water quality and temperature. The model predicts that coral populations with a high frequency of bleaching-resistant alleles at locus C29226S281 are found in regions that correlate with poor water quality and high temperature ranges. Spatial mapping of the predicted allele frequencies reveals that bleaching tolerance alleles are most likely to be concentrated near the mouths of the two largest rivers (i.e. Burdekin and Fitzroy Rivers) flowing into the Great Barrier Reef lagoon.

Specifically, the model predictions revealed the significance of individual water quality parameters driving high frequencies (0.85-0.93) of the C29226S281 bleaching tolerance allele (Fig. 5.4). For example, inshore regions of the central and southern GBR that are characterised by very high levels of chlorophyll concentrations (0.74-1.67 ug/l) are more likely to be associated with high frequencies (58%) than low frequencies (3%) of the tolerant allele. Reefs with high nitrate concentrations (0.359 and 1.262 umol/l) show the strongest association with high frequencies of the bleaching tolerance allele. The probability of corals having high allele frequencies was 8 times more than the probability of corals having low frequencies in regions with very high levels of either PP

(0.125-0.236 umol/l) or PN (2.113-3.137 umol/l) (42 and 39%, respectively). Similarly, probability of finding high allele frequencies was highest (51%) with a very high level of SS (2.887-9.696 mg/litre) among other states of the SS node. There was a 30% probability of finding high allele frequencies when secchi depth measurements were low (2.8-4.68 m), whereas there was a 49% probability of allele frequencies being low when secchi depth measurements were very high (13-27.9 m). The probability of finding tolerant alleles was more likely (47%) in regions within 6km of coasts. In contrast, in offshore regions (more than 40km away from coastlines), the probability of corals having low frequencies of the allele was 51%. There were reductions in probability of finding both high and low allele frequencies when SST range was low, but to different degrees (i.e. 84% and 22% for low and high allele frequencies, respectively).

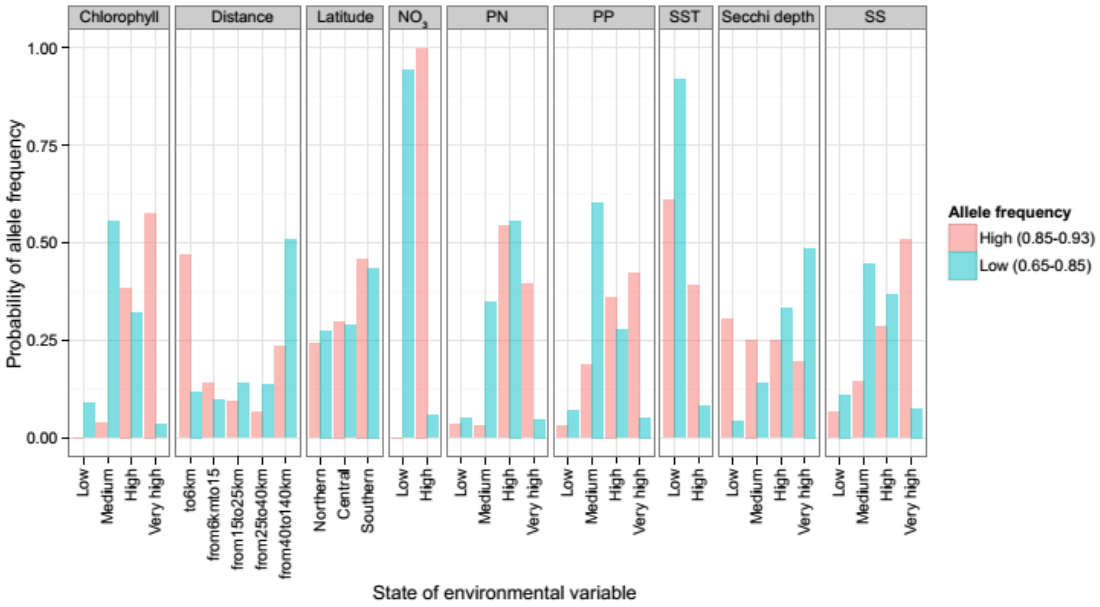

Fig. 5.4 Marginal probabilities of the environmental variables given the likelihood of finding high or low allele frequencies. Pink and blue bars show inferred probabilities for each state of the environment nodes. For example NO_3 concentrations are likely to be significantly higher (i.e. NO_3 variable state is high) when the allele frequency at locus C29226S281 is high, and much lower (i.e. NO_3 variable state is low) when the allele frequency is also low.

When poor water quality or high temperature range were specified alone or in combination, the probabilities for high allele frequencies increased by 30-100%

compared to when good water quality or low temperature range were used (Table 5.2). For example, high levels of SST range and NO_3 increased the probabilities by 25.3% and 62.2%, respectively. When very high levels of PP and PN were considered jointly, probabilities increased by 60.4%. With very high levels of SST range and NO_3 in combination, probabilities increased to 100%.

Table 5.2. Likelihood of finding high allele frequencies at locations associated with good (top) or poor (bottom) water quality, and low (top) or high (bottom) temperature range states.

Variable	Evidence entered	Probability of high allele frequency (%)
Distance from shore	40-140km	4.14
SST range	Low	5.84
Secchi depth	Very high	3.61
SS	Low	5.23
PN	Low	6.17
PP	Low	3.77
PP + PN	Low + Low	1.18
Chlorophyll	Low	0
NO3	Low	0
SST range + NO3	Low + Low	0
Variable	Evidence entered	Probability of high allele frequency (%)
Distance from shore	6 km	27.1
SST range	High	31.1
Secchi depth	Low	39.9
SS	Very high	38.6
PN	Very high	43.8
PP	Very high	43.7
PP + PN	Very high	61.6
Chlorophyll	Very high	61.8
NO3	High	62.2
SST range + NO3	High + High	100

BBN-based spatial mapping shows that high frequencies of the bleaching tolerance allele at locus C29226S281 occur in locations near the mouths of the Burdekin and Fitzroy rivers, which are both characterised by high levels of NO_3 concentrations and large SST ranges (Fig. 5.5A). Sites with high SST range and high NO3 were found in 8% of the study area (Fig. 5.5B). The Bayesian simulation approach demonstrates that high SST range does not increase the probability of finding high allele frequencies in isolation, but when combined with high NO_3, the influence of SST is significant, indicating the synergism between SST range and NO_3 (Fig. 5.5C).

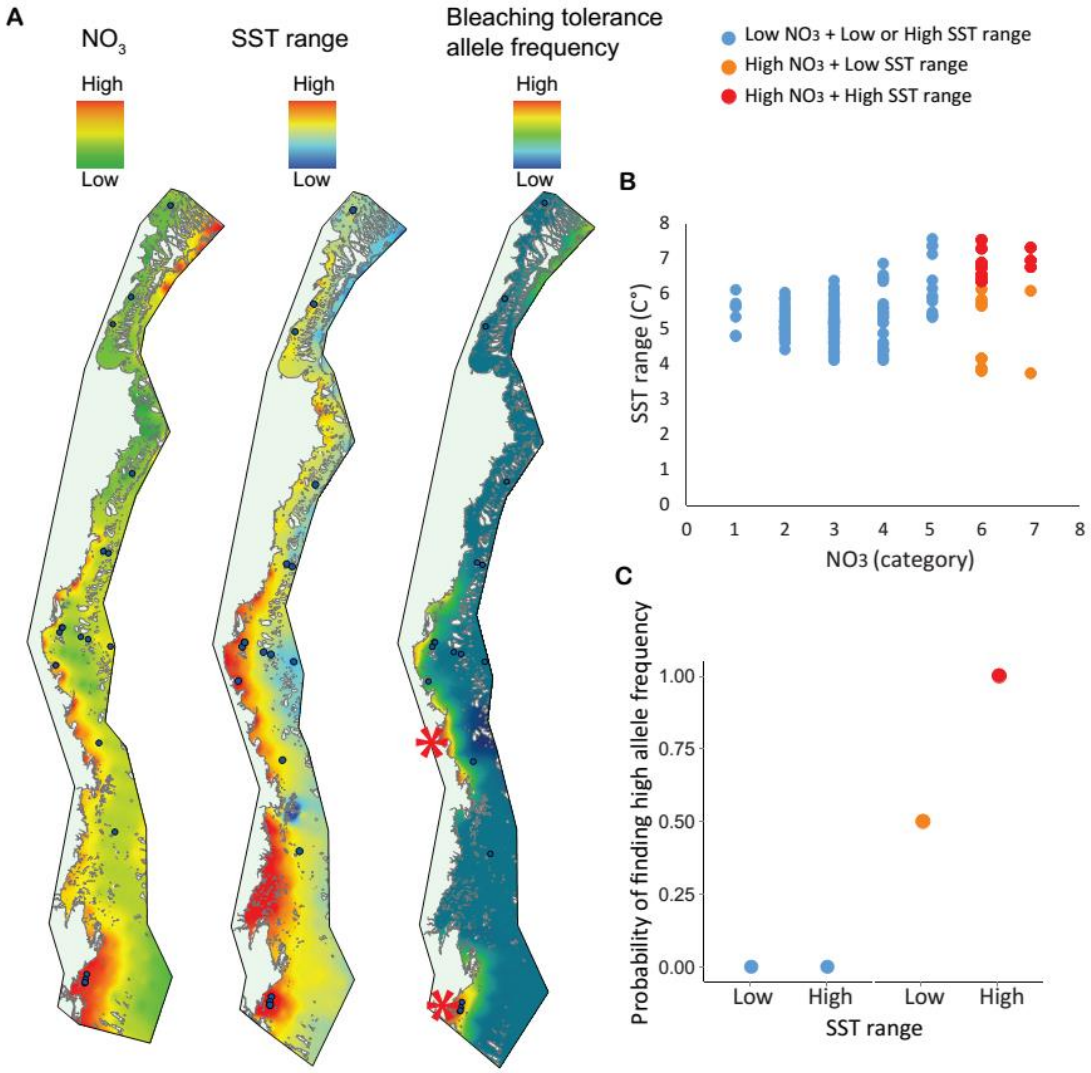

Fig. 5.5 BBN-based predictions of GBR-wide frequencies of a bleaching tolerance allele in *Acropora millepora*. (A) Individual strip maps (left to right) show spatial variation in: NO_3, SST range, and predicted likelihood of finding high frequencies of the bleaching tolerance allele. Areas of high allele frequency were predicted through the BBN model (Fig. 5.3). Sampling sites for genetic data are indicated by blue circles. Red asterisks show the locations of mouths of the Burdekin and Fitzroy Rivers. **(B)** The distributions of sites with combinations of different SST range and NO_3 levels on the GBR. **(C)** BBN simulation analysis of the synergistic effects of SST range and NO_3 on probabilities of finding high frequencies of the bleaching tolerance allele.

5.3.2 Sensitivity

With no evidence entered, sensitivity analysis showed that NO_3 is the most influential factor in predicting high frequencies of the bleaching tolerance allele, while SST range had a weak influence (Fig. 5.6). Despite the low significance of SST range by itself, with high SST range entered as evidence, the sensitivity of the allele frequency

node to water quality variables increased significantly, as was found in the analyses described above. The highest gain in sensitivity was observed for NO_3, increasing it to 100 %.

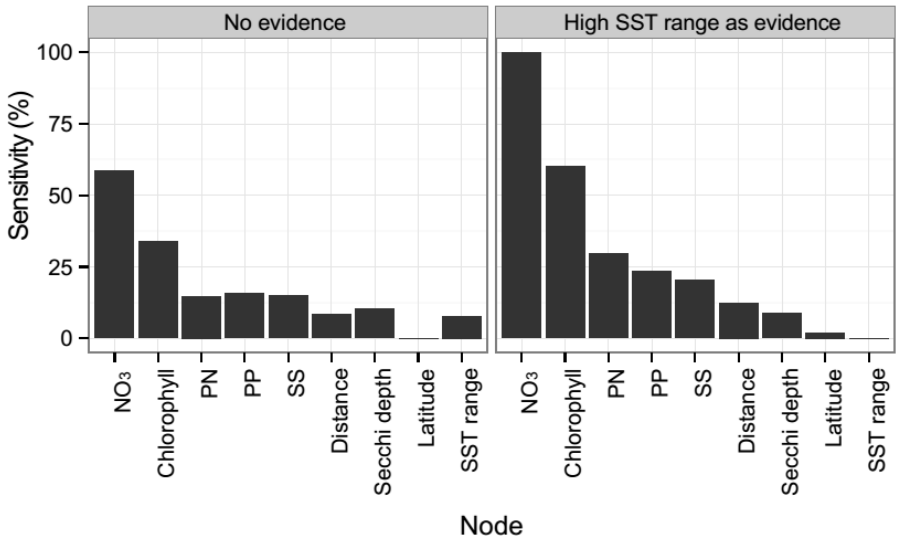

Fig. 5.6 Sensitivity of allele frequency predictions at locus C29226S281 to environmental variables with no evidence (left) and high SST range entered (right).

5.3.3 Predictability (error rates)

The BBN model correctly assigned all the predicted values to the observed allele frequency class, after 33 % (n=6) of the allele frequency data were removed (Fig. 5.7). The models with NO_3 and SST range nodes having 4 states had low accuracy in allele frequency predictions. The robustness of the model with 2 states for the two nodes, despite the limited sample size, indicates the clearly defined nature of the correlation structure of allele frequencies across populations in the data set. In particular, allele frequencies of populations near the Fitzroy river mouth are highly correlated with the key environmental variables of NO_3 and SST range.

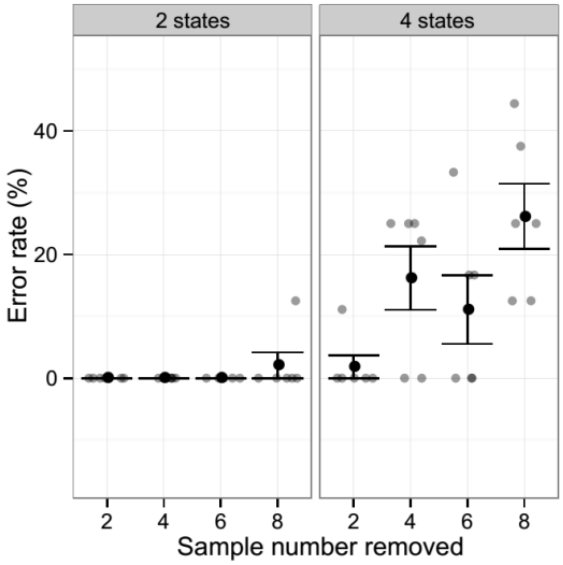

Fig. 5.7 Error rates (predictability) of the BBN models. Error rates were compared among models with 2 and 4 states of NO_3 and SST range nodes and with 2 to 8 samples removed from allele frequency data.

5. 4 Discussion

A probabilistic model of allele frequency distributions across environmental gradients indicates that, spatially, high frequencies of a bleaching tolerance allele are concentrated in populations of *Acropora millepora* located in poor water quality conditions, especially near the mouths of the two largest rivers in Queensland (Burdekin and Fitzroy Rivers; Fig. 5.5A). These two rivers export the largest amounts of suspended sediment, total phosphorus and total nitrogen into the GBR lagoon (Brodie et al. 2003). Moreover, reef regions containing populations with four of the highest allele frequencies found (i.e. Magnetic Island, Half-Tide Rocks, Halfway Island and Humpy Reef) are frequently exposed to flood plumes during wet seasons (Devlin et al. 2012), during which nutrient concentrations can increase by two orders of magnitude compared with non-flood periods (Devlin et al. 2001). Contrary to the notion that isolated pristine reefs represent refuges in the face of climate change, my results suggest that future 'rescue' effects may originate from 'polluted' regions.

Although long-term average data used in this study can underestimate episodic changes in environmental conditions, such data can implicitly identify both chronic conditions and regions with frequent flood plumes. Also, long-term and repetitive selective pressures may be more important forces in shaping detectable selection signatures in genomes when gene flow is significant, because genetic connectivity can quickly dilute genetic patterns formed by strong but episodic selective pressures. Changes in allele frequency can happen in one generation (Gagnaire et al. 2012), but also depend on effects of the loci on fitness traits and effective population size (Le Corre & Kremer 2003)

My results are consistent with the findings of studies on the exacerbating effects of NO_3 on heat-stressed *Symbiodinium* in corals (Nordemar et al. 2003; Schloder & Croz 2004). In particular, a recent study that investigated the combined effects of temperature and water quality in GBR coastal waters, which contained substantial concentrations of nutrients from soluble plant fertiliser and sediment enrichment, found that cumulative adverse effects of high temperatures and organically enriched coastal seawater significantly affected the survival of *A.millepora* (Fabricius et al. 2013). Organically enriched coastal seawater is expected to be more similar to the composition of seawater found on GBR reefs than experimental combinations of filtered seawater+NO_3. This is because naturally-occurring NO_3 gradients in the GBR are likely to encompass changes in values of other environmental variables. Indeed, the BBN and environmental network analyses suggest that throughout the GBR, nutrient parameters vary along with other variables. In combination, variation in more than one environmental variable probably affects symbiosis and holobiont fitness. For example, Humphrey et al. (2008) showed that when high levels of nutrients (inorganic nitrogen and phosphate) and SS co-occur, fertilisation is significantly reduced

in *A.millepora.* Zooxanthellae density increased but growth rates decreased in *Stylophora pistillata* when NO_3 and Fe were used in experimental combinations (Ferrier-Pages et al. 2001).

Although thermal stress is a primary driver of bleaching, poor water quality may act synergistically to induce bleaching. For example, nutrients associated with poor water quality may increase densities of algal symbionts, which then increase the susceptibility of the coral host to thermal stress and subsequent bleaching because high symbiont densities become a liability when *Symbiodinium* are photo-inhibited (Cunning & Baker 2012). This implicitly suggests that thermal stress and high nutrient concentrations promoting algal growth may act synergistically to cause bleaching.

Acute large shifts in nutrient level in coastal waters can occur in summers due to El Niño-Southern Oscillation (ENSO) events that are linked to inter-annual variability in rainfall and river flow (Lough 2007). Tolerance to such rapid environmental changes may be based on constitutively higher stress tolerance in corals. In Chapter 4, corals with bleaching tolerance alleles showed consistently higher antioxidant capacity. A transcriptomic study also demonstrated that thermally resilient corals constitutively frontload expressions of thermal and oxidative stress genes under control conditions (Barshis et al. 2013). Constitutively higher tolerance may allow corals to maintain their health (or store more energy reserve) in chronically stressful environments, such as those near the river mouths identified in this study, and hence to be resistant to future episodic disturbances.

Spatial information on the distribution of tolerance-associated alleles opens up the possibility of new approaches to coral conservation and management. Large-scale mapping of the likelihood of tolerant allele frequency distributions developed in this study allows managers to visually investigate allele distributions across many more

populations than could be realistically sampled. Such projections enable managers to target populations with high proportions of bleaching tolerance alleles for conservation. These populations are important as a source of: advantageous alleles that could be spread to affected populations, candidate colonies for transplant programs for reef restoration purposes, and breeding stock for selective breeding programs. Furthermore, knowledge of the distributions of bleaching tolerance alleles enables more efficient design of marine protected areas to ensure connectedness among populations that promote adaptive processes.

Coral bleaching is a complex process governed by various physiological and environmental factors (Lesser 2011), suggesting that bleaching tolerance is likely to be controlled by multiple quantitative trait loci (QTLs). Such polygenic traits can be under the influence of a number of small-effect alleles rather than a few large-effect alleles (Koornneef et al. 2004). Therefore, incorporation of allele frequency data for multiple genetic markers would strengthen the model and enhance its capacity to detect tolerant coral populations. However, it should be noted that there is likely to be a trade-off between stress resistance and other fitness-related traits, such as reproduction and growth (Little et al. 2004).

Spatial mapping of alleles at loci that are linked to bleaching-associated responses inform researchers about the potential locations of populations where other oxidative stress genes may also be under selection. Additionally, combinations of sharp gradients in bleaching-associated stressors can be used as indicators of the capacity of corals to deal with the environmental stressors. Such knowledge of allele distributions provides excellent opportunities to search for more loci associated with bleaching tolerance. This makes high-throughput genome-wide scans for other divergent loci more

efficient, as candidate populations or locations that are undergoing divergence in the presence of gene flow can be targeted.

Capturing population-level evolutionary processes in complex ecological systems is inevitably constrained by limitations of time, cost and logistics. BBN is very useful in dealing with small data sets, as the conditional probabilities of a BBN model can be estimated in domains that are fraught with uncertainty because they are based on limited available evidence (i.e. events that have actually been observed). The conditional probabilities of other nodes are linked through the model structure via Bayes law and hence complex associations can be explored in a quantitative manner.

In this study, a BBN maximised available information to make inferences about spatial patterns of bleaching tolerance allele frequencies, by using a combination of long-term monitoring environmental data, allele frequency data, and an expert knowledge-based causal web. A future study could further refine our beliefs in the model as additional observational data are identified and subsequently added to the model to enhance confidence in the correlations.

