Signatures of selection in the coral holobiont reveal complex adaptations to inshore environments driven by Holocene climate change

Ira Cooke[*,1,2], Hua Ying[3], Sylvain Forêt[3,4], Pim Bongaerts[5], Jan Strugnell[6,7], Oleg Simakov[8], Jia Zhang[1,9], Matt Field[2,10], Mauricio Rodriguez-Lanetty[11], Sara Bell[12],David Bourne[12,13],Madeleine van Oppen[12,14],Mark A. Ragan[15],David Miller[1,2,9]

[1]Department of Molecular and Cell Biology, James Cook University, Townsville, Queensland, Australia

[2]Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, Townsville, Queensland, Australia

[3]Research School of Biology, Australian National University, Canberra, Australia, ACT

[4]ARC Centre of Excellence for Coral Reef Studies, Australian National University, Canberra, ACT, Australia

[5]California Academy of Sciences, Golden Gate Park, San Francisco

[6]Centre of Sustainable Tropical Fisheries and Aquaculture, James Cook University, Townsville, Queensland, Australia

[7]Department of Ecology, Environment and Evolution, School of Life Sciences, La Trobe University, Melbourne, Australia

[8]Department of Molecular Evolution and Development, University of Vienna, Austria

[9]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD, Australia

[10]Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, Queensland, Australia

[11]Department of Biological Sciences, Florida International University, Miami, FL, USA

[12]Australian Institute of Marine Science, Townsville, Queensland, Australia

[13]Department of Marine Biology and Aquaculture, James Cook University, Townsville, Queensland, Australia

[14]Department of Marine Biology, University of Melbourne, Melbourne, Australia

[15]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

Abstract

The inshore region of the Great Barrier Reef (GBR) became inundated at the start of the Holocene when sea levels rose. Like many inshore reefs worldwide it harbours extensive coral assemblages despite being subject to relatively high turbidity, freshwater input and thermal fluctuations. To investigate how the coral holobiont has adapted to these conditions we used shallow whole genome resequencing of 148 *Acropora tenuis* colonies from five inshore locations to model demographic history, identify signatures of selection and profile symbiont communities. We also assembled the genome of *Acropora tenuis* from long-read sequencing, providing one of the most contiguous and complete coral genomes to date. We show that corals from Magnetic Island are genetically distinct from reefs further north reflecting a Pleistocene (250-600Kya) split, whereas photosymbiont genotypes differed between reefs in a pattern more likely to reflect contemporary (Holocene) conditions. We also identified loci in the coral host genome with signatures of positive selection in the northern population and demonstrate using coalescent simulations that these are unlikely to be accounted for by demographic history. Associated with these we find genes with roles in a complex range of processes including heterotrophy, osmotic regulation, skeletal development and establishment and maintenance of symbiosis. Our results show that *A. tenuis* holobionts from the inshore GBR have been significantly shaped by their environment and provide an example of complex adaptations in response to past climate change.

Introduction

Coastal marine ecosystems worldwide are threatened by the effects of climate change including warming[1,2], acidification, altered patterns of freshwater input and changing species interactions due to range-shifts[3]. Coral reefs are especially vulnerable because their structural integrity depends on the symbiosis between corals and photosymbionts (family Symbiodiniaceae[4]), but this relationship breaks down under stress leading to bleaching and widespread mortality during marine heatwaves[5,6]. The projected severity of these impacts therefore depends on whether coral holobionts (coral, photosymbionts and associated bacteria[7,8]) can adapt fast enough to maintain the integrity of reef ecosystems[9,10]. A key factor that might improve the adaptive outlook for corals is the existence of stress-tolerant populations which could fuel adaptation in conspecifics through assisted or natural gene flow[11–14]. Ironically, many of these resilient coral populations are threatened because they are found in inshore environments[13,15,16] where they face additional pressures from reduced water quality[17–19] that can exacerbate susceptibility to elevated temperature[20,21].

Although several studies have identified genetic correlates of the thermal stress response or divergent loci that characterise thermal resilience[12,22,23] very little is known about adaptive variation related to other traits that may be important for survival under different local environmental conditions. Understanding the totality of local adaptation is important because a wide variety of stresses are likely to accompany climate change and because some local adaptations may explain reduced survival between translocated genotypes[24].