Chapter 6.0 General discussion

Multiple anthropogenic pressures have caused extensive loss of coral cover in many regions of the world's tropical oceans (e.g. Bruno and Selig 2007; De'ath et al. 2012). The capacity to identify individual corals and populations that are resistant to bleaching, a common sign of environmental stress in corals, is becoming increasingly urgent to provide reef managers with tools that might contribute to maintaining and enhancing the resilience of coral reef ecosystems. In the last decade, considerable research has focused on the endosymbiotic algal partner in the coral holobiont, revealing that differences in the physiology of *Symbiodinium* among clades and types affect the physiology of the coral host (e.g. Little et al. 2004; Berkelmans & van Oppen 2006; Stat et al. 2008; Cantin et al. 2009), as does the density of *Symbiodinium* cells (Cunning and Baker 2012) and their adaptation to local environmental regimes (Howells et al. 2013). In comparison, contributions of the coral host genotype to variation in stress tolerance of the holobiont have been less well studied (but see recent studies by Bay & Palumbi et al. 2014; Dixon et al. 2015). Correlations between allelic diversity at bleaching-associated stress response genes and gradients in temperature and water clarity have suggested candidate loci as Qantitative Trait Loci (QTL) for environmental stress tolerance in corals (Lundgren et al. 2013), but before now, QTLs have not been validated for coral populations. Knowledge of standing coral host genetic diversity and its functional significance is important for understanding the capacity of coral populations to adapt to a rapidly changing climate (Barrett & Schluter 2008).

6.1 Major findings of this thesis and their implications

The primary aim of my PhD research was to investigate whether single nucleotide polymorphisms (SNPs) that correlate with environmental gradients can be

used to predict stress tolerance in corals. The major findings of my research are as follows:

- Allele frequencies in pooled DNA samples can be accurately estimated by a quantitative High Resolution Melting technique, drastically improving the cost and time efficiency of analysing large sample sizes required for estimating QTLs (Chapter 2).
- Allele frequencies at loci C29226S281 and C70S236 in populations of the common coral *Acropora millepora* spread throughout the Great Barrier Reef are correlated with multiple environmental gradients relating to temperature and water quality (Chapter 3).
- Colonies *of A. millepora* carrying G and T alleles at loci C29226S281 and C70S236, respectively, were significantly less affected by natural bleaching events associated with thermal and water quality stress, and also by experimental thermal stress (Chapter 4).
- Both loci showed constitutive differences in antioxidant capacity among genotypes in response to experimental thermal stress (Chapter 4).
- The markers are located on different chromosomes and closely linked to a range of genes involved in bleaching-associated processes, such as thermal, hypoxic and oxidative stress, suggesting they are indicative of tolerance levels to a multitude of environmental perturbations (Chapter 4).
- A Bayesian belief network (BBN) model developed in this thesis predicts that populations of *A. millepora* with high frequencies of the bleaching tolerance allele are found at sites with both poor water quality and large seasonal temperature fluctuations. Spatial mapping of predicted allele frequencies showed that the bleaching tolerance allele was concentrated near the mouths of the Burdekin and Fitzroy Rivers, where nutrient levels are chronically high (Chapter 5).

- The Bayesian simulation approach demonstrates that there is a synergism between high ranges in seasonally-fluctuating sea temperatures and high nitrate concentrations that may be the primary driver in selecting alleles at this locus (Chapter 5).

In summary, this thesis presents the first genetic markers underpinning bleaching-associated stress tolerance in reef-building corals. These markers are closely linked to genes involved in responses to thermal and oxidative stresses. My results provide opportunities for the development of novel approaches for the conservation and restoration of coral reefs, such as spatial mapping of stress tolerance for use in Marine Protected Area design, the identification of stress tolerant colonies for assisted migration, and marker-assisted selective breeding to create more tolerant genotypes for restoration of denuded reefs.

6.2 Further considerations and recommendations arising from thesis results

Gene by environment association analysis is based on the notion that alleles associated with higher fitness are more likely to be passed on to subsequent generations and spread within a population, and that alleles differ in their performance in different environments. Spatial patterns of allele frequencies are commonly influenced by the interacting effects of selective forces, including environmental, genetic, biological and demographic factors, as well as non-selective forces. High levels of variation in environmental factors throughout the Great Barrier Reef, particularly water quality and temperature that act synergistically to cause coral bleaching, make populations of corals spread throughout the GBR good candidates for the detection of QTLs. However, as bleaching tolerance is likely to be a polygenic trait (Seneca et al. 2010; Pinzón et al.

2015), patterns of allele frequencies at loci conferring bleaching tolerance (including loci underpinning oxidative and thermal stress tolerance) are likely to be complex.

Multi-faceted regulation of the bleaching response and loci of small effects are likely to lead to large variation in allele frequencies at bleaching-associated loci among populations exposed to similar environmental conditions. Therefore, gene by environment association studies in complex genetic and environmental systems like the GBR may give rise to spurious findings if sampling schemes do not take these issues into account. Increasing the range of environmental variation investigated along selective gradients, as well as the number of populations along the gradients, such as the sampling of 18 populations completed in this study (Chapter 3), will improve results where allele frequency ranges are large among populations experiencing similar environmental conditions. In support of this recommendation, a simulation study by Mita et al. (2013) indicates that using more populations will give better results than more samples per population when testing correlations between allele frequencies and environmental gradients. Hale et al. (2012) demonstrate that accurate estimation of allele frequencies requires 25-30 samples per population using real and simulated data sets obtained from a range of taxa with different levels of allele diversities and population sizes. However, it is often difficult to individually genotype many samples from a number of populations due to time and financial constraints.

To circumvent problems associated with time and financial constraints, DNA samples from multiple individuals can be pooled to estimate the proportion of alleles in a population. In Chapter 2, I demonstrate the high accuracy of allele frequency estimates that can be obtained using the pooled sample single nucleotide polymorphism (SNP) genotyping method I developed. Tools for accurately estimating allele frequencies

are important for detecting small differences in allele frequencies among populations in different environments.

Because quantitative trait loci controlling polygenic traits like bleaching tolerance may have small effects on a phenotype, their mean allele frequencies are likely to have small ranges among different environments. In Chapter 3, I found that allele frequencies at five loci used for outlier detection analysis vary by 21%, with the two environmentally-correlated loci targeted for marker development having frequencies that differed by 18% and 27% across the range of populations sampled. Consequently, it was important that the method developed had small error rates. In particular, a pooling method that causes high error rates in allele frequency estimation (e.g. ~25% error rates found for the peaks method; Chapter 2) may not detect environmentally-correlated patterns of allele frequencies at QTLs of small effects in a study where the number of populations sampled per environmental level is limited. The curves method I developed in Chapter 2 has <1% error rates, thus is the preferred option when allele frequencies at QTLs are likely to have small ranges.

Collectively, my research is the first to tease apart the effects of acclimatisation versus genetic adaptation in gene by environment associations. Genetic associations with underlying phenotypes are often complicated by acclimatisation that enables organisms to plastically change their phenotypes according to the environmental conditions they experience. Such non-genetic information can be transmitted over some generations (Youngson & Whitelaw 2008). For this reason, Dixon et al. (2015) were not able to exclude the possibility that acclimatisation effects underlay differences in heat tolerance among larvae raised from reciprocal crosses between parental colonies from reefs separated by 5° of latitude. The use of samples from the same environment to examine genotypically-determined quantitative traits, such as the use of SNP markers to

examine the link between genotype and its response to bleaching-associated stressors within a single population (Chapter 4), is a powerful way to separate the effects of acclimatisation versus genetic adaptation in gene by environment associations.

6.3 Future research directions

I found that corals with higher bleaching-associated stress tolerance are characterised by constitutively higher antioxidant capacity and expressions of thermal and oxidative stress response genes (Chapter 4; also see Barshis et al. 2013). However, gains in stress tolerance are often linked to energetic trade-offs with other fitness-related traits, such as reproduction and growth. The presence of a trade-off in phenotypes linked to the marker loci examined in my study is also indicated by the occurrence of both alleles at the two loci in all populations sampled, regardless of the environmental conditions they experience (Chapter 3). Knowledge of such trade-offs would allow managers to take appropriate precautions when using genetic markers for reef conservation and management to avoid fitness reductions in natural populations due to the introduction of maladapted alleles. Given the constitutively small differences in phenotypic response found among genotypes of *A. millepora* (Chapter 4), future studies should investigate genotypic differences in fitness using long-term experiments that involve measuring survival, growth and reproduction under both ambient (control) and moderately stressful environments. Such experiments could be conducted in both aquarium and field settings using juveniles of known genotypes derived from selective crossing of parents from the same environment to exclude acclimatisation effects.

In Chapter 5, I generated a spatial prediction map based on a Bayesian belief network model to predict the spatial distributions of allele frequencies for bleaching tolerance. Such maps can inform researchers and managers about reefs that are potentially susceptible to bleaching events. However, it should be noted that tolerance

alleles can have double-edged effects on fitness: stress tolerance can improve fitness in one environment but decrease it in another. Further studies of trade-offs associated with tolerance alleles would allow such knowledge to be taken into account when prediction maps of coral susceptibility (or tolerance) are generated based on allele frequency data. This would enable conservation efforts to be directed towards rescuing susceptible populations, as well as maximising the fitness of populations. Ideally, the goal of future research would be to identify candidate populations for management interventions based on identifying populations with allele frequencies that are lower than expected relative to predicted environmental conditions so such actions do not compromise optimal fitness achieved by natural populations (Fig. 6.1). These maladapted populations may be rescued by seeding or translocating corals with appropriate alleles. However, if high bleaching tolerance is linked to reduced fitness in a trait that is equally important for survival in future scenarios (e.g. disease resistance), then populations that have higher than expected frequencies of tolerance alleles in benign environments might also be considered for management interventions (Fig. 6.1). This dilemma highlights the need for future research on the roles of trade-offs in fitness to determine allele frequency thresholds for reefs that warrant conservation attention, as well as the need to identify more QTLs associated with bleaching tolerance to improve our capacity to appropriately define these thresholds.

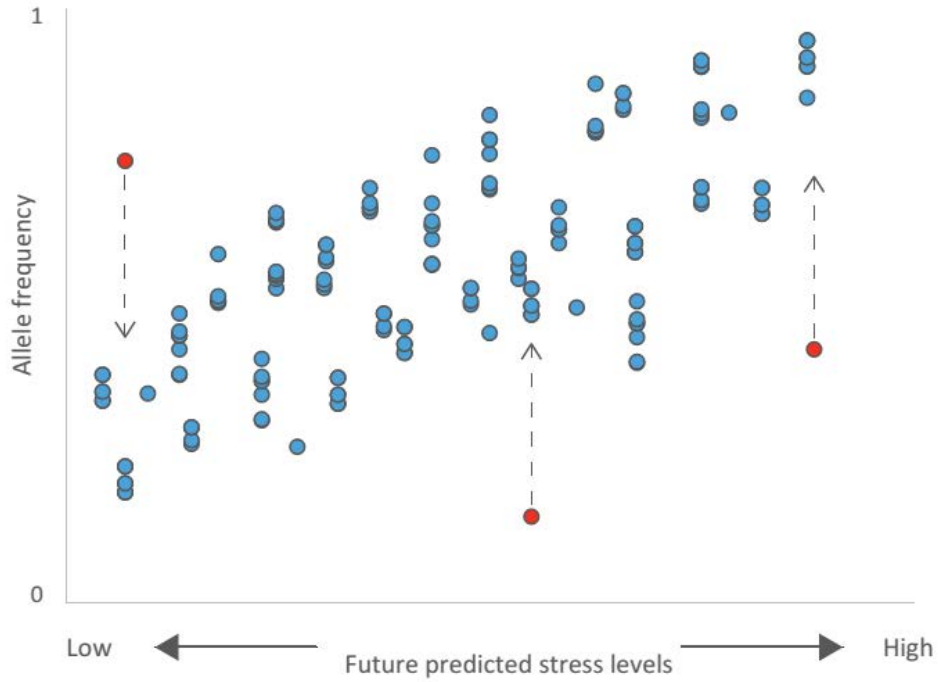

Fig. 6.1 Scatterplot showing hypothetical patterns of allele frequencies at a bleaching tolerance locus in relation to a gradient of future environmental stress. Circles indicate populations with different allele frequencies (Y-axis) along future stress levels (X-axis). Red circles indicate populations with allele frequencies that deviate from the projected pattern.

In addition to genetic diversity of coral hosts, allele frequencies of tolerance alleles are likely to be affected by a number of factors that contribute to variation in tolerance among holobionts, for example *Symbiodinium* type and stress history (Middlebrook et al. 2008). These additional factors may affect spatial variation in bleaching occurrence. Therefore, observational data of bleaching potentially encompass a range of information associated with the bleaching response. In future studies, the addition of information about these factors (i.e. trade-offs, *Symbiodinium* type, stress history and bleaching observations) to a BBN model may further refine the process of identifying populations that are susceptible to bleaching, and hence to long-term adverse effects on fitness and possibly mass mortality. A strength of BBN is its ability to incorporate both qualitative and quantitative information, as well as expert knowledge. For example, the model structure constructed in Chapter 5 is based on prior knowledge of the synergistic effects of temperature and NO_3 on coral bleaching and the putative

functions of genes that were shown to be linked to the QTL markers. Future mapping studies should incorporate further data on linked genes, which would improve the spatial prediction model and enable it to better predict and understand the susceptibility of coral populations to climate change.

In conclusion, my research raises new opportunities for the management of coral reefs in a changing climate. Spatial information on the distribution of tolerance-associated alleles opens up the possibility of new approaches to coral conservation and management. Large-scale mapping of the likelihood of tolerant allele frequency distributions developed in this study allows managers to visually investigate allele distributions across many more populations than could be realistically sampled. Such projections enable managers to target populations with high proportions of bleaching tolerance alleles for conservation. These populations are important as a source of: advantageous alleles that could be spread to affected populations, candidate colonies for transplant programs for reef restoration purposes, and breeding stock for selective breeding programs. Furthermore, knowledge of the distributions of bleaching tolerance alleles enables more efficient design of marine protected areas to ensure connectedness among populations that promote adaptive processes.

References

Abal EG and Dennison WC. (1996) Seagrass depth range and water quality in southern Moreton Bay, Queensland, Australia. *Marine and Freshwater Research* 47, 763-771

Anthony KR and Hoegh-Guldberg O (2003) Kinetics of photoacclimation in corals. *Oecologia* 134: 23-31.

Anthony K, Marshall PA, Abdulla A, et al. (2015) Operationalizing resilience for adaptive coral reef management under global environmental change. *Global Change Biology* 21: 48-61.

Aranda M, Banaszak AT, Bayer T, et al. (2011) Differential sensitivity of coral larvae to natural levels of ultraviolet radiation during the onset of larval competence. *Molecular Ecology* 20: 2955-2972.

Assmann V, Jenkinson D, Marshall JF, et al. (1999) The intracellular hyaluronan receptor RHAMM/IHABP interacts with microtubules and actin filaments. *Journal of Cell Science* 112: 3943-3954.

Baird A and Marshall P. (2002) Mortality, growth and reproduction in scleractinian corals following bleaching on the Great Barrier Reef. *Marine Ecology Progress Series* 237: 133-141.

Balic M, Pichler M, Strutz J, et al. (2009) High Quality Assessment of DNA Methylation in Archival Tissues from Colorectal Cancer Patients Using Quantitative High-Resolution Melting Analysis. *Journal of Molecular Diagnostics* 11: 102-108.

Bansal V, Tewhey R, LeProust EM, et al. (2011) Efficient and Cost Effective Population Resequencing by Pooling and In-Solution Hybridization. *Plos One* 6.

Barrett RDH and Schluter D. (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* 23: 38-44.

Barshis D, Stillman J, Gates R, et al. (2010) Protein expression and genetic structure of the coral Porites lobata in an environmentally extreme Samoan back reef: does host genotype limit phenotypic plasticity? *Molecular Ecology* 19: 1705-1720.

Barshis DJ, Ladner JT, Oliver TA, et al. (2013) Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America* 110: 1387-1392.

Bay RA and Palumbi SR. (2014) Multilocus adaptation associated with heat resistance in reef-building corals. *Current Biology* 24: 2952-2956.

Belles-Boix E, Babiychuk E, Van Montagu M, et al. (2000) CEF, a Sec24 homologue of Arabidopsis thaliana, enhances the survival of yeast under oxidative stress conditions. *Journal of Experimental Botany* 51: 1761-1762.

Berkelmans R, De'ath G, Kininmonth S, et al. (2004) A comparison of the 1998 and 2002 coral bleaching events on the Great Barrier Reef: spatial correlation, patterns, and predictions. *Coral Reefs* 23: 74-83.

Berkelmans R and Van Oppen MJ. (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 of London B: Biological Sciences* 273: 2305-2312.

Berkelmans R and Willis B. (1999) Seasonal and local spatial patterns in the upper thermal limits of corals on the inshore Central Great Barrier Reef. *Coral Reefs* 18: 219-228.

Berteaux D, Reale D, McAdam AG, et al. (2004) Keeping pace with fast climate change: Can arctic life count on evolution? *Integrative and Comparative Biology* 44: 140-151.

Boswell MG, Wells MC, Kirk LM, et al. (2009) Comparison of gene expression responses to hypoxia in viviparous (Xiphophorus) and oviparous (Oryzias) fishes using a medaka microarray. *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology* 149: 258-265.

Bradburd GS, Ralph PL and Coop GM. (2013) DISENTANGLING THE EFFECTS OF GEOGRAPHIC AND ECOLOGICAL ISOLATION ON GENETIC DIFFERENTIATION. *Evolution* 67: 3258-3273.

Bradshaw WE and Holzapfel CM. (2008) Genetic response to rapid climate change: it's seasonal timing that matters. *Molecular Ecology* 17: 157-166.

Brodie J, Devlin M, Haynes D, et al. (2011) Assessment of the eutrophication status of the Great Barrier Reef lagoon (Australia). *Biogeochemistry* 106: 281-302.

Brodie JE, McKergow LA, Prosser IP, et al. (2003) Sources of sediment and nutrient exports to the Great Barrier Reef World Heritage Area.

Brown B and Cossins A. (2011) The Potential for Temperature Acclimatisation of Reef Corals in the Face of Climate Change. In: Dubinsky Z and Stambler N (eds) *Coral Reefs: An Ecosystem in Transition.* Springer Netherlands, 421-433.

Bruno JF and Selig ER. (2007) Regional decline of coral cover in the Indo-Pacific: timing, extent, and subregional comparisons. *Plos One* 2: e711.

Burman JL, Bourbonniere L, Philie J, et al. (2008) Scyl1, mutated in a recessive form of spinocerebellar neurodegeneration, regulates COPI-mediated retrograde traffic. *Journal of Biological Chemistry* 283: 22774-22786.

Butcher LM, Meaburn E, Liu L, et al. (2004) Genotyping pooled DNA on microarrays: A systematic genome screen of thousands of SNPs in large samples to detect QTLs for complex traits. *Behavior Genetics* 34: 549-555.

Campino S, Auburn S, Kivinen K, et al. (2011) Population Genetic Analysis of Plasmodium falciparum Parasites Using a Customized Illumina GoldenGate Genotyping Assay. *Plos One* 6.

Cantin NE, van Oppen MJ, Willis BL, et al. (2009) Juvenile corals can acquire more carbon from high-performance algal symbionts. *Coral Reefs* 28: 405-414.

Carroll SP, Hendry AP, Reznick DN, et al. (2007) Evolution on ecological time-scales. *Functional Ecology* 21: 387-393.

Chazottes V, Le Campion-Alsumard T, Peyrot-Clausade, et al. (2002) The effects of eutrophication-related alterations to coral reef communities on agents and rates of bioerosion (Reunion Islnad, Indian Ocean). *Coral Reefs* 21: 375-390.

Cochat P, Pichault V, Bacchetta J, et al. (2010) Nephrolithiasis related to inborn metabolic diseases. *Pediatric Nephrology* 25: 415-424.

Coles SL and Jokiel PL. (1977) EFFECTS OF TEMPERATURE ON PHOTOSYNTHESIS AND RESPIRATION IN HERMATYPIC CORALS. *Marine Biology* 43: 209-216.

Cook LM and Saccheri IJ. (2013) The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity* 110: 207-212.

Csardi G and Nepusz T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems* 1695: 1-9.

Cunning R and Baker AC. (2013) Excess algal symbionts increase the susceptibility of reef corals to bleaching. *Nature Climate Change* 3: 259-262.

Davey JW, Hohenlohe PA, Etter PD, et al. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12: 499-510.

De'ath G, Fabricius K and Authority GBRMP. (2008) *Water quality of the Great Barrier Reef: distributions, effects on reef biota and trigger values for the protection of ecosystem health*: Great Barrier Reef Marine Park Authority.

De'ath G. (2007) The spatial, temporal and structural composition of water quality of the Great Barrier Reef, and indicators of water quality and mapping risk. *Report to*

the Marine and Tropical Sciences Research Facility. Reef and Rainforest Research Centre Limited, Cairns, Queensland, Australia 71pp (available at http://www. rrrc. org. au/publications/downloads/115-AIMS-Death-2007-WQ-on-GBR. pdf).

De'ath G, Fabricius KE, Sweatman H, et al. (2012) The 27–year decline of coral cover on the Great Barrier Reef and its causes. *Proceedings of the National Academy of Sciences* 109: 17995-17999.

Demokan S, Chuang AY, Chang XF, et al. (2013) Identification of guanine nucleotide-binding protein gamma-7 as an epigenetically silenced gene in head and neck cancer by gene expression profiling. *International Journal of Oncology* 42: 1427-1436.

Dempster AP, Laird NM and Rubin DB. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*: 1-38.

Desalvo MK, Voolstra CR, Sunagawa S, et al. (2008) Differential gene expression during thermal stress and bleaching in the Caribbean coral Montastraea faveolata. *Molecular Ecology* 17: 3952-3971.

Devlin M, Waterhouse J, Taylor J, et al. (2001) *Flood plumes in the Great Barrier Reef: spatial and temporal patterns in composition and distribution*: Great Barrier Reef Marine Park Authority.

Devlin MJ and Brodie J. (2005) Terrestrial discharge into the Great Barrier Reef Lagoon: nutrient behavior in coastal waters. *Marine Pollution Bulletin* 51: 9-22.

Devlin MJ, McKinna LW, Alvarez-Romero JG, et al. (2012) Mapping the pollutants in surface riverine flood plume waters in the Great Barrier Reef, Australia. *Marine Pollution Bulletin* 65: 224-235.

Diaz-Pulido G, McCook LJ, Dove S, et al. (2009) Doom and Boom on a Resilient Reef: Climate Change, Algal Overgrowth and Coral Recovery. *Plos One* 4.

Dimitrakopoulos L, Vorkas PA, Georgoulias V, et al. (2012) A closed-tube methylation-sensitive high resolution melting assay (MS-HRMA) for the semi-quantitative determination of CST6 promoter methylation in clinical samples. *Bmc Cancer* 12.

Dixon GB, Davies SW, Aglyamova GA, et al. (2015) Genomic determinants of coral heat tolerance across latitudes. *Science* 348: 1460-1462.

Doostzadeh J, Shokralla S, Absalan F, et al. (2008) High Throughput Automated Allele Frequency Estimation by Pyrosequencing. *Plos One* 3.

Dunlap WC, Starcevic A, Baranasic D, et al. (2013) KEGG orthology-based annotation of the predicted proteome of Acropora digitifera: ZoophyteBase - an open access and searchable database of a coral genome. *Bmc Genomics* 14.

Dunn SR, Pernice M, Green K, et al. (2012) Thermal Stress Promotes Host Mitochondrial Degradation in Symbiotic Cnidarians: Are the Batteries of the Reef Going to Run Out? *Plos One* 7.

Edmundson J and Mazumder A. (2002) Regional and hierarchical perspectives of thermal regimes in subarctic, Alaskan lakes. *Freshwater Biology* 47: 1-17.

Elliott MR, Zheng SQ, Park D, et al. (2010) Unexpected requirement for ELMO1 in clearance of apoptotic germ cells in vivo. *Nature* 467: 333-U114.

Er TK and Chang JG. (2012) High-resolution melting: Applications in genetic disorders. *Clinica Chimica Acta* 414: 197-201.

Fabricius KE and De'Ath G. (2001a) Environmental factors associated with the spatial distribution of crustose coralline algae on the Great Barrier Reef. *Coral Reefs* 19: 303-309.

Fabricius KE, Wild C, Wolanski E, et al. (2003) Effects of transprent exopolymer particles and muddy terrigenous sediments on the survival of hard coral recruits. *Estuarine, Coastal and Shelf Science* 57: 613-621.

Fabricius KE. (2005) Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Marine Pollution Bulletin* 50: 125-146.

Fabricius KE, Cooper TF, Humphrey C, et al. (2012) A bioindicator system for water quality on inshore coral reefs of the Great Barrier Reef. *Marine Pollution Bulletin* 65: 320-332.

Fabricius KE, Cseke S, Humphrey C, et al. (2013) Does Trophic Status Enhance or Reduce the Thermal Tolerance of Scleractinian Corals? A Review, Experiment and Conceptual Framework. *Plos One* 8.

Feero WG, Guttmacher AE and Manolio TA. (2010) Genomewide association studies and assessment of the risk of disease. *New England Journal of Medicine* 363: 166-176.

Fernandez-Canon JM, Baetscher MW, Finegold M, et al. (2002) Maleylacetoacetate isomerase (MAAI/GSTZ)-deficient mice reveal a glutathione-dependent nonenzymatic bypass in tyrosine catabolism. *Molecular and Cellular Biology* 22: 4943-4951.

Fernandez-Jimenez N, Castellanos-Rubio A, Plaza-Izurieta L, et al. (2011) Accuracy in copy number calling by qPCR and PRT: a matter of DNA. *Plos One* 6: e28910.

Ferrier-Pages C, Schoelzke V, Jaubert J, et al. (2001) Response of a scleractinian coral, Stylophora pistillata, to iron and nitrate enrichment. *Journal of Experimental Marine Biology and Ecology* 259: 249-261.

Filzmoser P and Todorov V. (2013) Robust tools for the imperfect world. *Information Sciences* 245: 4-20.

Folke C, Carpenter S, Walker B, et al. (2004) Regime shifts, resilience, and biodiversity in ecosystem management. *Annual Review of Ecology, Evolution, and Systematics*: 557-581.

Foll M and Gaggiotti O. (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 180: 977-993.

Fraser C, Hanage WP and Spratt BG. (2007) Recombination and the nature of bacterial speciation. *Science* 315: 476-480.

Friedlander AM and Parrish JD. (1998) Habitat characteristics affecting fish assemblages on a Hawaiian coral reef. *Journal of Experimental Marine Biology and Ecology* 224: 1-30.

Frieler K, Meinshausen M, Golly A, et al. (2013) Limiting global warming to 2 [thinsp][deg] C is unlikely to save most coral reefs. *Nature Climate Change* 3: 165-170.

Furnas M, Alongi D, McKinnon D, et al. (2011) Regional-scale nitrogen and phosphorus budgets for the northern (14 S) and central (17 S) Great Barrier Reef shelf ecosystem. *Continental Shelf Research* 31: 1967-1990.

Furnas M, Mitchell A, Skuza M, et al. (2005) In the other 90%: phytoplankton responses to enhanced nutrient availability in the Great Barrier Reef Lagoon. *Marine Pollution Bulletin* 51: 253-265.

Gagnaire PA, Normandeau E, Cote C, et al. (2012) The Genetic Consequences of Spatially Varying Selection in the Panmictic American Eel (Anguilla rostrata). *Genetics* 190: 725-U703.

Gaits F, Shiozaki K and Russell P. (1997) Protein phosphatase 2C acts independently of stress-activated kinase cascade to regulate the stress response in fission yeast. *Journal of Biological Chemistry* 272: 17873-17879.

Gala A, Fang Y, Woltedji D, et al. (2013) Changes of proteome and phosphoproteome trigger embryo-larva transition of honeybee worker (Apis mellifera ligustica). *Journal of Proteomics* 78: 428-446.

Ganopoulos I, Sakaridis I, Argiriou A, et al. (2013) A novel closed-tube method based on high resolution melting (HRM) analysis for authenticity testing and quantitative detection in Greek PDO Feta cheese. *Food Chemistry* 141: 835-840.

Garvin MR, Saitoh K and Gharrett AJ. (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* 10: 915-934.

Gates RD, Baghdasarian G and Muscatine L. (1992) Temperature stress causes host cell detachment in symbiotic cnidarians: implications for coral bleaching. *The Biological Bulletin* 182: 324-332.

Gautier M, Foucaud J, Gharbi K, et al. (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology* 22: 3766-3779.

Germer S, Holland MJ and Higuchi R. (2000) High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Research* 10: 258-266.

Golbuu Y, Victor S, Wolanski E, et al. (2003) Trapping of fine sedimetn in a semi-closed bay, Palau, Micronesia. *Estuarine, Coastal and Shelf Science* 57: 941-949.

Graham NA, McClanahan TR, MacNeil MA, et al. (2008) Climate warming, marine protected areas and the ocean-scale integrity of coral reef ecosystems. *Plos One* 3: e3039.

Gratwicke B and Speight M. (2005) The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. *Journal of Fish Biology* 66: 650-667.

Gray JC and Goddard MR. (2012) Gene-flow between niches facilitates local adaptation in sexual populations. *Ecology Letters* 15: 955-962.

Gruber JD, Colligan PB and Wolford JK. (2002) Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing. *Human Genetics* 110: 395-401.

Guerin R, Arseneault G, Dumont S, et al. (2008) Calnexin Is Involved in Apoptosis Induced by Endoplasmic Reticulum Stress in the Fission Yeast. *Molecular Biology of the Cell* 19: 4404-4420.

Guest JR, Baird AH, Maynard JA, et al. (2012) Contrasting Patterns of Coral Bleaching Susceptibility in 2010 Suggest an Adaptive Response to Thermal Stress. *Plos One* 7.

Guinotte J, Buddemeier R and Kleypas J. (2003) Future coral reef habitat marginality: temporal and spatial effects of climate change in the Pacific basin. *Coral Reefs* 22: 551-558.

Gundry CN, Dobrowolski SF, Martin YR, et al. (2008) Base-pair neutral homozygotes can be discriminated by calibrated high-resolution melting of small amplicons. *Nucleic Acids Research* 36: 3401-3408.

Hale ML, Burg TM and Steeves TE. (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies.

Halekoh U, Hojsgaard S and Yan J. (2006) The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software* 15: 1-11.

Hallas R, Schiffer M and Hoffmann AA. (2002) Clinal variation in Drosophila serrata for stress resistance and body size. *Genetical Research* 79: 141-148.

Harvell D, Kim K, Quirolo C, et al. (2001) Coral bleaching and disease: contributors to 1998 mass mortality in Briareum asbestinum (Octocorallia, Gorgonacea). *The Ecology and Etiology of Newly Emerging Marine Diseases.* Springer, 97-104.

Heibo E, Magnhagen C and Vollestad LA. (2005) Latitudinal variation in life-history traits in Eurasian perch. *Ecology* 86: 3377-3386.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, et al. (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* 11: 123-136.

Hendry AP, Taylor EB and Benkman C. (2004) How much of the variation in adaptive divergence can be explained by gene flow? An evaluation using lake-stream stickleback pairs. *Evolution* 58: 2319-2331.

Hilbish T. (1985) Demographic and temporal structure of an allele frequency cline in the mussel Mytilus edulis. *Marine Biology* 86: 163-171.

Hilbish TJ and Koehn RK. (1985) The physiological basis of natural selection at the LAP locus. *Evolution*: 1302-1317.

Hilbish TJ and Koehn RK. (1987) The adaptive importance of genetic variation. *American Scientist*: 134-141.

Hirayama H and Suzuki T. (2014) GDP-Mannose Pyrophosphorylase A,B (GMPPA,B). In: Taniguchi N, Honke K, Fukuda M, et al. (eds) *Handbook of Glycosyltransferases and Related Genes.* Springer Japan, 1599-1606.

Hoegh-Guldberg O. (1999) Climate change, coral bleaching and the future of the world's coral reefs. *Marine and freshwater research* 50: 839-866.

Hoegh-Guldberg O, Mumby P, Hooten A, et al. (2007) Coral reefs under rapid climate change and ocean acidification. *Science* 318: 1737-1742.

Hoeghguldberg O and Smith GJ. (1989) INFLUENCE OF THE POPULATION-DENSITY OF ZOOXANTHELLAE AND SUPPLY OF AMMONIUM ON THE BIOMASS AND METABOLIC CHARACTERISTICS OF THE REEF CORALS SERIATOPORA-HYSTRIX AND STYLOPHORA-PISTILLATA. *Marine Ecology Progress Series* 57: 173-186.

Holland PM, Abramson RD, Watson R, et al. (1991) DETECTION OF SPECIFIC POLYMERASE CHAIN-REACTION PRODUCT BY UTILIZING THE 5'- 3' EXONUCLEASE ACTIVITY OF THERMUS-AQUATICUS DNA-POLYMERASE. *Proceedings of the National Academy of Sciences of the United States of America* 88: 7276-7280.

Holmes KE (2000) Effects of eutrophication on bioeroding spongen communtieis with the description of new West Indian sponges *Cliona* spp. (Porifera: Hadromerida: Clionidae). *Invertebrate Biology* 119: 125-138.

Howells EJ, Berkelmans R, van Oppen MJH, et al. (2013) Historical thermal regimes define limits to coral acclimatization. *Ecology* 94: 1078-1088.

Hughes AO, Olley JM, Croke JC, et al. (2009) Sediment source changes over the last 250 years in a dry-tropical catchment, central Queensland, Australia. *Geomorphology* 104: 262-275.

Hughes TP, Baird AH, Bellwood DR, et al. (2003) Climate change, human impacts, and the resilience of coral reefs. *Science* 301: 929-933.

Humphrey C, Weber M, Lott C, et al. (2008) Effects of suspended sediments, dissolved inorganic nutrients and salinity on fertilisation and embryo development in the coral Acropora millepora (Ehrenberg, 1834). *Coral Reefs* 27: 837-850.

Hunter CL. (1993) Genotypic variation and clonal structure in coral populations with different disturbance histories. *Evolution*: 1213-1228.

Johnson MT, FitzJohn RG, Smith SD, et al. (2011) Loss of sexual recombination and segregation is associated with increased diversification in evening primroses. *Evolution* 65: 3230-3240.

Johnston SE, Lindqvist M, Niemela E, et al. (2013) Fish scales and SNP chips: SNP genotyping and allele frequency estimation in individual and pooled DNA from historical samples of Atlantic salmon (Salmo salar). *Bmc Genomics* 14.

Jones A and Berkelmans R. (2010a) Marine'Refugia'in the Keppel Region of the Great Barrier Reef.

Jones AA and Berkelmans RR. (2010b) Keppel Islands reefs: Baseline study 2008-2010. A report prepared for the Fitzroy Basin Association.

Jones AM, Berkelmans R, van Oppen MJ, et al. (2008) A community change in the algal endosymbionts of a scleractinian coral following a natural bleaching event: field evidence of acclimatization. *Proceedings of the Royal Society of London B: Biological Sciences* 275: 1359-1365.

Jouaville LS, Pjnton P, Bastianutto C, et al. (1999) Regulation of mitochondrial ATP synthesis by calcium: Evidence for a long-term metabolic priming. *Proceedings of the National Academy of Sciences of the United States of America* 96: 13807-13812.

Jump AS and Penuelas J. (2005) Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters* 8: 1010-1020.

Kendrick GA (1991) Recruitment of coralline crusts and filamentous turf algae in the Galapagos archipelago: effect of simulated scour, erosion and accretion. *Journal of Experimental Marine Biology and Ecology* 147: 47-63.

Koornneef M, Alonso-Blanco C and Vreugdenhil D. (2004) Naturally occurring genetic variation in Arabidopsis thaliana. *Annu. Rev. Plant Biol.* 55: 141-172.

Langerhans RB, Layman CA, Langerhans AK, et al. (2003) Habitat-associated morphological divergence in two Neotropical fish species. *Biological Journal of the Linnean Society* 80: 689-698.

Larcombe P, Ridd P, Prytz A, et al. (1995) Factors controlling suspended sediment on inner-shelf coral reefs, Townsville, Australia. *Coral Reefs* 14: 163-171.

Lasfargues C, Martineau Y, Bousquet C, et al. (2013) Changes in Translational Control after Pro-Apoptotic Stress. *International Journal of Molecular Sciences* 14.

Lavebratt C and Sengul S. (2006) Single nucleotide polymorphism (SNP) allele frequency estimation in DNA pools using Pyrosequencing (TM). *Nature Protocols* 1: 2573-2582.

Le Corre V and Kremer A. (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* 164: 1205-1219.

Lesser M. (2011) Coral Bleaching: Causes and Mechanisms. In: Dubinsky Z and Stambler N (eds) *Coral Reefs: An Ecosystem in Transition.* Springer Netherlands, 405-419.

Lesser MP. (2006) Oxidative stress in marine environments: biochemistry and physiological ecology. *Annu. Rev. Physiol.* 68: 253-278.

Lie TJ, Costa KC, Lupa B, et al. (2012) Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 109: 15473-15478.

Liew M, Pryor R, Palais R, et al. (2004) Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical Chemistry* 50: 1156-1164.

Lin XB and Gänzle MG. (2014) Quantitative high-resolution melting PCR analysis for monitoring of fermentation microbiota in sourdough. *International Journal of Food Microbiology* 186: 42-48.

Little AF, Van Oppen MJ and Willis BL. (2004) Flexibility in algal endosymbioses shapes growth in reef corals. *Science* 304: 1492-1494.

Loeb VJ, Hofmann EE, Klinck JM, et al. (2009) ENSO and variability of the Antarctic Peninsula pelagic marine ecosystem. *Antarctic Science* 21: 135-148.

Lough J. (1994) Climate variation and El Nino-Southern Oscillation events on the Great Barrier Reef: 1958 to 1987. *Coral Reefs* 13: 181-185.

Lough JM. (2007) Tropical river flow and rainfall reconstructions from coral luminescence: Great Barrier Reef, Australia. *Paleoceanography* 22.

Lundgren P, Vera JC, Peplow L, et al. (2013) Genotype - environment correlations in corals from the Great Barrier Reef. *Bmc Genetics* 14.

Lutz A, Motti CA, Freckelton ML, et al. (2014) Simultaneous determination of coenzyme Q and plastoquinone redox states in the coral–Symbiodinium symbiosis during thermally induced bleaching. *Journal of Experimental Marine Biology and Ecology* 455: 1-6.

Lynch M, Conery J and Burger R. (1995) Mutation accumulation and the extinction of small populations. *American Naturalist*: 489-518.

Lynn BD, Turley EA and Nagy JI. (2001) Subcellular distribution, calmodulin interaction, and mitochondrial association of the hyaluronan-binding protein RHAMM in rat brain. *Journal of Neuroscience Research* 65: 6-16.

Ma C and Staudt LM. (1996) LAF-4 encodes a lymphoid nuclear protein with transactivation potential that is homologous to AF-4, the gene fused to MLL in t(4;11) leukemias. *Blood* 87: 734-745.

Mader E, Ruzicka J, Schmiderer C, et al. (2011) Quantitative high-resolution melting analysis for detecting adulterations. *Analytical Biochemistry* 409: 153-155.

Malentacchi F, Forni G, Vinci S, et al. (2009) Quantitative evaluation of DNA methylation by optimization of a differential-high resolution melt analysis protocol. *Nucleic Acids Research* 37.

Marubini F and Davies P. (1996) Nitrate increases zooxanthellae population density and reduces skeletogenesis in corals. *Marine Biology* 127: 319-328.

Marubini F and Atkinson MJ. (1999) Effects of lowered pH and elevated nitrate on coral calcification. *Marine Ecology Progress Series* 188: 117-121.

Marubini F and Thake B. (1999) Bicarbonate addition promotes coral growth. *Limnology and Oceanography* 44: 716-720.

Mayer MP and Bukau B. (2005) Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences* 62: 670-684.

Mazor M, Alkrinawi S, Chalifa-Caspi V, et al. (2011) Primary Ciliary Dyskinesia Caused by Homozygous Mutation in DNAL1, Encoding Dynein Light Chain 1. *American Journal of Human Genetics* 88: 599-607.

McCulloch M, Fallon S, Wyndham T, et al. (2003) Coral record of increased sediment flux to the inner Great Barrier Reef since European settlement. *Nature* 421: 727-730.

McKergow LA, Prosser IP, Hughes AO, et al. (2005) Sources of sediment to the Great Barrier Reef World Heritage Area. *Marine Pollution Bulletin* 51: 200-211.

Meyer E, Aglyamova G and Matz M. (2011) Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Molecular Ecology* 20: 3599-3616.

Meyer E, Davies S, Wang S, et al. (2009) Genetic variation in responses to a settlement cue and elevated temperature in the reef-building coral Acropora millepora. *Mar Ecol Prog Ser* 392: 81-92.

Middlebrook R, Hoegh-Guldberg O and Leggat W. (2008) The effect of thermal history on the susceptibility of reef-building corals to thermal stress. *Journal of Experimental Biology* 211: 1050-1056.

Miller J, Muller E, Rogers C, et al. (2009) Coral disease following massive bleaching in 2005 causes 60% decline in coral cover on reefs in the US Virgin Islands. *Coral Reefs* 28: 925-937.

Mita S, Thuillet AC, Gay L, et al. (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* 22: 1383-1399.

Mitchell-Olds T, Willis JH and Goldstein DB. (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* 8: 845-856.

Morin PA, Martien KK and Taylor BL. (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* 9: 66-73.