One way to understand the genetic architecture and longer-term adaptive responses of corals is to examine signatures of past selection, but this is challenging, because selection can be difficult to distinguish from demographic factors such as population bottlenecks. Here we propose an approach to address this issue using dense allele frequency information from whole genome sequencing to detect signatures of selection while using explicit demographic modelling to calculate a neutral distribution and control for false positives. A beneficial side effect of the approach is that it also permits simultaneous genotyping of the dominant photosymbiont in each coral colony.

We apply our approach to identify signatures of selection and profile symbiont communities in *Acropora tenuis* populations from inshore reefs of the GBR. This species is widespread throughout the Indo-Pacific where it is common on both inshore and offshore reefs, and has become an important model for research into water quality stress[25,26], adaptive changes in symbiont composition[27] and the feasibility of assisted evolution approaches[28,29]. The reefs sampled in this study provide an ideal context for local adaptation because they were recolonised between 14 and 6Kya when sea levels rose resulting in marine flooding of the GBR shelf[30]. This event changed the available environment for coral growth dramatically, shifting from a single line at the edge of the shelf to the distinct outer and inner reefs we see today[30]. Our results highlight the complex molecular changes associated with corals that have colonised these inshore habitats where estuarine influence and turbidity due to resuspended sediments is relatively high. In addition, we investigate differences between inshore sites with contrasting water quality due to relative proximity to large river systems[31] and find distinct differences in their relationships with photosymbionts.

Results

De-novo assembly of the *Acropora tenuis* genome

To maximise our ability to model demographic history and detect selective sweeps we generated a highly contiguous and complete reference genome for *Acropora tenuis* from PacBio long-read sequencing. The resulting assembly of 487Mb in 614 contigs with an N50 of 2.84Mb, and largest contig of 13.5Mb, represents a considerable improvement in contiguity compared with most previously published coral genomes (Supp Table S1) including a recent assembly from the same species[32]. Importantly, over 80% of the genome was in fragments larger than 1Mb which allowed us to focus purely on large contigs and thereby minimise a range of potential problems including ambiguous alignments from unmerged haplotypes or misassemblies at the ends of contigs. Large genomic fragments also provide the necessary length to capture recent selective sweeps which can distort allele frequencies over centimorgan sized regions[33,34].

Annotation of this genome assembly with Illumina RNA-seq data resulted in 30327 gene models of which over 91% (27825) belonged to one of 14812 ortholog groups (gene families) shared with other cnidarian genomes. Of these gene families, 97% (14403) were shared with other published *Acropora* genomes (*A. millepora* and *A. digitifera*). BUSCO analysis identified a highly complete 91.8% set of metazoan genes including a small fraction

that were duplicated (2.6%) or fragmented (1.8%). A genome size estimate of 470Mb obtained from kmer analysis of short read data with sga.preqc[35] was very close to the final assembly size confirming the high completeness of the assembly and suggesting that there are few artificial duplications remaining after removal of unmerged haplotypes.

The high completeness of our long-read assembly allowed us to resolve a substantially larger proportion of interspersed repeats (41.65%) compared with published acroporid assemblies from short reads (*A. digitifera*[36]: (30.15%), *A. millepora*[37] (31.58%), *A. tenuis*[32] (34.32%). The only other available coral genome with comparable repeat resolution is a recently released long-read assembly for *A. millepora*[38] which identified a very similar proportion (40.69%) of interspersed repeats to our *A. tenuis* assembly. This increased ability to resolve repeats is reflected in a substantially increased proportion of unclassified elements in long-read assemblies (20.8% *A. tenuis*; 21.2% *A. millepora*) compared with short read assemblies (16.4% *A. digitifera*; 16.5% *A. millepora*), but the overall composition of known repeat classes was remarkably similar between *Acropora* genomes (Supp Figure S1).

As an additional check on the accuracy of our *A. tenuis* assembly we compared its high-level genomic organisation with that of its congeneric, *A. millepora* for which a chromosome-scale linkage map is available[38]. As expected given their small evolutionary distance (15Mya[32]), macrosynteny analysis between the two *Acropora* genomes revealed a high degree of collinearity and limited rearrangements between linkage groups (an example is shown with arrows in Figure 1A). For comparison we also examined synteny between *A. millepora* and the much more distantly related anthozoan, *Nematostella vectensis* (greater than 500Mya[39]). Despite this long divergence time the number and basic structure of linkage groups was remarkably well maintained, but with near complete rearrangement within groups (even distribution of points within blocks in Figure 1B).