Mullen LM and Hoekstra HE. (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution* 62: 1555-1570.

Mumby PJ, Hastings A and Edwards HJ. (2007) Thresholds and the resilience of Caribbean coral reefs. *Nature* 450: 98-101.

Munder S, Kasbauer T, Prexl A, et al. (2010) Notch signalling defines critical boundary during budding in Hydra. *Developmental Biology* 344: 331-345.

Muscatine L, Falkowski PG, Dubinsky Z, et al. (1989) THE EFFECT OF EXTERNAL NUTRIENT RESOURCES ON THE POPULATION-DYNAMICS OF ZOOXANTHELLAE IN A REEF CORAL. *Proceedings of the Royal Society Series B-Biological Sciences* 236: 311-324.

Ning W, Chu TJ, Li CJ, et al. (2004) Genome-wide analysis of the endothelial transcriptome under short-term chronic hypoxia. *Physiological Genomics* 18: 70-78.

Niyogi KK. (1999) Photoprotection revisited: genetic and molecular approaches. *Annual review of plant biology* 50: 333-359.

Nordemar I, Nystrom M and Dizon R. (2003) Effects of elevated seawater temperature and nitrate enrichment on the branching coral Porites cylindrica in the absence of particulate food. *Marine Biology* 142: 669-677.

Nornes S, Newman M, Wells S, et al. (2009) Independent and cooperative action of Psen2 with Psen1 in zebrafish embryos. *Experimental Cell Research* 315: 2791-2801.

Norton N, Williams NM, Williams HJ, et al. (2002) Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Human Genetics* 110: 471-478.

Noseda M, Fu YX, Niessen K, et al. (2006) Smooth muscle alpha-actin is a direct target of Notch/CSL. *Circulation Research* 98: 1468-1470.

O'connor MI, Piehler MF, Leech DM, et al. (2009) Warming and resource availability shift food web structure and metabolism. *PLoS Biol* 7: e1000178.

Opperdoes FR and Coombs GH. (2007) Metabolism of Leishmania: proven and predicted. *Trends in Parasitology* 23: 149-158.

Paaby AB, Blacket MJ, Hoffmann AA, et al. (2010) Identification of a candidate adaptive polymorphism for Drosophila life history by parallel independent clines on two continents. *Molecular Ecology* 19: 760-774.

Pachauri RK, Allen M, Barros V, et al. (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Packett R, Dougall C, Rohde K, et al. (2009) Agricultural lands are hot-spots for annual runoff polluting the southern Great Barrier Reef lagoon. *Marine Pollution Bulletin* 58: 976-986.

Palais RA, Liew MA and Wittwer CT. (2005) Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. *Analytical Biochemistry* 346: 167-175.

Pandolfi JM, Bradbury RH, Sala E, et al. (2003) Global trajectories of the long-term decline of coral reef ecosystems. *Science* 301: 955-958.

Peterson G, Allen CR and Holling CS. (1998) Ecological resilience, biodiversity, and scale. *Ecosystems* 1: 6-18.

Pfaffl MW. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 29.

Pinzón JH, Kamel B, Burge CA, et al. (2015) Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. *Royal Society Open Science* 2: 140214.

Polato NR, Voolstra CR, Schnetzer J, et al. (2010) Loation-specific responses to thermal stress in larvae of the reef-building coral *Montastraea faveolata*. *Plos One* 5

Polato NR, Naomi SA, Iliana BB. (2013) Variation in the transcriptional response of threatened coral larvae to elevated temperatures. *Molecular ecology* 22: 1366-1382.

Polovina JJ, Howell EA and Abecassis M. (2008) Ocean's least productive waters are expanding. *Geophysical Research Letters* 35.

Pratchett MS, McCowan D, Maynard JA, et al. (2013) Changes in Bleaching Susceptibility among Corals Subject to Ocean Warming and Recurrent Bleaching in Moorea, French Polynesia. *Plos One* 8.

Pratchett MS, Munday PL, Wilson SK, et al. (2008) Effects of climate-induced coral bleaching on coral-reef fishes - Ecological and economic consequences. In: Gibson RN, Atkinson RJA and Gordon JDM (eds) *Oceanography and Marine Biology: An Annual Review, Vol 46.* 251-296.

Puklowski A, Homsi Y, Keller D, et al. (2011) The SCF-FBXW5 E3-ubiquitin ligase is regulated by PLK4 and targets HsSAS-6 to control centrosome duplication. *Nature Cell Biology* 13: 1004-U1291.

Rao RV, Ellerby HM and Bredesen DE. (2004) Coupling endoplasmic reticulum stress to the cell death program. *Cell Death and Differentiation* 11: 372-380.

Räsänen K and Hendry AP. (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters* 11: 624-636.

Rayner N, Parker DE, Horton E, et al. (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres (1984–2012)* 108.

Renaut S, Nolte AW, Rogers SM, et al. (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (Coregonus spp.). *Molecular Ecology* 20: 545-559.

Riegl B and Branch GM. (1995) Effects of sedimetn on the energy budgets of four scleractinian (Bourne 1900) and five alcyonacean (Lamouroux 1816) corals. *Journal of Experimental Marine Biology and Ecology* 186: 259-275.

Rogers CS. (1979) The effect of shading on coral reef structure and function. *Journal of Experimental Marine Biology and Ecology* 41: 269-288.

Saier MH. (2000) Families of transmembrane sugar transport proteins. *Molecular Microbiology* 35: 699-710.

Sakaridis I, Ganopoulos I, Argiriou A, et al. (2013a) A fast and accurate method for controlling the correct labeling of products containing buffalo meat using High Resolution Melting (HRM) analysis. *Meat Science* 94: 84-88.

Sakaridis I, Ganopoulos I, Argiriou A, et al. (2013b) High resolution melting analysis for quantitative detection of bovine milk in pure water buffalo mozzarella and other buffalo dairy products. *International Dairy Journal* 28: 32-35.

Schaffelke B, Thompson A, Carleton J, et al. (2009) Reef Rescue Marine Monitoring Program-Final Report of AIMS Activities 2008/09. *Report to the Reef and Rainforest Research Centre. Australian Institute of Marine Science, Townsville* 146.

Schemske DW and Bierzychudek P. (2007) Spatial differentiation for flower color in the desert annual Linanthus parryae: was Wright right? *Evolution* 61: 2528-2543.

Schloder C and D'Croz L. (2004) Responses of massive and branching coral species to the combined effects of water temperature and nitrate enrichment. *Journal of Experimental Marine Biology and Ecology* 313: 255-268.

Schmidt PS, Serrao EA, Pearson GA, et al. (2008) ECOLOGICAL GENETICS IN THE NORTH ATLANTIC: ENVIRONMENTAL GRADIENTS AND ADAPTATION AT SPECIFIC LOCI. *Ecology* 89: S91-S107.

Schneider CJ, Smith TB, Larison B, et al. (1999) A test of alternative models of diversification in tropical rainforests: ecological gradients vs. rainforest refugia. *Proceedings of the National Academy of Sciences* 96: 13869-13873.

Schweighofer A, Hirt H and Meskiene L. (2004) Plant PP2C phosphatases: emerging functions in stress signaling. *Trends in Plant Science* 9: 236-243.

Seneca FO, Forêt S, Ball EE, et al. (2010) Patterns of gene expression in a scleractinian coral undergoing natural bleaching. *Marine Biotechnology* 12: 594-604.

Serrano R, Gaxiola R, Rios G, et al. (2003) Salt stress proteins identified by a functional approach in yeast. *Monatshefte Fur Chemie* 134: 1445-1464.

Sezgin E, Duvernell DD, Matzkin LM, et al. (2004) Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in Drosophila melanogaster. *Genetics* 168: 923-931.

Sham P, Bader JS, Craig I, et al. (2002) DNA pooling: A tool for large-scale association studies. *Nature Reviews Genetics* 3: 862-871.

Shick J, Lesser M, Dunlap W, et al. (1995) Depth-dependent responses to solar ultraviolet radiation and oxidative stress in the zooxanthellate coral Acropora microphthalma. *Marine Biology* 122: 41-51.

Shick JM and Dykens JA. (1985) OXYGEN DETOXIFICATION IN ALGAL-INVERTEBRATE SYMBIOSES FROM THE GREAT BARRIER-REEF. *Oecologia* 66: 33-41.

Shinzato C, Shoguchi E, Kawashima T, et al. (2011) Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* 476: 320-U382.

Siegel PM and Massague J. (2003) Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer. *Nature Reviews Cancer* 3: 807-820.

Singh KK. (2004) Mitochondria damage checkpoint in apoptosis and genome stability. *Fems Yeast Research* 5: 127-132.

Smith CJ, Nedwell DB, Dong LF, et al. (2007a) Diversity and Abundance of Nitrate Reductase Genes (narG and napA), Nitrite Reductase Genes (nirS and nrfA), and Their Transcripts in Estuarine Sediments. *Applied and Environmental Microbiology* 73: 3612-3622.

Smith L, Barshis D and Birkeland C. (2007b) Phenotypic plasticity for skeletal growth, density and calcification of Porites lobata in response to habitat type. *Coral Reefs* 26: 559-567.

Smith TB, Mila B, Grether GF, et al. (2008) Evolutionary consequences of human disturbance in a rainforest bird species from Central Africa. *Molecular Ecology* 17: 58-71.

Smith TB, Wayne RK, Girman DJ, et al. (1997) A role for ecotones in generating rainforest biodiversity. *Science* 276: 1855-1857.

Souter P, Willis BL, Bay LK, et al. (2010) Location and disturbance affect population genetic structure in four coral species of the genus Acropora on the Great Barrier Reef. *Marine Ecology Progress Series* 416: 35-45.

Stanek MT, Cooper TF and Lenski RE. (2009) Identification and dynamics of a beneficial mutation in a long-term evolution experiment with Escherichia coli. *BMC evolutionary biology* 9: 302.

Stat M, Carter D and Hoegh-Guldberg O. (2006) The evolutionary history of Symbiodinium and scleractinian hosts—symbiosis, diversity, and the effect of climate change. *Perspectives in Plant Ecology, Evolution and Systematics* 8: 23-43.

Staus DP, Blaker AL, Taylor JM, et al. (2007) Diaphanous 1 and 2 regulate smooth muscle cell differentiation by activating the myocardin-related transcription factors. *Arteriosclerosis Thrombosis and Vascular Biology* 27: 478-486.

Steiner CC, Römpler H, Boettger LM, et al. (2009) The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Molecular Biology and Evolution* 26: 35-45.

Stinchcombe JR and Hoekstra HE. (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100: 158-170.

Stockwell CA, Hendry AP and Kinnison MT. (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution* 18: 94-101.

Stumpf MPH and McVean GAT. (2003) Estimating recombination rates from population-genetic data. *Nature Reviews Genetics* 4: 959-968.

Stuopelyte K, Daniunaite K, Laurinaviciene A, et al. (2013) High-Resolution Melting-Based Quantitative Analysis of RASSF1 Methylation in Breast Cancer. *Medicina-Lithuania* 49: 78-83.

Szmant A and Gassman N. (1990) The effects of prolonged "bleaching" on the tissue biomass and reproduction of the reef coral Montastrea annularis. *Coral Reefs* 8: 217-224.

Telesnicki GJ and Goldberg WM (1996) Effects of turbidity on the photosynthesis and respiration of two south Floridat reef coral species. *Bulletin of Marine Science* 57: 527-539

Thomas CD. (2011) Translocation of species, climate change, and the end of trying to recreate past ecological communities. *Trends in Ecology & Evolution* 26: 216-221.

Timmermann A, Oberhuber J, Bacher A, et al. (1999) Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature* 398: 694-697.

Tishkoff SA, Reed FA, Ranciaro A, et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* 39: 31-40.

Toggweiler J and Russell J. (2008) Ocean circulation in a warming climate. *Nature* 451: 286-288.

Tonne JM, Sakuma T, Deeds MC, et al. (2013) Global gene expression profiling of pancreatic islets in mice during streptozotocin-induced beta-cell damage and pancreatic Glp-1 gene therapy. *Disease Models & Mechanisms* 6: 1236-1245.

Trench R. (1979) The cell biology of plant-animal symbiosis. *Annual Review of Plant Physiology* 30: 485-531.

Turunen M, Olsson J and Dallner G. (2004) Metabolism and function of coenzyme Q. *Biochimica Et Biophysica Acta-Biomembranes* 1660: 171-199.

Umina PA, Weeks AR, Kearney MR, et al. (2005) A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. *Science* 308: 691-693.

Vaillant E, Vatin V, Boutin P, et al. (2002) Rapid SNP allele frequency determination in genomic DNA pools by Pyrosequencing (TM). *BioTechniques* 32: 1138-1142.

Valerius MT, Li H, Stock JL, et al. (1995) GSH-1 - A NOVEL MURINE HOMEOBOX GENE EXPRESSED IN THE CENTRAL-NERVOUS-SYSTEM. *Developmental Dynamics* 203: 337-351.

van Oppen MJH, Oliver JK, Putnam HM, et al. (2015) Building coral reef resilience through assisted evolution. *Proceedings of the National Academy of Sciences of the United States of America* 112: 2307-2313.

van Oppen MJH, Peplow LM, Kininmonth S, et al. (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: 4899-4914.

van Oppen MJH, Puill-Stephan E, Lundgren P, et al. (2014) First-generation fitness consequences of interpopulational hybridisation in a Great Barrier Reef coral and its implications for assisted migration management. *Coral Reefs* 33: 607-611.

Van Woesik R and Authority GBRMP. (1991) *Immediate Impact of the January 1991 floods on the Coral Assemblages of the Keppel Islands*: Great Barrier Reef Marine Park Authority.

Vega Thurber RL, Burkepile DE, Fuchs C, et al. (2014) Chronic nutrient enrichment increases prevalence and severity of coral disease and bleaching. *Global Change Biology* 20: 544-554.

Voss JD and Richardson LL. (2006) Nutrient enrichment enhances black band disease progression in corals. *Coral Reefs* 25: 569-576.

Wang AL, Fu MC, Jiang XQ, et al. (2014) Evolution of the F-Box Gene Family in Euarchontoglires: Gene Number Variation and Selection Patterns. *Plos One* 9.

Wang S, Zhang LL, Meyer E, et al. (2009) Construction of a high-resolution genetic linkage map and comparative genome analysis for the reef-building coral Acropora millepora. *Genome Biology* 10.

Weiss A and Leinwand LA. (1996) The mammalian myosin heavy chain gene family. *Annual Review of Cell and Developmental Biology* 12: 417-439.

Werner M, Sych M, Herbon N, et al. (2002) Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. *Human Mutation* 20: 57-64.

Wiedenmann J, D'Angelo C, Smith EG, et al. (2013) Nutrient enrichment can increase the susceptibility of reef corals to bleaching. *Nature Climate Change* 3: 160-164.

Wilson SK, Graham NAJ, Pratchett MS, et al. (2006) Multiple disturbances and the global degradation of coral reefs: are reef fishes at risk or resilient? *Global Change Biology* 12: 2220-2234.

Wittwer CT, Reed GH, Gundry CN, et al. (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical Chemistry* 49: 853-860.

Wooldridge S, Brodie J and Furnas M. (2006) Exposure of inner-shelf reefs to nutrient enriched runoff entering the Great Barrier Reef Lagoon: Post-European changes and the design of water quality targets. *Marine Pollution Bulletin* 52: 1467-1479.

Wooldridge S and Done T. (2004) Learning to predict large-scale coral bleaching from past events: A Bayesian approach using remotely sensed data, in-situ data, and environmental proxies. *Coral Reefs* 23: 96-108.

Wooldridge SA. (2009) Water quality and coral bleaching thresholds: Formalising the linkage for the inshore reefs of the Great Barrier Reef, Australia. *Marine Pollution Bulletin* 58: 745-751.

Wooldridge SA, Brodie JE, Kroon FJ, et al. (2015) Ecologically based targets for bioavailable (reactive) nitrogen discharge from the drainage basins of the Wet Tropics region, Great Barrier Reef. *Marine Pollution Bulletin*.

Wu LIN and Kruckeberg AL. (1985) COPPER TOLERANCE IN TWO LEGUME SPECIES FROM A COPPER MINE HABITAT. *New Phytologist* 99: 565-570.

Yan S, Zheng C, Chen ZQ, et al. (2012) Expression of Endoplasmic Reticulum Stress-Related Factors in the Retinas of Diabetic Rats. *Experimental Diabetes Research*.

Youngson NA and Whitelaw E. (2008) Transgenerational epigenetic effects. *Annu. Rev. Genomics Hum. Genet.* 9: 233-257.

Zhao RM and Houry WA. (2005) Hsp90: a chaperone for protein folding and gene regulation. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* 83: 703-710.

Zhen Y and Ungerer MC. (2008) Clinal variation in freezing tolerance among natural accessions of Arabidopsis thaliana. *New Phytologist* 177: 419-427.

Zhou LM, Myers AN, Vandersteen JG, et al. (2004) Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. *Clinical Chemistry* 50: 1328-1335.

Appendices

Supplementary S.1. R scripts for peak height finder

```
library(IDPmisc)
library(gtools)
library(pastecs)

par(mfrow=c(2,1))
par(mar=c(2, 3, 1, 3))

file<-("normalized_fluorescence_melt_curves_file_name.txt")  #this file comes off the Roche machine but the y-axis has been normalized within each reaction.  See example file for reference.
a <- c(4)  # a = which SNP/column number to use; only makes sense when this value is even

df<-read.table(file,header=TRUE)
colhead<-names(df)

#identify x,y coords of snp of interest
coli<-colhead[a]
xsub<-df[,a-1]
ysub<-df[,a]


#identify x, y coords of snp of interest
#identify coord of amplicon peak and slice probe window based on it
# note, maybe you should not slice if you want to see the whole window including calibrators...  but for titration exps, everything is in the same plate -> temp shifting is reduced
dfsub<-data.frame(cbind(ysub,xsub))
maxy<-1
maxx<-dfsub[ysub==maxy,2]  #Identify the xcoord of max y
dfsub<-dfsub[(xsub>55) & (xsub<maxx-3),] #slice out data points below 55deg and above 4 degrees below the amplicon peak, both columns
xsub<-dfsub[,2]
ysub<-dfsub[,1]

#Now the snp of interest (even values of a) are defined as xsub, ysub; dfsub holds the probe peaks range of data for the snp of interest
#identify local maxima of, minima bounding probe peaks using pastecs::turnpoint
y_turns<-turnpoints(ysub)
y_peaks<-summary(y_turns)$peaks
y_pits<-summary(y_turns)$pits
tops <- which(y_peaks, arr.ind=TRUE)
bottoms <- which(y_pits, arr.ind=TRUE)
# coords of all peaks: ysub[tops], xsub[tops]; coords of all pits: ysub[bottoms], xsub[bottoms]

#pick two (three) probe peaks and two minima bounding the probe peaks from the potentially many identified:
####***************####
t<-tops[c(2,3)]
b<-bottoms[c(1,2)]
t<-c(24,31)
b<-c(1,47)
```

```
####***************####

#find topline (tline) and background line (bline)
tline<-lm(ysub[t]~xsub[t])
bline<-lm(ysub[b]~xsub[b])

#subtract the background from the peaks:
bkg<-xsub[t]*bline$coefficients[2]+bline$coefficients[1]  #finds the spot on the bottom
line where the top of the peak is located.  y= slope of bline*coord of top + intercept of
bline
diff<-ysub[t]-bkg           # note, this is not the hypotenuse that you are seeking; that is
under 'dist' later on

#finding equation of the line(s) perpendicular to bline:
oldm<-as.numeric(bline$coefficients[2])  #oldm = old 'm' in y=mx+b
oldb<-as.numeric(bline$coefficients[1])  #oldb = old 'b' in y=mx+b
newm<-(-1/(oldm))  #slope of perpendicular line (new 'm')
newb<-ysub[t]-(newm*xsub[t])  #b = y-mx

#find coords of the line intersection/s with peaks:
interx<-as.numeric((oldb-newb)/(newm-oldm))
intery<-as.numeric((oldm*interx)+oldb)

#find distance between peak tops and the line between bounding minima (pythag.
theorem) (find the hypotenuse)
dist=sqrt((ysub[t]-intery)^2+(xsub[t]-interx)^2)

plot(ysub~xsub, type="b", ylim=c(0, 1), xlab="temp", ylab="-(d/dt) fluorescence",
main=colhead[a])
points(ysub[t]~xsub[t],col=2,pch=20)      # peaks
points(ysub[b]~xsub[b],col=3,pch=20)    # pits
points(bkg~xsub[t],pch=19,col="darkgreen")
text(round(dist,4),x=xsub[t],y=ysub[t]*1.2)
abline(tline, lty=2, col=2)  #top line
abline(bline, lty=2, col=3)  #bottom line
segments(interx, intery, x1=xsub[t],y1=ysub[t],col="slateblue",lwd=2,lty=3)


# copy and paste this into the console to print out the peak heights:
#stuff=c(title,round(interx,4),round(dist,4))
#line<-matrix(data=stuff,nrow=1,byrow=FALSE)
#write(line,file="peak_heights_titration_snp22162_29Apr10.txt",sep="\t",ncolumns=8,a
ppend=TRUE)
```

Supplementary Table S.2. 384 SNP loci from the coral linkage map selected from Wang et al. 2009.

SNP_name	SNP_number	Forward Primer	Reverse Primer	Probe	Gene Name
C60613S230	1	AGTATGTTGCTGAGAATGTTGCT	ATCAAAAGCATCCTCTACACCAT	ACAGGAATCACCcGAGCTGATGACTCT	Coatomer
C16774S791	2	CTGGAAGTTCATGCTTAGCCTT	AATGTTGTGCCGTAGTGTTCGA	CCTGCACAAACAGAcGTTAACACCTTTTTCA	Mn superoxide dismutase
C50281S478	3	GGAAGATGCAGGTGTTTACGAT	CCATTATACATGTCTTGGACACA	GTAACCGCAGaCGCTTGGGCCAC	Arginine kinase
C52394S280	4	TGTCCTTTAGTGGTGTTGATGA	GACCCTACAGGATGAATTGGA	GGTGAAATAAATCCACGaTCAAATTTCATTCCTTCTT	Hsp60
C63538S709	5	CAATGCAGCATCAACTGCATCT	GGACGGTAACAAAGACATCCAA	GGATCCTCGGcCATCGCAAGAAA	Ligand of numbX2
C3255S483	6	GCCAAACGGGCCAAGCAAGT	ACAAAGTGACCATTGAATTGCCA	GCCCAGCCGAAATCtCTTTCTTTTCATACAG	Sclerostin_domain-containing_protein_1_precursor_(SOSTDC1)
C7134S210	7	AATGATTCACCAATTAAGCCTCTT	GAATCCACCCAACGATGGAGT	CCAACGTGCACCcTCCAGAAACAAGGG	Dynein_heavy_chain_domain-containing_protein_2_(DNHD2)
C15620S247	8	CATAACATTGCTCCACGGAAGA	ATAAACAAAATGGGAGGAAATCGA	AAAGCTCGGTGCAtTGCTTACTAACAAGA	Transforming_acidic_coiled-coil-containing_protein_(Tacc1)
C18443S396	9	CATGTTATCCACACGATGGACA	ATGGCTGCCAAAGCACAGCT	CTTCACCTTCGCaCGGGAGTAACTGAC	Nucleolar_protein_5A_(NOL5A)
C20479S292	10	GTGTAAATCACTATGCGTGCTT	AGAAATCCATTTTCATGCCACAT	CTGGTCACTAGGGGTtGTCATAAGAGATTTATC	Circadian_locomoter_output_cycles_protein_kaput_(CLOCK)
C21618S209	11	CTTCTGTGCGATGGGGTGTT	GATGAGTCGCACGATATGCAA	GAAACAGAAAGATACcAGCAAAAGGATCGAATT	Guanine_nucleotide-binding_protein_G(s)_subunit_al_(G-s-alpha-60A)
C25628S525	12	TTAGAGGCCGACAATGTTGGA	TTACCCTAAGCCCGCTTCCT	GGTGTGCATTTGGTgTTGAATACACGACCAC	Amiloride-sensitive_cation_channel_2-C,_neuronal_(accn2c)
C28868S363	13	AGACGCTCATCACGGCCATA	CAAGATTGGCTCATTGTACACA	TCACAATGCTCTgCAGGAGAATGTGCG	Tumor_protein_p63-regulated_gene_1_protein_(TPRG1)
C29060S309	14	CAAAATTCCCACAAGGAGCTGA	CATCTCCTTCCTTGCACAACT	AACCCCATACAGCgTACAAGTCCTGAGCT	SH2_domain-containing_protein_5_(SH2D5)
C35020S147	15	GATGAGGTTTATGTATGATCACA	CAAAGTACTTATTCACCTCAGCT	CTGAATGGCCAAAaCCTTTCCAGCCTAGG	Transmembrane_emp24_domain-containing_protein_3_pr_(Tmed3)
C45199S349	16	CTAATGTTACATGGAGAATTTGCA	AGGCGACAGAGACACAATCGT	CATTCTTGAGTTCAcTTTATCTCGCCTCCCG	Achaete-scute_homolog_5_(ASCL5)
C49658S304	17	AATGGAGTTCCTACAATCGTGT	GCAGGCGATAGTGTAGGAGA	GCAGATCTACTAcTCACCTCCGCCTAG	RNA-binding_protein_24_(rbm24)
C55644S292	18	TGGTACTTGCGAGCCACTGA	CTTTGGACGAGGCTATCGACT	TGAAGATGGCCGTcCCATTTCCATAAGCC	Probable_methylmalonate-semialdehyde_dehydrogenase
C59049S135	19	GAAGAACCGAACGTCAACTACT	TTTAAAGCGATGTGTCAAAGTGA	AGAAACCCATTGcCGGCGTTAAACAGT	Tubulin_polyglutamylase_TTLL7_(TTLL7)
C70S236	20	GTTAAGTGACCCTTGCTACCA	GACATGCATATGCTCCAACTCT	GTACTTTTATGACGcGCTAAATGCATGACTT	0
C76S562	21	GTTGTTGGCTATTGCCACCTT	CGACGTTGTTCATAATACCTCTT	CTTGGCATTGCTgATATCGCAAAGAAA	0
C166S563	22	TCAAGCACTGTTCCACCACAT	AAGAATCCGTTGAGGGAGCAA	GACGTCCACTGGTTcACTTTGTTGTTGTTGCA	0
C188S318	23	AATTCACATGGTCTCTTCCAGT	CCCTGAAGGTAGAGTTGAGCT	GTCAACTTGGCTgCTAATGGAAGCTCC	0
C237S473	24	TCAAGCCTTGCTATTGAACTGA	CCAATTGAAATGAGTGGGTGAA	CATACTTTGAATACAGgAAAGGAAATCAAGATCTA	0
C288S173	25	TTCATGCTGTCCATAAGCAAGA	CAACTTCCTCACCTTGCACAA	GCTTTATAAAATTTTgCCCAAAAGCCTGCTCAA	0
C490S693	26	AACTAAAAGTAGAGGCATTCCTA	GGTCTCAAACTAAACACGTACA	GAAGGGTGAAAGAGgTAGAATGAACGACCGA	0
C841S459	27	TACGGCTTCTTGTGACGTACA	AAATGTTTGATGAGGTTCAGACA	TTTCTTCCACATCgTCTGTTTCAGGAGCC	0
C915S149	28	GAGAAAAATGCTTTCCTGGTTGT	CGTACTGGCTGACACTTCCT	AAGAGGACTGGGaACAATACGCCGTTG	0
C969S127	29	TGTTTCCTCCAGTTCTGAACAT	TGGTATTTTGTTGCTTAGAAGCT	TGCAGATCAAGTTGaGTGTGCTGTTTATTAC	0
C986S247	30	TTCTAAGGAGGGAACAGCGAA	AAGGGCTTAAGCTAATGCATCT	TAAAAATGAGCTGTTAaTTAGGGCCTTTATACCGA	0
C989S461	31	CAACATGTACGACAATGATGGA	CATCTTGTAGATAGCTCCCACA	ATGGCCTTATCACcAAGGACGAGATGTAC	0
C1023S218	32	GTCTTACATTCACATGTAGAGTA	TCTGCCAGCGCGTGGTGTGT	GTCTTCTGTGTGaGCGACACACACGGC	0
C1024S157	33	CACAAGGAAATATTACTCGTCCA	ATGTGTAGCTTAATTAGGACACT	GCGACATCGAAATAAAaTATTATGAACACTAGAAG	0
C1063S181	34	TATAAACCGAGAGTGACCTCAA	CACCGCAACCTATACAACACA	AAGTTGTGAAGTTAAcCACGATTAGCTCTTCAG	0
C1114S124	35	CACAGGACAGGAAGTGTGGT	GTGACTGGATCTCGTGGACT	TACGACATAATCACTaACACTACCGAGGATACA	0
C1136S272	36	CAGTGTCCTGTGCACCAGAT	ATATGGTTCTCCGCTGGAACA	GGTAATTGAGGATgAGTGGCACAGTGAGG	0
C1328S290	37	TAGTTGTCACCAGTGCAGGAT	ACAAGGAAATGGTCCCAGTGT	CTTTGCCACTCgCTTTGGCATTCTA	0
C1405S258	38	GTTGAAGAACTCTGTTTGTGGTT	GTCTCTAAAAGCAATGGGGACA	GTTTAGCAATGGAAaTCAGAAGTCAGGGTCT	0
C5239S208	39	GTGAAAACTAAACTACAAGAGTGA	TAAAGGAACCTCGTTTAAAGTGA	CAGCTTTCATGTgCGATGCTGACATTG	H/ACA_ribonucleoprotein_complex_subunit_1_(nola1)
C9608S288	40	CAGTGAAGGTCCTCTGGTGA	TGAACGAACTTGCAAAAGCGAA	TCCGCTTTCCTCgGTGACTTCGTATAC	Double-stranded_RNA-binding_protein_Staufen_homolo_(stau2)
C14259S283	41	TGTTTATCTGTGACTTTGCTCTA	CAGATGTTAGGGCAGACTACT	GAGCTCTTGCTgTTGGTCGCTCTAA	Vesicle_transport_protein_USE1_(use1)
C19944S225	42	TCTACGACGTCATTGAACAAGT	GAGGCGTTGTACGTCCTGAA	GAGCTGATGGAgGGTCTCACCGCGT	DNA_topoisomerase_1_(TOP1)
C23209S177	43	AGGCTACGTAACCAAGCTCTT	GCCTGGTGACTACATCTGTGT	TGGCCCACAACTcCTGGCACTATTGCT	Rhamnosyl_O-methyltransferase_precursor
C26852S307	44	TCTGAAGATGGTTGATGACTGT	AATCGTCTTTCGTAGAATTTCGT	GCTTTATGGAATGGaTAGATGGTTCCGATCC	Lactoylglutathione_lyase_(GLY I)

C28447S501	45	CAGGTAAAGCCTTTTGGAGGTT	TCCAGTTTGACGCGGACAGA	TCACAGTCACATCaGTAATCGGCTCCATT	C-myc_promoter-binding_protein_(DENND4A)
C29226S281	46	CACAGGAATAAGAACTTGAGACA	CAGTGCAAAGGTGGTGGTAGA	GGTCCTGGTTCTTTaTTTGCAGGTGAAGGCA	Vacuolar_protein-sorting-associated_protein_36_(vps36)
C29463S468	47	GTTGACGGAAGGCGTCAAGT	CGGAACGTGACTAAACAACACA	AGGAGATAAGCATTcTACATGACGAGATGTT	Low_density_lipoprotein_receptor_adapter_protein_1_(ldlrap1-A)
C45133S676	48	ACAGGTGTTTGGGCTTTGTCA	CGCAAACTAAACTGTCATATCCA	AAACGCTGTTCACcTCTTTTGAGCTATGT	WD_repeat-containing_protein_53_(WDR53)
C52176S400	49	TGTCCCGTCTCGAGTGACGA	GCGCATATTCTGTGCTGATGT	CTAAGGGTGATTcCGCTCTCATCGGGC	Extracellular_protease_precursor_(eprA1)
C2435S173	50	TGGGGATTCGCCGCCAAGCA	CACGAAGTTATAAAAGGCGTCTA	GTCTCACCCACAcCAGACAAAGCGCGC	Protein_XRP2_(rp2)
C6659S249	51	TCCCTGCGCCGATACGTCTT	GAAGGCAGACGAATGGGTCAA	ACATTTGCCTCCCAaTAAATCAGAAACTGTG	Alanyl-tRNA_synthetase,_cytoplasmic_(AARS)
C6723S318	52	GTGTTGAGTCAACTAACGAGAAT	TCTTGTTCTTCCTCTTCTCCAT	GGACATGCTGTTaGGGCTTTGGGAATA	Presenilin-2_(psen2)
C13550S341	53	CTTGGTCATTTTGTAAACCCAGT	GAAGGATGGAGCGTCCTTCA	CCCACAGCAATAAGgTATCGCATAATTGACA	Protein_phosphatase_1_regulatory_subunit_12B_(Ppp1r12b)
C21470S842	54	CCTGATCCTTGAACAAATCAAGT	TTGAGGTCGTTTCTCAGAAGCT	CGCTCCTTTTAGTTTtACATTGCTCTTTCCTAC	Solute_carrier_family_25_member_44_(Slc25a44)
C22162S248	55	GGCATTACGCTTATGGGAACT	GTCCTTGTGATGGTCCAGGT	GATCTATCTGGAAAGAaAAATCATCGGGTTTTATT	Amiloride-sensitive_sodium_channel_subunit_gamma-2_(scnn1g-B)
C22405S305	56	GGGTGTTTTGCTAGATGTCACA	GGATCAATTGATCTCCAGCTGT	GATGTACCTCATTCGgACATTTAGAACTAGATT	Transmembrane_protein_72_(Tmem72)
C27925S129	57	CCATGGTTTGAGCGCCTTGA	TTTGCACCACGCCTGAGTGA	GCAGCCCCAGgCTTTGTCCGACA	Glutathione_transferase_omega-1_(Gsto1)
C29080S200	58	TTCATTCCAACTGCGATAAGCT	CTCGTTTGTCTTGGACACTGA	AGGCAGATGCAACaAGTTTAATTGGCGCA	Kelch-like_protein_7_(Klhl7)
C29432S370	59	TTACCGCGCGCATGCCCATT	CTGATGATGTTCTTGTAATTTGCT	GGACACCAAGCGaCGTGAACCTTGGAA	Neutral_alpha-glucosidase_AB_precursor_(Ganab)
C30854S314	60	CACGGATGACGTTGCTTAGGA	CTTTCACTGTTGACAGAAGAGA	CGCGTTTGAGATTCcCCACTTCAAATTTATC	Chondroitin_sulfate_N-acetylgalactosaminyltransfer_(Csgalnact2)
C36218S165	61	TCTTCAGCCACAGCGCAGTA	CCTATTCTGTGGTTTGTGATGT	AGAGGTCTGGGCaCCACCATTGGTGCA	Retinol_dehydrogenase_13_(Rdh13)
C38503S228	62	CCTTGGCATTTGTTAAGTCTCT	TCCAAGATATCAAACAGGCTGA	ATCAGTGGTGCCGTaACATTGACACTAAACC	NADP-dependent_L-serine/L-allo-threonine_dehydroge_(ydfG)
C49448S110	63	GTAATCTGCTCAAGAAAGCTGTT	GGTCATATTTCAGACTGGATGT	GCCCACCATTCTGTtAACAAAACAGGTACAA	Zinc_finger_CCCH_domain-containing_protein_14_(Zc3h14)
C55647S531	64	AGTCAGAGGTGGTCTGCACA	ACTGTTGCGTGATCCACAAGT	AAATTTGTGTTCCAGgGGATGAGGAACTTTAAT	DNA-directed_RNA_polymerase_II_subunit_RPB11_(Polr2j)
C63883S448	65	ATCACAACAAAGCTCAGGCAAT	AACTTTGAGGTTGACTGGTTCA	GCAGAGTCTCTCtCAAGATGCTTTCT	WD_repeat-containing_protein_47_(Wdr47)
C3724S507	66	TACAAGTTCTTAATCTGGGCTTA	ATGACATGCATGAGACAGAACA	CAAAGGTGATTGTaTCAGCTGCCTCTTTA	F-box/LRR-repeat_protein_14_(FBXL14)
C6250S141	67	CATCCTTCACTGGGGCTTTCT	GAAAACCAAAGGTGCCACTGT	CTATCATTGCATCgGCGCTGTAATTTTGG	Zinc_finger_CCHC_domain-containing_protein_9_(ZCCHC9)
C6267S266	68	GCTGGGGACCTCCTTGTAGA	CCAAGAAGAAGAAGTGGTCCA	TATCATAAGTAGCTTTgTCAAAGAGAACCAAGTAC	40S_ribosomal_protein_S25_(RPS25)
C11461S560	69	TGATTCTTGTGGCGACATGCA	GACGACTGTCGAAGTGAGAGT	ATTCAGCAGTTAAaCCGCGACATGTTCTG	Bardet-Biedl_syndrome_7_protein_homolog_(Bbs7)
C19797S331	70	CAAGCAATGCTGCCCGTTACA	AAATTTCACATCCTCAAAAGCGT	TTATTTGGCCTCgATCCCTGGGATTGT	ELMO_domain-containing_protein_2_(ELMOD2)
C27071S243	71	TCAAATCTGTTGAATGTGGATCA	CCATTTCTAGCTCTACAAAGGA	CGCATGCATTTTGaAAGATGTGGTCTCAA	MICAL-like_protein_2_(MICALL2)
C27153S258	72	TCGATTTTCTTCGCTAGCACTT	CGAGAAACAACTCGAAGATCCT	TTCAGGGTCCTGaCCTCCCAGAAGAGC	Coiled-coil_domain-containing_protein_146_(CCDC146)
C28595S225	73	GACAAAAGAGTTTGGTAGTGACA	CAGAATCAGAAAGCAAATCATCA	GCCATTTCAACCAcGAGTCTTCTCTTTAC	Activating_signal_cointegrator_1_complex_subunit_3_(ASCC3)
C43885S203	74	GTGAAGGATATCTGCCAGTCT	GGCTGAGTGCCATTGCAGCA	AGATCCTCGGTTGAcTTGTCTTCAAGTTCCC	Chromodomain-helicase-DNA-binding_protein_2_(CHD2)
C52436S128	75	CCAGGATATAATTCAATAGGTCT	GGCCAGGCTCAAGTTGTTCA	CATTTGCACACTGgATAGCTCCATCCAAG	Gamma-secretase_subunit_PEN-2_(psenen)
C63602S197	76	TCACCAGTATTGCCGAGATCA	TTGCATATGCGGTGGGAAAGA	ACACAAGCACCAcCAAACCATCAGCTT	Tolloid-like_protein_1_precursor_(tll1)
C27337S504	77	CCTGGTTACATTAATCCACTTGA	GATAGTTGTTGGAACAATCTTGA	AAAGTAAGTGGAGCgTTAATCATGTCACTCT	Endoplasmic_reticulum-Golgi_intermediate_compartme_(ergic2)
C48806S304	78	TGCATAGTGTAGGTATCAAGTGT	TTCCATTGGGCGATTTCCACAT	ACAAAGGTCCCTGTtGCTTAAGAGTAACTCG	Isocitrate_dehydrogenase_[NADP]_cytoplasmic_(IDH1)
C1166S172	79	TGAGTAACATCCTCCCCTTCA	CTTGGAAATGCTTTCCTTTCACA	GTCACAAAGCAAAATTCcAAATAAACAAGTTTACAGA	0
C1379S679	80	GCACCAGAAATGGCAGCAGT	ATTCAATGGCTGTTATACCAACT	TAGGTTTGGTTTGAACgTACTCAATCAGATAACTA	0
C1419S315	81	ACAGCTATTTCGTCTGCATGT	CCTTTAGTGTCATCACGGAAT	CCGGCCACGCAGtTAAACTTGTTGGT	0
C1723S422	82	CATCATTATGCTTGGTTGTTACT	TTTGGTGTTGGTTTCTACTCTT	CGGTCAGCAACAAGaAATGAAGAGTAGATT	0
C2348S700	83	CTTTATATCATGCTCATCCTTAT	AAGGTTATGTATGAGAAAGAAGT	GCTTTCATCTTATCgATTTTTGCTGATTGCA	0
C2365S347	84	ACTGGGATACTCCTGTGGAA	TTCACTTTTACGAGATGTACTTA	GTGGTAATGTCACCgTTACGAAACTTGGCT	0
C3633S408	85	GCGTTCTTCGAAATTCTTGCT	AATCGGGTAGGCGAGGTGTT	GTCTGTTGGTAGTcGGAAGATAGCACACT	Synaptotagmin-10_(Syt10)
C4134S257	86	GCTGGACATTTCTGCAGAAT	AGCGGTCACATGCAAATGGT	AAGATACACGAAAGaACTCACACAGGAGTT	Zinc_finger_protein_554_(ZNF554)
C8085S432	87	GAATGTTTGGAAGTTCGTTAAT	GATGAAGTATTCGGACCTTCT	CGCTCCTCTGGAGaATCTTAAAGTTCAG	DNA-directed_RNA_polymerase_I_subunit_RPA43_(twistnb)
C16965S252	88	TGTGTCCCACACATTACAGTT	GAATAACGTCAGTAGCTTTGT	CCCGCCAATAgCAGACGGCCAGA	Guanine_nucleotide-binding_protein_G(I)/G(S)/G(T)__(gnb1)
C17287S307	89	CTGAAATCGTGCCACAGTCT	TAGAGTTGCTGTCATATATGGA	ATTCTAGCCACAAcTCCCACGTGACCTT	Pumilio_domain-containing_protein_C14orf21_(C14orf21)
C26329S310	90	CAGTACGCAACATCACAAGAAT	AAGATTAAGATGCCCAGGAAAA	AATGCTTTGTCCTGgACAGACACTATCACT	Papilin_precursor_(Ppn)
C31340S160	91	GTTGCCACTTCCATCAGTAT	TGAATGCTCGCTGGCTTTCT	CTAATTCATAATCCTCcTCATTCATCCACTCAGA	SWI/SNF-related_matrix-associated_actin-dependent__(Smarcc2)
C34124S511	92	GTGGGAAAGGTGATTCCTGAA	CCACAGACACATCAGGAACT	TGGCAAACTCACgGGTATGGCATTGG	Glyceraldehyde-3-phosphate_dehydrogenase_(GPD)
C40003S97	93	GAGCTGTCCATCCAGTTCTAT	CAACAGCACTGCCATGTGAA	ACTGTTCACAACTcAGTGCCTGAAGGCA	Integrator_complex_subunit_10_(ints10)
C45380S826	94	TGACACCAGCCAAACTTGCT	TAAAGGTATGAGGATATCAAAGA	CAGTGATGGGTACTCCTgATCAGAATACACTA	Protein_FAM92A1_(FAM92A1)
C45851S374	95	AAGTAAAGCTTTCTTTGTTCCA	GCATTCACTCATGGCCAACAT	TGGACGTGCCAAAcACTGAACTTAGTAA	SPRY_domain-containing_SOCS_box_protein_3_(SPSB3)
C50909S225	96	CAAAACACACAGCAACGAATCA	CAGTTTGTTTCTGGTTGGATAT	AGTGACGAACGCTTaCAAGTGTAATGTTAC	DnaJ_homolog_subfamily_C_member_21_(dnajc21)

C10005S217	97	CAGATGTCGTTCGA ACACAACA	GAAAAAGACAAATATG GACGTCA	CTCCTTGACTTGcTTATCGCT TCCCAA	0
C10050S780	98	TGAAAGCCTGTTCT GACATGTT	AGAGGATTGCTGCACT CTACA	CATCCCTCAACTGAtGAGTG AGTATAACTAC	0
C10466S190	99	CCAACAACCATAAC ATCTTGTGT	TCATCAACTCCCAATA CAGGAA	GCAAAGCCTGAcGTGCTGG GAGAAA	0
C10475S502	100	GCTATCTCGATCCTT CGAGGT	CATACAGTACAAGACA GCTGGT	GGCAAAGGTATCTGTTgTAA TGAAATTCCTTCTGG	0
C10565S307	101	AATAGCTGATCCGA CGTGACA	GCCTATAAGAAATGGC GAGCAA	TGAGATTAGGTAGCcTTGTT TCATGATCTAA	0
C10625S161	102	GCATCAGGATCTGG AATCTTCT	ATCCCCCTAAAGAGAT TGATGA	TCATCCCAGTCTTCtGGTTTC TTATCATGAC	0
C10697S175	103	GCATGCCATCTGGA ACTATGA	GGTGGCTGTCTGTATG GGTT	GCCGCGCAGCAaCAGTCAC AAAGTA	0
C10773S305	104	GTTCCTTCTTCGAAA TTCCATCT	ATAACCTGCGGTTTCG ATCCA	CTCGTGGTCGCCtAATTCTA CTGCACT	0
C10810S897	105	TTTCAAACATGCCCT TGACCTT	TGTAGGTGAAGCTGAT GGTGA	ATCATTTGAGCTcGCGGAGG TTCACAA	0
C10862S253	106	GTCCACAAATGTTAT GATTGGAA	GAAGCTGCAGGATGTC TGGAT	CGCACAAACCGAcCTCCTTT AACAGTT	0
C10890S256	107	CACTATCTGAGACA CACTGCAA	ATCAAGAGTGCATTTG ATCACTT	CCCATTGACACTGAcGTTAT GAATGGAGAGT	0
C10924S223	108	GTTGACAAGTCATTT CCTCAAGT	GTCCCATCTGGAAGCC TTCA	GGCAGTCCAGTTGtTTCGTA GGATGTGTC	0
C11020S415	109	CATGACGTCCAACT CCAACCA	ATTCCAGTAAAATGCC TATCAGT	CAGTGCAATACTTcTCGTCA AATGCTTGT	0
C11040S312	110	TTGCTTAAACTTCTC GATTCGAT	CAATCGCCGCTGTTCT AGTCA	TCTGCCTGCTCGtAGAAGAG CATTAGG	0
C11099S398	111	GCATACTCTAATCTC GTGAACA	AACAAACCTTTCATATC GACACA	TGAATTTTGAATCGTTcCGAT AAATTTCACCTTAA	0
C11110S247	112	CTCGCATATGGGAA CCTACCT	TGGGCAACAATGTGTG TCCTT	GGCTCCCGGGTAgGAATTGT GGAACCT	0
C11242S364	113	GAAAGATGATGGCT TTGAAGGA	GCTCAATGATGCGCAT GAGCT	GACAAAGATTTATCACAgTAT AACACTGAACTGAATC	0
C11329S180	114	GTGAACACATCTAC ACCTTGGA	CTTATGTAGTGTCACA TTGACCT	AATGTTCACGATAgGGCGTG AAAAGAGCT	0
C11422S292	115	CACGTGATTTGCAA TTTCTCCTT	CAGTCGCGCAAACGC CATGA	AGCCAACAGTTCcATGAACC CGTTTTA	0
C11439S315	116	TATTGTTCCTCAGCT TCCTCTA	GTCAGACATTTGAACG GGTCAT	TCTTTCACCAAGGAtTTGATG ATGGCTTCAA	0
C11463S192	117	AGGCATATTTGAAC GACATACTA	CGAAAAAATCCTGTAC AAACTCTT	CTTGTAAATTATTTcATCGCA GAGTCGAGTT	0
C11470S398	118	CAGGATAAACACCA CCTGACA	GTTTAGCCGTTGACCA TGAGA	GCTTTTCCAAGAATgACGAT ATCTGGACGGT	0
C11520S633	119	TTTAAGAATGCTTCT TCGACTGA	TTACGTTTCTTTGTTGC ACATCT	TTGGACTCCAATGAaAGGGA GTTCATCCTCA	0
C11524S150	120	TTGACAGTGATTCA GTTTCTGTT	CTTCATTCTCAGATATT GGAGTT	GGATGGCGGGTcGAAAAGA CACCAT	0
C11535S517	121	TGGTCTTGACTAAA GGCATGTT	CAGCCATTCAGCTCCA CAGTA	GCCACGTTCCTCCaTTAACT TCTAGTGAA	0
C11638S270	122	CAGGTTTTCTCCGA GGCGTTA	CATGAAAACGGAGGAC GAAGT	CACCGAAAAACGACcCTCGT TTTTGATTGCA	0
C11670S169	123	TGTGAACCTAAAAG TGATGTAGA	GTTCTTCCACCATACA GGATGA	CACAGGCCAACcAGACCACC GATCT	0
C11715S299	124	TGGCATTTTCGCGA GGAGCA	AATATAAGCCCGTACA TTGATGA	TGTGGTTTCTGGAcCATGCA TTTGCAGGT	0
C11759S946	125	CTTTAACGAGGCTG TGTCGTA	AATCATCGGGCCTACT GGCA	GAACTCCTTTCGTTGGgTAA TTATGTTTTTGAAAG	0
C11797S545	126	CTGAGATGATTCCC CAAGAGT	GTGGGGTTTTCATCTT CTTGGT	CCTCACTGCTTTGTtCTAAAA GAATCCATCA	0
C11959S269	127	CAAAAATTGATGGC CAGCAACA	TTCAAGGGAATCCTGT ATCAGT	CTTTTGCTGGTCCaTTACAA GTGTGGCAT	0
C11999S90	128	TTTTTAGAAACAGTG ACAATCTCA	CACAGGTCTTCTATTA GGTCTT	ATCAACTATATTTTTTAcCCT CTTCACTGTTTCCTGT	0
C12093S318	129	TCAATGGTGTGGAA CGCAACT	GCAACAAAACGTCATC TTGCATA	GCGGCAGTTCCGtAAGATGT CATGTAT	0
C12097S324	130	CTTGATATCATGTTT GACCTCAT	GGAATTGAAAATGGCA ACCAAGA	TTGAGAAACGTCAgTTACAG CAGAAGTAC	0
C12118S364	131	ACTGGAAAATTACT CTGGAAGTT	CGCAGAACGAGTCAC CTTGT	AAGGAATACATTGGAtCGATT CAACTCAAGCTT	0
C12140S96	132	GGATTTGTGACCAA AGATGATGT	TCTGTTCCTTCCATGG CAAGA	ATCCAGCAGAGATAaAATGC GAAAAATTGAA	0
C12174S605	133	CAAATTTATCTCTTG CTGACGAA	TAAGGTCCTTGTCAAG TCCAAA	ACACGACACTCTtCTCTGAA GGCTCCT	0
C12216S415	134	GCTTTCCATTTTACT GACAGTAAT	TGGTCCTGCTACCTCT GTGT	ATGAAGTACATTAAAcTGATC AGCAGGAAAACT	0
C12219S331	135	AGATTTTCCCTGATC TCCTTGT	ACTTAGAAGATAGAGC CAAAAGT	ATCGGATCCTTCTTGaTTTTG TTCAAAAGAGGC	0
C12260S193	136	TGACAACCGTAGAG TACTGCT	GCTGAGGCGCTGTAAT TTTCA	CAGCAAGGACATGcTTTACT TCAGGATCT	0
C18064S518	137	AGCATCCACATCTC TGACCAT	CGGTGTTGTTATGAGG AGACT	TTTTCCAACTCCTTaGGTGAT GGATTCTGGG	0
C12395S564	138	GATGGCGACTTGTT CACGTCA	CCAGAGAACATCAAAG GAAGTA	CTATTACGAAGAcTTCGAGC TGCTGCC	0
C12464S260	139	TCACCTATTCTGCTT CAAGCAT	GCAACATTCTCTCTTG AAGCAA	TCACTTTCTTAGCcGAATGG GGAGATACT	0
C12479S421	140	TGAGCTGCCACTCC TCAGTT	GCCTGTACGTTTGACC AAGAT	CCCCCTGATCGTCgAATGGT ATATTGTCG	0
C12507S635	141	TACAGCAATATTCTG TGTGACAA	GCAGTGATACCAGTGC ACTCT	ACTTTGTGTTCACAgTGGGG TGTAAGATGTC	0
C12550S536	142	TCGATCGCTGGTGT GCTGGA	GCTTGCACATTACTGA ACATGA	AAAAATATGTTCGATATaGTT TCGTTGTAAAGATTGG	0
C12606S270	143	CATCCATCTGTGGG ACAGCTT	CCAGGAGTGTCACAG CTTCA	TCCATTCCATCTTGGaAAAT CGGTTTTATAGAG	0
C12677S188	144	GGTAAGCACATTAT TATGGACAA	TATCATCCAGCCATTC GCCAA	GCACCAATTCACCtTAATCGA GCTGGACT	0
C12729S314	145	AGCCTGCCATCCAA GCCTGT	TCATTTTCCTGGCCAT CGTCT	CGCTGATCAATGTaCAACGT CATTACGTG	0
C12902S674	146	AGCCGCTGTTTGTT TGATTCTA	ACGGAACAAAAGGCA GCTACA	AGAAGATTAAAAGgCTCCGG AAGATCGAC	0
C12987S419	147	GGGTTACAACTTCT AAGGGAAT	TCTCTTCGGCTGCTTT ACGTA	GATAGTGGGAATGAcACTGA AGAAGAGGCAC	0
C13142S250	148	GTTTAAACCGTTTGT ACTAACGA	AGAAAGAAGCTGCTGA GGACA	GATGGATCCTAGcGCTCCCT TCAGTAA	0
C13265S200	149	GCACATTTGATCAG CAAACAGT	CTAACTCCACAAGTAT GAGGCA	ACATATCCTTAGAgCAAAGA CCAAGCCGG	0
C13288S189	150	TCTCTGTGTCCCAA AATGAGAT	GTGTGTCCAGCATGGT GTTGT	TGTTTTTGCATAAGAaACAAC AAGGACAGAAGG	0
C13301S439	151	CTCTACAAACGAGC	GTGCACGAATTGCCAA	GCACCTCAGGTAAAtGAAGA	0

		GGTGTAT	GCGAT	CCAAAAGGTTT	
C13315S616	152	AAGATTCTTGCTGATTGAAAGTAA	AAAAACGTGATACCGTCGTCAT	CTTGACTAATTTTGACTgGACAACATTTCTGTAAATG	0
C13354S446	153	TTCCACTTACAAGTACTTCGAAA	GCCAGTTGATTCGCGGGTCA	TCTTGGTAGATCCgTTCCACAAGGCAAAA	0
C13394S333	154	AAGTTCCTACTTCTTTAACGGAT	CGGAACTCAAGTTGGCGTGTA	TCCGCGCGTTCTtTTGTTGATCACAAA	0
C13486S116	155	GGTTTTATCTCCTCTGGGTGA	ACTTAGAAAGAGCCGCGAGTA	TGTACATGTTTCTTTtGTCCAAGAGATCCTGAT	0
C13535S196	156	TTCCATATTCCGTGCAAGTGAA	AAAACGAAATCCTCGGTGAAGA	TTAGCACTGTAGGTGGgTTTCATTGAAATTTTCTT	0
C13648S225	157	GAAGTCTAATAGTCGCGTAGGA	CTTGATCGATCCAAAGACGATT	AAATAACAGGAATTGAAaACCTTTCTATTGCAAAACT	0
C13698S442	158	CTTCACAATTCTTCTCATGTCCA	ATGAAGACACTTTCACCAAACTA	TCGTAGCTAACTTTTtGAAGTTTGAGAACCATA	0
C13861S511	159	AAACTAATCCCTCCATCTCGTA	GATTTCACTGGTGAGCTGACA	GCTTGCGTGAAGcTCCTGTAGTTGTAG	0
C13905S483	160	CTTGGGCAAGACAGGCAGAT	GGACAAGAATTATTGGCTACAGA	AGGAAGAAATTTAAACAaAATTTTTTCCGACGATTGA	0
C13990S341	161	GAAACACACTATTATAACCGTCAT	TGTGAAGAGCTCCTTACACGA	CTAATCCTCCCATcTGACAATCTGGTGGT	0
C13992S181	162	TTGACCAAGAAGTAGACACCAA	GCGCGGTGAAAAGTTATCAAGA	CAGCCTCAGTGTTtGCTCAATGACATGAC	0
C14018S197	163	AAGGGGAAGCATCACTCCTTT	CGTCTGTTTGTCACCTTTCGA	CGTGAGGCCTTTtATAGCTTCCACGGT	0
C14154S231	164	TATGGATGAAATGTACCACAGTT	CTTCATAGCGGTCTGAGAGATT	GTTCGTTTAATAAACTaGAGGATAGGCCACTTAGA	0
C14161S301	165	GTTAACTGTATATCAGTGAAATGT	GCTTTGAAATGTGGAAACCAGA	AGTTAAATATCAAGCtCTCTGGCCATATTGGGA	0
C14226S523	166	GTGCATTATAAAAAGTGAATATGGA	TGGCAAAATTAACAGGAGGGAA	CTTGCTAGGATGAtGAGCCATGTCTTTCA	0
C14242S316	167	GGTCAGGGGGAGAGAGAGAT	CAGAATAAACAAAGTGTCCAGGT	AGTGCGATTTGGgGGTGTCACCTTTAA	0
C14246S887	168	CAATGGCCTCTACCAACAGAT	GTCAAGAATAACCGTGACAACA	CTATCTCACCAAcGGCAGAGTGGGGAA	0
C14269S102	169	GGTCTCAGTAAATATCCAGGTT	CCACTCTGAAGCTCTTCTTCAT	CAGTGGAATTGACTCaCTTTATTTGATGTCATG	0
C14319S510	170	GGCAGGAAATGGTGCAGATGA	GAAGGTGCTAACAGACGGCTT	GAAGGCGGATATCtGTCACTTTGTGGTGG	0
C14357S360	171	TGTGTATAGAGAATCGAGCCAA	AGAAATCTGAGACGACAAGAGT	CAACGGCCAGCAcCATTGGAAGTAATT	0
C14364S490	172	AGCGTTCCCTCGTTCTCGTT	GTATAACTTGAGTGACCCGCTT	CTTCTTCCTACCTtTGGGTGCAGTTCGTT	0
C14404S340	173	AATCATGCTTGAACCAACGGAT	TTTCGTAACTTGAATTTACAGGAT	CCTGTTAATTGTCTcAGGAAGAATTCCTGGT	0
C14455S306	174	GGGCTCTTGAGAGAATAGTAGA	CACCCACCAAGATCAAAGAGAT	TGTGTAGGTGGTGTcATAATGCCTTGACTCC	0
C14474S185	175	CAAACTTGGCATACAGTTTCACT	AGGAAACACAAGATTTCTTGGAA	TTAACTTCTCCAGAtTGTGAGTCAAGGGTAA	0
C14487S191	176	GCGAGGCTCCCTTACGTCAA	ACGTCACTCCCACGTCAAGT	TTGATGAGATGcGCAAGGTGCCTGG	0
C14532S618	177	CAAAGCTCTTAGCACGATAGAT	TTAGCGTTCCGAAAGGGGTAT	AAATAATTCATCCCAcTGCCGATTGAAATCGAA	0
C14641S195	178	TGAACACAATTCTGTTGCTTCAT	CGATGTTCTTTGAAATACAGAGA	GACTAGAATACTGTTTTgCACTTGTCATTTAACACAT	0
C14723S141	179	TAGATAGCACTGAAATGTAATGTA	AGTCATTCCCAAAGGCAGCAA	AGTCTCAATGTTGCgATTTTCGACAGTTTAA	0
C15011S233	180	CGCGCATCGGTTCCTGTTCT	CCTTTAAATCTCGTTTCGAGCTT	CTTTAATTGGTCCgCGCCATGTCCTATGG	0
C15021S282	181	CAGAGGGTGTGTGTCCTCCA	GAAACAGAAAATGCAGGCTCTT	TACAGTTTGGATGTcAGTGAGCTAACAGCGG	0
C15044S328	182	GGCGAGGGCTTTTCGCTTCT	CCAGTCCAAAGCAAGATGTCA	TCGGAGTTCTCAcCGCTTCTAGTTCTC	0
C15056S244	183	CAAAGTCTGGGATATTGTTCCT	AGTGTGGTAATTCTCAATGACTA	AAGAATAAATAGTCgCAGTCTACTTGCCAAA	0
C15084S136	184	GAATTCTTCAATTTGTAACTGGAA	CGAACAACTATCCTAACAGTGT	ACCGGGAAATATcGCCCACACACAACA	0
C15111S282	185	GCTTAGCTGGCAAGTATGGTA	CAATGGTCATGCACCAATGAGT	CATTTGTGACAGCaACTTCACAGGCTGGT	0
C15113S204	186	AATTGAATTAAAGTATGTTGGGATA	CAAGTTGAGAATGATTCAAAGCTT	ACTGAACAAGTTTgCAAGGTTGTCAGCTC	0
C15150S931	187	ATTCAAGTTAATTGTGCTGGTGA	AAAGGGGCCATTGAAACAGCT	CATTGATAAGAATgGCCAACCTTGGCTCT	0
C15176S465	188	CCTTGGTAGATGGAGACCTCT	TGGCCCATTTCTCCATGAACT	TGGGGAAGAGTAACAaATCTGTGACAATTGTGG	0
C15238S417	189	CATTGATTCCTTCAAGCATGAGT	GTCAAGTTCAAGCACCTAGACA	TTTGGACCATCTGGTAtAGCATTATAATTCCACTA	0
C15269S273	190	GCGCCAAGAGACGAGTGGAA	GTTTACTTTAGAATTTACAATGCCA	TTGGCATTCTCCACaTTGGTGATTGACATCC	0
C15286S686	191	CCTGATAAAGAGTCAGCCCATT	TGACAGAGTGTACTTACTGCAA	CAATGATGTAAACATTaTCAGGGGTAACACTCTGA	0
C15318S250	192	CCTCATGAGCATCAACCTCTT	GACAGGCCATGTGCAGATTGT	GGGGTGGACAATcGGTGTTGGCAAGCA	0
C15351S256	193	GATTGGTTCAGGGGTGGATGA	TTTGACCAGTGCCCATCTTCT	TATCTTGGTGCTCAtGCAATGATAATTGGGA	0
C15355S114	194	CACCGATTTCTTGCAAAGAAGT	ACAGCCGGGCATGCCGATGT	CACGGGTATTTCCgTCCTCTTAACAGGGC	0
C15415S232	195	CTGACTTCTTGATCGATACCAA	GTGCGTCAAAAGCACTTGCAT	CTTCTCCTTGTGCtAAATATCCCAGCAGT	0
C15493S507	196	GTGCACGCGGCACAGCCAAA	CGACCTACGGAACCGGAAGT	GAGGGCAAGGCATaTTGGGAAGTTCAATG	0
C15522S127	197	TTGACACCTGATTCATGTCTCA	GAAACAGCAAAATTGTAGAAGTGA	TTCAACCCAATGTTCaATTTCTTCTGGTGTCAA	0
C15670S505	198	ATTGATCAATGGGTGTGACCAA	CAAACGTTGCGGCAGCCATTA	GCCTTCATCTCTTtCCGGAGGACACAAAT	0
C15741S475	199	TCAATTCGAAATGACCGTTACTT	GAAGCCATCACCATGGTAGCT	ACTGATACAGCGGTtTGATTACCGATGACGA	0
C15774S399	200	GAATCTGCCAAACAGTATGTTGA	ATTTGGTGAAGCATTTCGCACA	CCACAGCATAACTcCCAAAGTATGCCACG	0
C15873S711	201	GTCATAATTTCCACTGGCGTGA	AGCAGGGTCTGTGAGGTACA	CATTATTTCGATGGTcTGTGACAAAGAAAAGAA	0
C15891S454	202	GAGGGAGGAACCCATCCACT	AGCATCGGCAAAATAAAATCACT	GGCGATTCCTCAATtTGTCTTAGTCTTTCAC	0
C15985S312	203	CTCTTGTAGATCTGCTTGCAAT	CTTATGAAGGAGTTTAACAAGAGA	TTTTGATTAACGTCTTTtAGTATGTGAATCTCTTCTT	0
C16096S170	204	TTTTCATCAGCCATGACTGGAA	GAGGAAAAGAACGGGAGAAGT	TCATCGCTTCTCgCCAAATATCCCGTA	0
C16127S174	205	ATTTCTTGTTCCTTCACTTGTCT	ATCAACAGCAGGACTACCTGA	GCTGAGCTTTCCAGgGTGAAATTTTCAAGAT	0

C16136S488	206	TGTCCTTCGCCGAG CTCTGA	AGATGAGAGTGGTCGA AGGTT	TACTCGAACAAAATaAAGCG ATCCATCAGCA	0
C16176S198	207	CGTTTTTCCAGTAAA CGTGGTA	CACCCGTTTGACGAGG CGAA	CGTTTCCACGTGtGATGGGC AGACAAT	0
C16181S885	208	GCCTTTGAAAGGAA CGAAGACT	AGACATTGTCGGGCCT GGAT	TGGCCATTCTTTAgTGATGTA CAGCCGTG	0
C16269S320	209	CTTGCCAAAGAGGA ATTTCAGT	ATAATCTAGAACAGGA GCTTTCA	CAAGAGACGTATTTCcCTTTT GTTCCTTTTGAG	0
C16279S643	210	CATTGTCTATTACTT ACTCTGCAA	CGGATAAAACATGTGG CTCCTA	CTCAGCAGAAATAATgAAAG TTCTTGACCTATC	0
C16387S343	211	CATAACAGTTCTCAA ATAAGCACT	TATTGCCTACAGCGTG GAACA	TTTGTTGTATTGCAgCAGTTT CTGAGGACGT	0
C16442S295	212	AGCAGCCATCCGCT CAGCAA	CCTTACTGCTCAAGAA ACAAGT	TGTCTTTTGATATCgAGGTCT TCCTGTACAA	0
C16449S173	213	TACATGGTGATCTA AGGGTGAT	GCTCATTAGTTACTGC TTAGCTT	GTTTCTTTTGGGAAGaAAGC TCATGAATTAGTT	0
C16458S418	214	GATGGGATTCATGA TGTTTCCA	GTTTGTACATTCACGC GTGAGT	AAACAAGAGTTGAGGAAaTA ACAATAATAGTAAAGTA	0
C16499S363	215	TATCCTTTGGTTCTG GACATCT	CTGTAGGTGTACGATC GACTT	GTTGACCCTTGAAcTGGACT TCTGCCCTC	0
C16549S511	216	GTCCTTGGAAAGAG AAGATTCA	GGCTACACGAAATCCA AAGGT	TTTTGGCTTCGAAgCCTTTC GTTTTGGTC	0
C16621S398	217	CTGCTTGATTGAGG TCACTGT	CCCTACTAAGATCAGT AAAAGGA	ATTGTTATTCAAATTgGGAAG AATCTCGCTGGT	0
C16634S406	218	TGTGAACACCGCTG GCATCA	ATAGCACCGTCATACC CATGT	GGGTAATATTGGTGTTcATA CTGTAGCTGATCCGT	0
C16716S153	219	GCCACTGGGATTAG GTGATGA	AGGAAATGATTCAATG TCAGTGA	TGACAATAATAATAACcAGTG TCAGGCTAATCCTA	0
C16867S473	220	TTGAGCAATGCAAA AGGGATCT	CCAACCAGTTCAGCAG CCAT	TGTGAGTGAACTgCAGAAAG AGCTGGT	0
C16912S265	221	TGCAGCAACGGTGA TTTGCTA	CACGCCTAGAGACAGT TGAGA	CCCTGTACTACCATcTGACT CACCAGTATTG	0
C16956S551	222	GGAGTGAACTGGAG CAAAGAA	AGTTCCTCTCCATCTG GTGTT	ACAGTAAACAGTTTGGcTGA AGAAATACAAACCTC	0
C17050S589	223	GGAAACCTAACACG CAAGTGA	ATAATAAACGTCTGCT GAAGCAA	CCATTCTCTAAACTGtTGAAT GTAGGCAAGTCT	0
C17077S225	224	AGCTCCTTTAGAAG ACTATCAAA	CGTTGATCGCCTGTTC GAAGA	TGCCTTAAATTCATcGATTTG TTTCGGCCTA	0
C17151S285	225	AGTGAAATGGTTGT CATTACACA	GAAGAGATCAACGCTG ACACA	CTCCTTGCACGTTaGTGATC TGAATGCAG	0
C17299S143	226	CTTTCGCCTGCTCTT TGTTGAA	ATCAAGCTCATGACAT GACAGT	CATACAGCGTTTCgTTCTTTC TCAGTCGT	0
C17330S121	227	GTCTCAGTTAGTTTG TTCTCCA	TCCTTAAGACTTGTGC CACAAT	CAAATTATCGCCTCaATCAAT TCCAAACGGA	0
C17438S197	228	CACACTGTCGATCT CCGTACA	GGAAACAAGAATTGTT TGGTCAA	GAGTCTGTCATAaGTCAGGC TGGCCCG	0
C17471S281	229	ATTTGGGAAAGGGC TCGGTAA	CCTTGAGGTAGCACCA TAGCT	ATGGCAGTGGGATaGTGGA ACCTGTTGTT	0
C17475S294	230	ACTGCGGAGGACGA GTACGA	CGCTAAAACAGGGCTA GAGCT	CAAATTTTCGTTTCCaAACCT GTCATGCTACTC	0
C17479S262	231	AAGATGAGGAAAAC TAACCAAGA	CCACCGTGAGCGACG GCATT	CAGAAGTAGTTGAgGGCGG GTTTCACTCT	0
C17498S226	232	AATCCTTTCCTGATT TGGTTTGA	TTGTTTCCTACGGTGC GCAAT	AATTATCCCTGGTGTtAGTCC CAATATTGGGGG	0
C17622S201	233	CGCCACAGAGTTAT GTCATGT	GCCCATCCGTCACCCA GCAT	TGAACTTCTAGGAgCATAGC TGAAAGCTG	0
C17741S312	234	CAATTGTTCTTATTT CTGTACTATCT	CGCGCAATAACGTATT AAATGCA	GGTAATTTTCCACTaCTTCG CTTCTTGCATT	0
C17828S268	235	ATTAAGCTCTGGCT ACTTGTCT	CTTAACTGTTGAGCAA AGCACA	TCAGCTCTTTCAaAACTCTG GCCATTT	0
C17912S202	236	CAACCCCAAGCTCT AGCCAAA	TAAATCTGGTGACAGT TGTGCT	TTTCTTTTTCTCTTCTaGACT TATGCCATAAGAGG	0
C17914S739	237	CTTCGGCGGCTTGG GAGGAA	CCGAAAAAGGAATTGG TGGCTT	TTTCTTGCGATTgTCGAGGT CCTCACC	0
C18084S286	238	TTATTCTGCCCAGCT TCTTTCT	ATCAAGACGAATACGA GGAGTA	CAGTCGCATGCAaGCTTTCG TCGTACA	0
C18165S232	239	CGACAGAAAGCATA AGCTCCA	CTCGCTCTGCTTTTCG CAGAA	CTGGGAAGAACGgCTGAACT CGATCCT	0
C18185S479	240	TCCATGTCCATCATA AACACCA	ATGGAGGATTCACATA CAGAGA	AGCAGTGCCCTTgTCTTCTC TGAGACT	0
C18231S140	241	CTTTCTTCGCCACTA AGCTGTA	GGACTGCGAACAAAG GAATGA	AGCAATTTACTCTCcCCAGA TAATGATGCAC	0
C18363S421	242	TTGGAGGCAAAGTG CTTGTGA	ATTTTGTCAACGTACC AGAAAGA	TATTTGATTTAGGTgGGTCG ACATCCTGAGA	0
C18366S189	243	GTGAATCAAAGGGT CCACTGA	GGTGGGAGATTTTGAC ACAGA	TCCAGTGAGAGACAtTCGTC AAATATCCACG	0
C18397S183	244	GAGAGCTTGGCATC TACTGTA	GTCTTAAACAATAGGA GAATGACA	GTACAGTGTACAGGAaTGGT GAACATTCTGTTA	0
C18442S324	245	AGTTAAATGTACGTA CTACTTTCA	ACAGGAGTATGTGATT TCCACT	CTAAGACCTTGAGATaCTTC ATGGATTTCAGAC	0
C18487S1302	246	CAGGAAACTTTGTA CTTCCCAA	GGATTTCAAGTTTTGTT GTTCGT	TGTCGAAGCCGTGtCAATGC TAGTAACTC	0
C18576S293	247	TACCACTAGACTCTT GAGTTCT	TAAATTACAAGTGTTC GAACTTCA	GCTTTATTCAAGtGAGCCGT GGGCCAC	0
C18580S230	248	CAAACCGTTCTATC GGACCAA	CAAACTTCCATGGTTT GAATCATA	CGGAAACTTGAGTTaTTTAC TCGCCACTCTC	0
C18603S149	249	GGTGCCCAGACCAC CATCCA	GCTCAGCAACTTCCTT TGGTT	TGCTGTGTCTGAgGACGTGG AGTCTCC	0
C18841S310	250	ACCCACTCTTGAAA GACTCAAA	GACGAGATCCAACATG CACTA	TATGAGCCAGCAgCCGTTGA GAATTTA	0
C18920S453	251	CTCGGCGAAGTCTA TCGCTTA	GTGTATTTACAAGTTCT ATCTCCA	CAAGAAACGTTTTATaAACG TGCAAGCAAAAAT	0
C18993S556	252	GTTAAACTGTTCCCA TTGCCTTT	TGAGAATCTTTGCAAC TGTGCT	ATGACAAAGCTATTAAaAGG ATTCCTAAAAGGATA	0
C19002S323	253	ATGCGCAGTCAAAC AAAGATGA	GCCAGCTGTCTTTAGA TGAGT	TTCCAAACCCAAGGCgGTAA AGGATATTCTTGG	0
C19092S284	254	GTATGGATACGTCA TCGTTAACA	GCATCGTGATAGAATG ATCCTT	GTAACATTTCAAGAGaATCG GTCGGAAGATTAC	0
C19178S536	255	CTCCAAGACCAAAG TGCGCAT	AGGCTAGAGCAAGGG CGACA	ATGCAGGGGCCCtGCTTTAA ACTCAGG	0
C19263S650	256	AATAACAAGGCATC GCCAGGA	TTTACATCATGCCACG GAAACT	GCCTTCAGTCCTATTCCCCT GAATCGACT	0
C19364S520	257	TCTGTGTCTTCATCT TCTAGGA	AGTTTCGCTTGAAATT GTACTGT	AGTGAAGAAGTAACTcAAGT TGAAAAACGGCTC	0
C19470S311	258	GATAGTTGTTTGAC CTGGTGTT	CAAATCCGAGCTCAGC GGAAA	GGTGGTCCCGGcGCCAATG AATTGT	0
C19478S130	259	TTTGCCAGAGAACC TCCACAT	ATGAAGTTGTAAGGTG GAGGAT	GGGAGTTTGAATCgCAACAA TATGCCACA	0
C19533S241	260	CATTGACCATGCAC	GCGAGTCTGATCCCTC	GTCACTGTGGTgCCACCGTG	0

		ATATACCA	AATGA	AAGGC	
C19552S147	261	AAGAACTAGTGCAA AAGCATTGT	CGTCTAAGTTATGACA TGCTTGT	CAACTTAAAGTAGAAGGtTAT TATTGAGGGTGTAGAA	0
C19560S178	262	CTAATACAATGTAAC AGTTTCAGA	TAGCACAAAGTGATGG AAATTCA	CGACACTGATCAATTCtATAT TTTGCCTAGTTTAG	0
C19713S134	263	CGATGTCGCCTTCT TCGTCAT	GACTTGAAATAGGACC CATAGA	CGGAAAGTGGAATgGTCAGC CAATCAGTG	0
C19740S286	264	CGGCCAGCCTTTAT CGTCCT	CTATAGGGCTGTGCAA TTCCT	TAAATTTAAATCcGCGCCGT GTCTCTG	0
C19862S335	265	TGGTGATCCACAAT AGGCCTT	GCCCCTTGGACAGCA CCAGA	GATGCATCTTGCaGCCCTCT GTACAAA	0
C19916S128	266	TCTTCCTCTGATTCA GATGTCA	GCTTCCACCAATCAAC ATGTCT	AGTCAACTCTTTGTCaCTGT CAATTGACACAGC	0
C19881S196	267	TTGAAGATGCTGGA ATGGAAGT	CAGATCCTTCATGCTG CTTGT	CCTGTCTAAAGAgGAGGCTG CAGAAAA	0
C19928S437	268	GAACACTAGTCTTG AGGTCTCT	AAAGAAAATAACGAGC TGCATCT	TGGTCAGTTTCTTCcCGTAG TTTTATGAGGA	0
C19982S400	269	CTCGTAAACTGGGT TTATTGGA	TCAGCGCCTTTTGTCA AAGGA	GGGTATAAGCCCTaTGCCAC GAGAAATCC	0
C20102S582	270	AGTTTTGTCCAAAG AAGGAATGT	AGTTAGAGGCATACGA TTCAGA	GTTGATCTTGATCTTcGTCTT CATCTTCATCTT	0
C20163S412	271	CATGGAAGTTAGTC AGACAGTT	GAATGATGAGGTTGCA TTCAGA	TTTGCCTGGACTTCtTCAACT ATCAGTTTTT	0
C20167S379	272	TCTGCCACACCAGC CCCTAA	GTCAGATTTGCAACAT TTGAGTA	ACATAAAAGTGTCTGaACTT GGGTCAGTTTGGG	0
C20274S537	273	GCAATAGCAATAAG AATGTCAGA	CAAGTGAAAGTTCATT TCATGCA	AAGCGATCTAGTTCcCTAAT CCATGTCTTGT	0
C20399S426	274	CAATTATCATGCCA CACAAGGA	TTTCGACTTGACTTAG GTCTGA	AAACTCTGCCAACaTCTGAA GCTGATTAA	0
C20407S208	275	GGAGAACTATTTGA ATTCTTGAGT	GACAACTTGCTTAGTG AATCCA	GAACAAGAGTATTTgTCAGA AAGTGAAGCCG	0
C20442S307	276	CCAGAGCCCGCCG GCAAAGT	GGTTGATCGCTCGGG GGCATA	GCAACAAGAGACTTaGCTAT CCATCTTGCTC	0
C20443S297	277	ATTGATGCTCAGGA GTCGTCA	GGAATTGTCGGGTATT CCGTTA	ATTTGTCGCACTtTCGCGGG AGGAAAG	0
C20570S83	278	GCTCCGGTGGTAAA GACTCAA	GGAAGAACGTGAGAA GAGACT	TTCAACGGGATCaAGTCCAC CTGGGGG	0
C20581S243	279	CGAAGATAGGTTCA ATGCAAACT	AAGAGGGGATAAGCC ATTGATA	TCTTAAAGCTTGGcCTCGAA CATCTGTTC	0
C20625S210	280	GTATTACACAGGTC GCCCTCA	TTTTGGGCAAGGGTGC GGAA	AGCATTTGAAGTCgCATTCA CTTCCATGA	0
C20763S245	281	TCTGACAAACATTTT TGGCACAA	AAACACGACACAATCA GACCTT	GCTCCAAGTATGTATTaCAA CACTTATCAACACTA	0
C20768S189	282	GAATGGCCAAAGCT TGGAAGA	CGATTGCTCCGATTGA TGTGT	TCTTTTCTGGTCCcAGACCA GTGTTTCGG	0
C20821S413	283	CATTTCCGCCAATG TGTTTCCA	CCACATCAATAATCTC CTGCAA	ATTTTTGCTCCGcTGCTTCA GCAGAGT	0
C20998S134	284	GCCAAAATGGATAC ATGAAGCT	GCACATTTGCTTGAAT TTTATCGA	AATAGGAGACTCAcGCTGTT GTAAGACCA	0
C21135S139	285	GAATCCTCAAGTTA CTGCACAA	ACCCGCGGAAGTTATA TCGAA	TCGCGTTCTGAGaAGCTGG GACACACC	0
C21186S526	286	AAAGCCTTTATTCAT TTCCATTGT	AACTAAACTTCGCACA GAAGGT	CGCGAAAGTTAGAcGAAACG TGAACACTT	0
C21244S233	287	AAAGACCTGATAAT CCCTTCCA	TTATCAGGATGTTGAG GTCGAT	CGAACTTGAAAAGTCtAAGA GAGAGAAAGAAGT	0
C21253S536	288	AGGATAAGCCACAG TAGACGT	AGTGCAACAACGCCG GTGAT	CTGAATAAGCTTGaGTCGGG TGTGTTTAG	0
C21349S456	289	CAAGGACATTTAAC GAAGGACAA	CGAAACGAAAAGCTGA GATGCT	GAGTTTTCGTTTCGGtAGAT GACAAAGAAACTG	0
C21833S285	290	AAGTCTCGCAACCT GAACACT	AAATTAGCAGAAGGCT GTGCTT	CATCATCATCGCCAcATCAA AGTCTTACACA	0
C21844S313	291	CCCATTTTTCCTCCT TGGTGTT	TGATTGATAGTCAAGG TGAATCT	GTGAAATATTAAAATGgCTTG GGGATATCAGATCC	0
C21914S231	292	GCACAAAGCATATT CTCCAAGA	TCTGAGGAAGAGACAA ATGACA	TTCCCTTCATAATCcAAATAA GCTGCTCTCT	0
C22100S336	293	ACTGGCTTCACCTT GTTGACA	TGCATATGCCTTACCA CCTCT	TGGCTCTTGATAGcTACTGG TGACTTAGG	0
C22109S391	294	CTTTATTATCGACCT CAGAGACA	AAGAGAGTTTTATTGG ATCTGGA	CCAAATTAGAGGATGTaTTTT GTTCCTCTAGAGGT	0
C22110S143	295	CACCGCAGCGAAG GTTGACT	GACCAGCAACTGTCAT GACGT	AAGTTGTCGCATgCGCAGAC AGGTTGA	0
C22138S164	296	GGCTTTTGCTGTGG ATCTCAT	GTGAGGATCAGTCTTG GTTTGA	GCATCGACTTTAGaGTTACA TGCTGGGGA	0
C22182S205	297	ATCGGTGTCTTCGG CTGCAA	CATTCACTGATTTTGAA ACGAACA	CTTGTTGTCCCATGCTaTTC AAAATCTTGTACTAT	0
C22306S240	298	ATATTTACGTGATTA TTTCCCAGT	GTGGATGGTATCCCAT GATGT	CTAATTCGAACAAAgAGACT CGATCCTGAGT	0
C22425S453	299	CGCACTTGACGACC TCGCAT	CATATCCGTGCAATAA CAAGGA	TTGTTTTTGAGATCTaTCCCC TTTGGTCCATGG	0
C22427S223	300	TCAGTCGCATAAAT ACTAACACA	GTAGATCCACGCTAAA TGTTCA	TAATATTAAACAGAAAaGAAA CAGCCAGAAAGGAC	0
C22464S266	301	CTTCCTAAACTCCTT GGTGCAT	TTCTCAGAATGTTATTG CTGTCA	AGAGCCAACAGGtGAATGAG GTCCTGC	0
C22489S363	302	GTTAAGAGAATAAG TCAATGCCAT	TTATGATAACAGGCCA GACCAA	TCATCAGGCGATtCACTGTA CAGGCAA	0
C22526S224	303	AAATTCTTTTCCGTT CTCTAGGA	CTACTTTGTTTCGTGG ATATCGT	CCCCAAGCACACCcCAACTTA AGTCGTT	0
C22545S1379	304	CTCCTGGTCTAAAG TTTGATTGA	GGCTATGGACAGGAA GGCTT	GTGCCAGAATGATgACCAAT CCTTGGAAA	0
C22633S340	305	CAATTCGTGCACGG CTTCCA	CGTTTTTCCCGAACAT GCACAA	AAAACGAACTTTGTCaCGAT TTTTGAATCTGAC	0
C22643S340	306	CTTGCTGCTGGGAA GTTCCT	CTGTTTCCTTGTTAGAT ACGGTT	TTTTTAACATCAACAACcTCT TCTGGTATACTTGACA	0
C22761S360	307	CATTGGACTGGAGC CAATAACT	GTTTCAATGTAACATTC AATTCCT	AAAAGCCGAGAAcTTGGCAA AACGATT	0
C22820S193	308	AGTTGTCTGTATATT AGACATGTA	TGGCCTGCATATGGCT TCCA	ACCCCGAAGTAATTtGAGTTA CCCTTGGTCG	0
C22821S388	309	TCTGCTCGATCTAG GTACTGA	CGATAAAATATGAAGC TCAAGGT	CATTTCTGTCGAATgCTCTC CTTAGATTTTA	0
C22826S366	310	GATTGAAGGCGCTT TGGTGTT	GAGCTGAACTTCGAGA TCTGA	CATTTTCATCTGTAAGtTTCT CTTTCTTCTCTTAG	0
C22875S709	311	CTCTTCTTTTGCTGC ATGTTGA	TGCAGCAAAGAGAGA GGAACT	TTGTTTCTCTTTCAAcACACC ACCAAACGATAG	0
C22900S198	312	TGACTTTTGCACTAG CCTTACA	ACTAAAAATATGTGGG ACTCTCT	GCATATACAAGAAATATcATA ATGAAGGCTAGTGTAC	0
C22973S285	313	ATTCCATCATGCGTT TTCCTCT	GAAAATTTTGAACACA GAAGAACA	CTTGCCACCGAAaATCCTCT TTCCTGA	0
C22982S334	314	CTCAGCATCGTATC TCTGGTT	TTAGCTCACAAAATAC TGATTTCA	TACTTTTCGAGTGtTTCCAGT CCTTCGGA	0

C22993S160	315	CTTCTGTGAAGGGT GTGAGGT	GGGTATTGACACAGAT CCACT	TGGATGAAAGCGATGAtGAA TTTTATAGGAAATAA	0
C23019S237	316	AACAAGACAAAAGT GATCACGTA	CTTCTTTCCTTCCTCA GCTTCT	ACACCTGGTGTgCCTGTGAC CCCCG	0
C23083S345	317	CTTGGGCAGGAATG GACTGA	GAAGACAAGGGCTGC TTCTGA	AAAGGCTCAGGAAtTAGGCA CTGTTGTCC	0
C23085S183	318	TCGCACACTCCGCA CTTCCA	GGAATAACGATAAGTG TCTCCAT	AGCAGCAGATGCATAgTATT ATGAAAAAAACAG	0
C23126S678	319	GTTTTTACAGTTTCT TCCTGTTCT	CACAAAACCTTGATTG GGTTGA	TGGTCTTCCTCTACaGTGAT GTCCATTCTGA	0
C23210S557	320	GATTGTCCTTGCTG AAGGAGTT	GCATTTTGTCCTTTGA CTTGTGT	GAAGCACTGAGAGAGAAcAA AATTTTGAAAATATTAT	0
C23327S599	321	TCCTGAGGCAATGT GTTTCCA	CCAGCCAGGAATGTG CAGAT	CGACTATCTGGTCTgAGTTC TTCAAAGTAGT	0
C23375S174	322	GCTGTGAATCAGAA CTTTCAGT	AGCGAGGAAGCAAGT GCAGA	ACTGAGGAGTTTGATcATTA CTCTCTTTAGGCA	0
C23489S194	323	CAGTCAACGCAAGA AATATTCTA	GACCATCTTTCATGAA ACACAGT	AGCACAAGCTTTTAaGCACT AGAGTATGTCC	0
C23502S311	324	CTGACATATAAGGG TAGTCCGA	TCAGTCCATAGCCGCA AGGT	CTCAAGGAGGATaCGGAGT CGCATTCG	0
C23508S203	325	AAGATTTTCAGGATT TATTTGAACA	GCCTCATGAAAAACTG CTCTGT	GTCACAGAACACTCgAGCTT TCAAAAATATA	0
C23525S293	326	ATAGCTTTCTTACTT GTACAATCA	CAAATAAAACAGCTTTT GAAAGGT	CTTTTCTTATCTTCCAAcCAA ATGGTAGAAACAAAGT	0
C23566S420	327	GTCATGTAGGTGAG CAAGTAGA	CAAACAGATAATGAAG TCATCACA	AGCTTTTTCAGATCcTGTCTT CTGAACAATG	0
C23734S391	328	GTCAAAGATAGAGC ACTCAAGA	GCTAAATATTTTGTTCT CTGTCAAT	CCATACCAAATAACAcAGGA CGCGTATATAGGC	0
C23738S719	329	GCCTGAACCTTCTT GTCTTTGT	AGGCAACCAATATAGG AAGAGA	TTTTCCAATACATGTcCCCAC ACTTATGAGGAA	0
C23950S250	330	GAAACTAACTGCAA ATGATTCGT	ATTTTCTTCGTGAACTT AGAGGA	ATTAATTTTCAAATCgGTCAC GTCTGACTCAAC	0
C23978S544	331	CTAAACCCAAAAGA TCCAAGAAA	ACAAAGTTTAAGAGCT GTCAGAT	ACAGAGACCAAAGcGAACCT AAACCAGTC	0
C24058S463	332	CGCCCCATTTTGTAT AGATTCT	GTCTCAACGAATACCG ATGCT	TCTTGCCTTTATAATtTGTCC CTTCTTGACGGC	0
C24096S618	333	CCATTACAGTTGTG CCCAAGA	GGAAGGCAGAATACC GTGCAT	GTAAGATGTGCTTaCATGGC TGGTCTGCA	0
C24216S175	334	GGTCCTTATTTCCA GTCAATGT	CAGAATAGTAGACAAG ACTCAGA	AAATGTCCAGTCAAAtGGTTT GATGACTTCTAA	0
C24129S242	335	AGCAAGGATTGTGT TCGTGTTA	GTGGGTGAGACGTATA TGCTT	CAGTCAGATCAATaTTTGGG GATTCGCGC	0
C24140S397	336	AGCATCCAACAGCA TTCATGAT	TGACTTGGTGTTACGT GCACA	GACGCATTGTGGATgACTTG GTAAGGAATAA	0
C24238S242	337	TGAACACACTCGGG CTCCTT	CGATGTCGATCCTTGT GTGCA	TTTTACTAAACTACAGtTATG CGAACATCCACCTA	0
C24321S173	338	AATTGAGGCTGTGG TGTATTCA	CATAGAACATCCTATT GTACACT	CAAGAGTTTTTGTGATAaTTA TTGGCAATTTTTCCAC	0
C24388S705	339	AGCCTGCTAATAAT ACGAGTCA	AGAAATCCTATGGCCA CAGCT	GCTGTCAAAGTTCGTgTAAT ATTCGCTTATTAA	0
C24438S225	340	CAAGCACCCCCTTT GGTGAAA	TGCCATTCAATAAACA TCTCTCT	TCATTTAACATTCGGaAAACT TAACGTGCAGTT	0
C24582S267	341	TTTATGGATTTAGCT GTGCATCT	CCTATGTGACTAGCTT AATCAGA	AGTTAGAAAGGCAaTTGGGA GTCTGCTTT	0
C24813S193	342	GTACTCTAGCTGAA GGTTTTACA	CAGGATTGGAAATGCG TAACCA	CACGCGAAAAAGGGaAAAC CTGTTAAATATT	0
C24856S313	343	CTTTTTCTATATTTT GGGTCATGA	GAGACTGATCAGATTT CTAAAGT	CATCTTGTACTTCaTGTACTT GCCGGCCT	0
C24897S240	344	CTCCTTTTCACAATC ATTCTCTT	GAAAGTAGAGACACCA GTTCTT	GTACTCTCTTCTTCcGTTTCT TTCCATTGGT	0
C24932S258	345	GTATTTGTCCACAAC CTCTTGAA	CGGACGTTTTAGCGAC AGGA	AATTTGAAAACTTTaACGCC GTCTTGCTCAC	0
C25131S634	346	ACCAGTAATCATTAC TTCATGGA	ATCTGTATCCGTAAAA TGCCTTA	GAGTGATGAGGAAGaTACAG CCATTGAAAAT	0
C25187S178	347	CATTCCCTCACATAA ACTGCAAT	CAAAAAGAAAGGCTTT GCTGGA	TTGCCTTCATGTTCTtGTCCC TAAAATTCAGTA	0
C25192S305	348	GGTTGTTTCCTTCG GTGTTGA	CAGAGACCCCAACAAC GACA	TGTTGAGGTCCcTCTGGTCG CAGCA	0
C25225S451	349	CTCCTAGCCATGAT GTACAGT	CCTCGTTACAAGGAAA ACTTGTT	GCACTGACAGGTGTcTGATG CATTTTAATTC	0
C25234S280	350	GAAATGAAATAATAA GGAGACCAT	GACATTATCATGCACA TTACAACA	ATCCACCATTACCgCTGCCA TATCTCTAC	0
C25302S260	351	GCTACGTGATATCT GTCGACTT	ACGGCAGATTCGCCCT TTGA	CATTAAAATAACACcTCCCG GGAACCTTGGA	0
C25351S196	352	GACATCCAACTATTA GAAGCACA	CAAGTTCGTGGGCATG CTCT	GATAGATGAATTAGAgAAGG AAAAAGAGCAGGC	0
C25425S128	353	TTTAATTTCTGTTGC TCATCCCA	TGAACCACCACCTACT ACACT	AGATGCTTTTTGCcAAACCC TCTCTCAAC	0
C25536S620	354	TGTAGGAGGTGGCT CGCTTT	TGAGCCAACAGTTCTT AAGCAA	GTTTTGATCTTGCaCTTCGG GAACAGAAA	0
C25568S279	355	TGAGGATCCAGTTA AGGACCA	AAACTGGAGATCGTAA CTCACA	AGTTCACTTGCCAgGGACAC ATCAATGTC	0
C25652S324	356	GGGAAAAAGCGAAA CACGAACA	CGGATTGCAAAACTCT GGAGA	CCGACCATCCGTtAGCAGTC TCGTTAA	0
C25677S330	357	GCTTCTTTTCCATCT CCGTCA	TCTGGGAGAACTTTAG TGTCAA	CTATCGCTTTCACtATCAGAA CCAGGTCT	0
C25688S405	358	TAGGGTTGTGGTTG TGGGTAA	GGCTTTCAAGGTGCAG GCAA	TGTCGGAGGTGGcATGTAAT TCTCACC	0
C25713S318	359	CTTGTGTTTCGATAG TGCTAAGA	CTCGTGTCCAAAGATC ATCCT	TTACATTCCAAGTTAAgTTTT CCGTACATGTGAAT	0
C25725S230	360	CGACCAGCAGACAA ATGAAGA	GACAGAGGATACAGA GGAAGT	ACCGAGGAGATCtTCTCTGC CGTTCTG	0
C25946S829	361	GAACGTGAACGGAG GAGGATA	CCGAAGGGGTCAAAAT GTACT	ATTTACCTTTGTCGCtGAATT TCTTGAGATCTG	0
C26043S200	362	GAAATCTACGGCTA GGAGTTGA	CCATTGAAACCCGAAC TGGTA	TATCAAGTTGCATAaGTGAC TGCAAGTTTCG	0
C26116S342	363	CTGTTCAAACTGTC CCAGCTT	CTCAGTGGATCTGATG CTGAAT	CATATAAGCTGCcGCCCGTT TTAAGTT	0
C26140S243	364	CGTCTCCCTTGGCT GCTAGT	TGAGGTATCATCTATA TGCATCA	GCGGCAAACTTTGTgTACCA AATTGCAACTA	0
C26271S403	365	GGCGGCAAGATTCG CTAGAT	CCCCGTGACGTTTTCT CTCA	TCGACGGTGGGGcCATTTTC TCCATGT	0
C26275S382	366	CTTGTAGTCTTGGTA GTGACATT	GGCTCAAGTAAAGGAT TTGCAT	ACATAGCTCCCAATgTGTTTT TCAGGATCTA	0
C26311S424	367	CGTCTGTACCATGG CCATCA	CAAGCGTTAGGTATCG GAGGA	CTGAGCTAACTGaGCTGTCA AACCAGA	0
C26478S226	368	TCGATTCTCTACGAA TTTCTTCA	TGTTAGTGGATTCCAG TACGAA	CGGTTTGCCGTCgACTTCTA AAGTGAT	0
C26794S214	369	GGATGCAATTAATG	GGCAACAGAACTTTGC	CGGGAAAAGAAGCcACACAA	0

		GAGAGATCA	TGCTT	ATTAGCACG	
C26831S450	370	GGCCACATAGGCAC ACGCAT	AATCTTCGGCTTACTG GTATGA	ATCTTTTCTACATATCgTCGT GAATCAATTCAAAT	0
C49697S354	371	AGATTCGATGGACG CCATTCT	TTTGTGAGATGGCAGC TGTGA	GCTTGATCATCAATgTGAATT GCTTCTGAGT	0
C26997S204	372	GACATCAGAAACTA TGACGACAA	ATCATCCAAAGTGAAT CGGTGA	TTCAAGTGGAAAAAGGaGAC TTGATTATTCTCCGA	0
C27026S472	373	GATGTCGTTGTTCA AGTCTTCA	CAAGAACACATGGAAG ACACCT	AAACCAAGCAACtTTCCTCC GCGGAAG	0
C3729S182	374	GCCAAAAGTGTAGC ATTGTTCA	GTTTTGTAAGTAAAGC ATAAGGAA	TGTTTGGATTGAGACTTaAAT AAAGTTGTCATTATAG	0
C5145S66	375	CCATGTTATGGTCTT GGCAGT	GTGTTGGAGCCTGGG GACAA	GGGCTATTTTGCAAATgATG ATTAAAAAGTGAAAA	0
C8136S163	376	CTTCATCGCCTTCC AGTGGT	GAAATGGATGCTGGC GAAAGT	TATCTTCTGTAGgGCCTTCA GCTGCTC	0
C11076S81	377	AATCAAGGAATTTCA AAAGTCTCT	GAAGCGCCTTCCCGC AGAAT	GGTCTTTCCTGAGGAaAGTG ATAATCAATGATT	0
C14848S1085	378	CAGGCCAAGACAAT TTAAGTTGA	GATAGGATTTTTTGCT ATCGAGAA	GCTAACTAAAGAGCgCTTTA ACTTGAAGATT	0
C25444S173	379	TTGATGTTGACACC GCCTGAA	AAATAGGAATAACCAG TGTGGAA	AATGGATAGCTgGCGCTGGT AGCCT	0
C14755S556	380	GTCTCCAAGGTGCA AATAGCA	GCCCTGGACAGTTACT CACA	ATCGCTGGACAAGaCAGTCT TTGTATCGA	Tetratricopeptide_repeat_protein_25_(Ttc25)
C54074S403	381	CGACAACACCCTTG CCAGCT	CCGAAGAGTTTGACAA GGTCAT	GCGAATCACATTCATgGTTC CTTTGATGTTGTG	3-oxoacyl-[acyl-carrier-protein]_reductase_(fabG)
SI15295S1881	382	CAGTTAAAAGATGT AAGGGGATA	CAATACAACTTTATTGT GCATAGA	TTATTTGCTATTTTAATaATA GAAAACGTGTAACAGG	Catalase_(cat)
C17723S124	383	GCGAAATGCAATAA TGGTACAGT	TAGTCCTAGCTTATGG ATCACA	CTAAACTAGCAAACAcGGTC TTCAAACATGAGT	0
SI13055S1796	384	GATATTGCAAGACC GCAAGGA	TCTCATTTCGTACACAT AGGCT	GAAAAAGGAATTGCcAAAAA TGCTGTGGATG	Heat_shock_protein_105_kDa_(HSPH1)

Supplementary table S.3. Data sources for environmental variables used in this study.

Category	Variable	years - data collection	Owner institution	Reference papers	URL
Water quality	Chlorophyll	1992-2006	Australian Institute of Marine Science	De'ath 2007 De'ath & Fabricius 2008	http://e-atlas.org.au/content/gbr-lagoon-water-quality-data-aims
	SS				
	PN				
	PP				
	NO_3				
	NO_2				
	NH_4				
	DIP				
	TDN				
	TDP				
	Silicate				
	Secchi depth	1992-2006	Australian Institute of Marine Science & Queensland Department of Primary Industry and Fisheries (QDPI-F)	De'ath 2007 De'ath & Fabricius (2008)	http://e-atlas.org.au/content/secchi-disk-depth-measure-water-clarity
Temperature	Mean SST	2000-2006	Hadley Centre for Climate Change, UK Meteorological Office	Rayner et al. 2003	http://e-atlas.org.au/content/sea-surface-temperature-hadisst-11
	Min SST (Satellite data)	1990-2003	National Oceanic and Atmospheric Administration	Wooldridge & Done 2004	N.A.
	Max SST (Satellite data)				
	Range SST (Satellite data)				
	Min SST (Temperature logger data)	1989-2014	Australian Institute of Marine Science	N.A.	http://www.aims.gov.au
	Max SST (Temperature logger data)				
	Range SST (Temperature logger data)				
Others	Soft Corals: Richness of zooxanthellate taxa	1997-2008	Australian Institute of Marine Science	Fabricius & De'ath 2008	http://e-atlas.org.au/content/octocorals-great-barrier-reef-0
	Soft Corals: Richness of azooxanthellate taxa				
	All seagrass species combined	1994-1995	Queensland Department of Primary Industries and Fishery (QDPI-F)	N.A.	http://e-atlas.org.au/content/subtidal-seagrass-distribution-qdpi-f
	Species richness of interreefal fish communities (seabedbiodiversity)	2003-2006	Australian Institute of Marine Science	N.A.	http://e-atlas.org.au/content/seabed-biodiversity-continental-shelf-great-barrier-reef-world-heritage-area-fish-communitie

Supplementary Table S.4. The numbers of samples (fragments) and colonies that were used to measure three phenotypes (Fv/Fm; chlorophyll; CoQH2).

Locus	Genotype	Sample number	Colony number	Treatment	Measurement
C29226S281	AA	2	1	27	Fv/Fm
C29226S281	AB	5	4	27	Fv/Fm
C29226S281	BB	11	5	27	Fv/Fm
C70S236	AA	5	2	27	Fv/Fm
C70S236	AB	9	5	27	Fv/Fm
C70S236	BB	4	3	27	Fv/Fm
C29226S281	AA	2	1	32	Fv/Fm
C29226S281	AB	7	3	32	Fv/Fm
C29226S281	BB	7	4	32	Fv/Fm
C70S236	AA	3	1	32	Fv/Fm
C70S236	AB	7	4	32	Fv/Fm
C70S236	BB	6	3	32	Fv/Fm
C29226S281	AA	0	0	27	Chlorophyll
C29226S281	AB	2	2	27	Chlorophyll
C29226S281	BB	4	3	27	Chlorophyll
C70S236	AA	2	1	27	Chlorophyll
C70S236	AB	2	2	27	Chlorophyll
C70S236	BB	2	2	27	Chlorophyll
C29226S281	AA	0	0	32	Chlorophyll
C29226S281	AB	4	3	32	Chlorophyll
C29226S281	BB	2	2	32	Chlorophyll
C70S236	AA	1	1	32	Chlorophyll
C70S236	AB	2	2	32	Chlorophyll
C70S236	BB	3	2	32	Chlorophyll
C29226S281	AA	0	0	27	CoQH2
C29226S281	AB	5	4	27	CoQH2
C29226S281	BB	7	4	27	CoQH2
C70S236	AA	4	2	27	CoQH2
C70S236	AB	5	3	27	CoQH2
C70S236	BB	3	3	27	CoQH2
C29226S281	AA	0	0	32	CoQH2
C29226S281	AB	7	2	32	CoQH2
C29226S281	BB	4	2	32	CoQH2
C70S236	AA	2	1	32	CoQH2
C70S236	AB	4	2	32	CoQH2
C70S236	BB	5	2	32	CoQH2

Supplementary Fig. S.1. Strip maps (left to right) showing spatial variation in: chlorophyll, PN, PP, SS and secchi depth on the Great Barrier Reef. The maps were generated using spatial prediction maps based on long-term monitoring data gathered by the Australian Institute of Marine Science (http://eatlas.org.au/data/uuid/df7012eb-a23f-4cf2-b92d-6fff77718987).