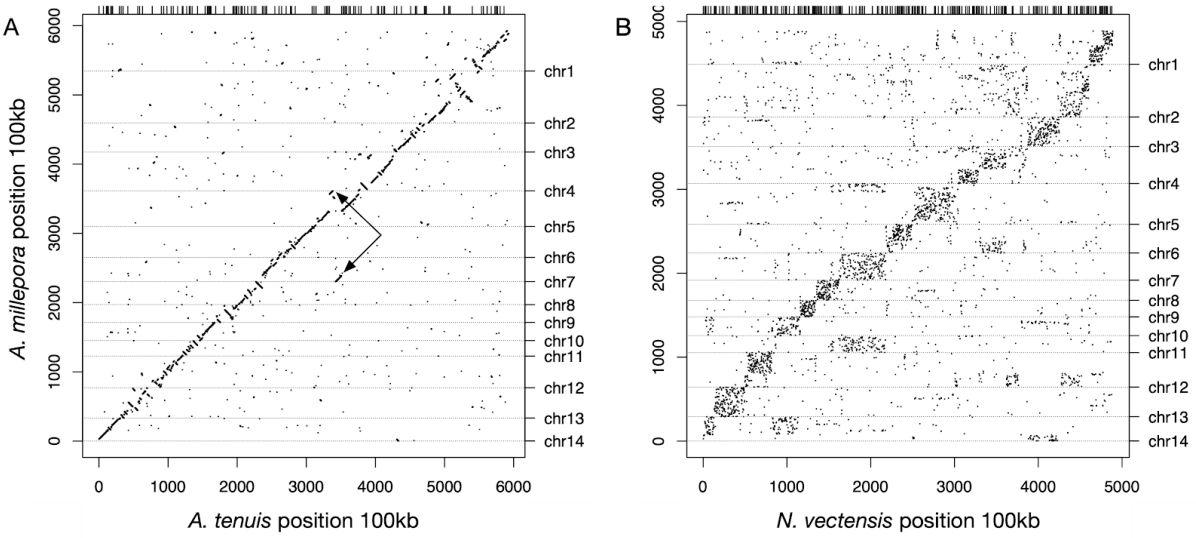

Figure 1: Macrosynteny conservation between phylogenetically close (*A. tenuis*) and distant (*N. vectensis*) relatives of *A. millepora*. Dots represent blocks of collinear groups of genes and margin lines represent boundaries between contigs (top) or *Acropora millepora* linkage groups (right).

Sampling locations and population structure

River discharges create hyposaline, turbid and sometimes nutrient rich conditions which are generally suboptimal for coral growth, and major rivers can therefore have a significant influence on coral reef development[17]. In the central GBR the inshore current flows predominantly northward along the coast, meaning that regions to the north of large river systems such as the Burdekin are heavily influenced by plume waters. To investigate potential adaptive changes in the *A. tenuis* holobiont in relation to water quality we chose reefs from opposite ends of two plume influence gradients, Burdekin in the south and Wet Tropics in the north (Figure 2A). These gradients capture modelled long-term frequency of secondary plume waters which are a strong driver of water quality[40], though many other factors also play a role (see methods for full details and rationale). For the Burdekin gradient we sampled two sites with high river influence ("plume"; Magnetic Island and Pandora Reef) and one site with low riverine influence ("marine"; Pelorus Island), whereas for the Wet Tropics gradient we sampled one "plume" (Dunk Island) and one "marine" site (Fitzroy Island).

Superimposed on these water quality gradients we found strong coral population structure based on latitude. *Acropora tenuis* colonies from Magnetic Island were genetically distinct from those inhabiting the four reefs north of Magnetic Island. Analysis of population structure with PCAngsd showed these as two clear clusters (Supp Figure S2) and admixture analysis showed near complete assignment to either of these clusters for all but three individuals (Figure 2B). We refer to these clusters as north (all reefs other than Magnetic Island) and Magnetic Island. Genome-wide average Fst values between Magnetic Island and northern reefs were much larger (0.19) than between reefs within the northern cluster (0.008). Though clearly distinct, the genetic difference as measured by Fst between Magnetic Island and northern reef populations is within the range of previously discovered geographically structured populations for this species[41]. We also observed evidence of recent gene flow between Magnetic Island and northern reefs with three colonies showing mixed ancestry (Figure 2B). We also checked to determine if this population structure was captured by differences in mitochondrial haplotypes between colonies but found that, in contrast to the nuclear data, mitochondrial haplotypes were completely unstructured (Supp Figure S3). This most likely reflects the relatively recent divergence (see below) combined with a slow mutation rate which is a well-documented feature of anthozoan mitochondrial genomes[42–44].

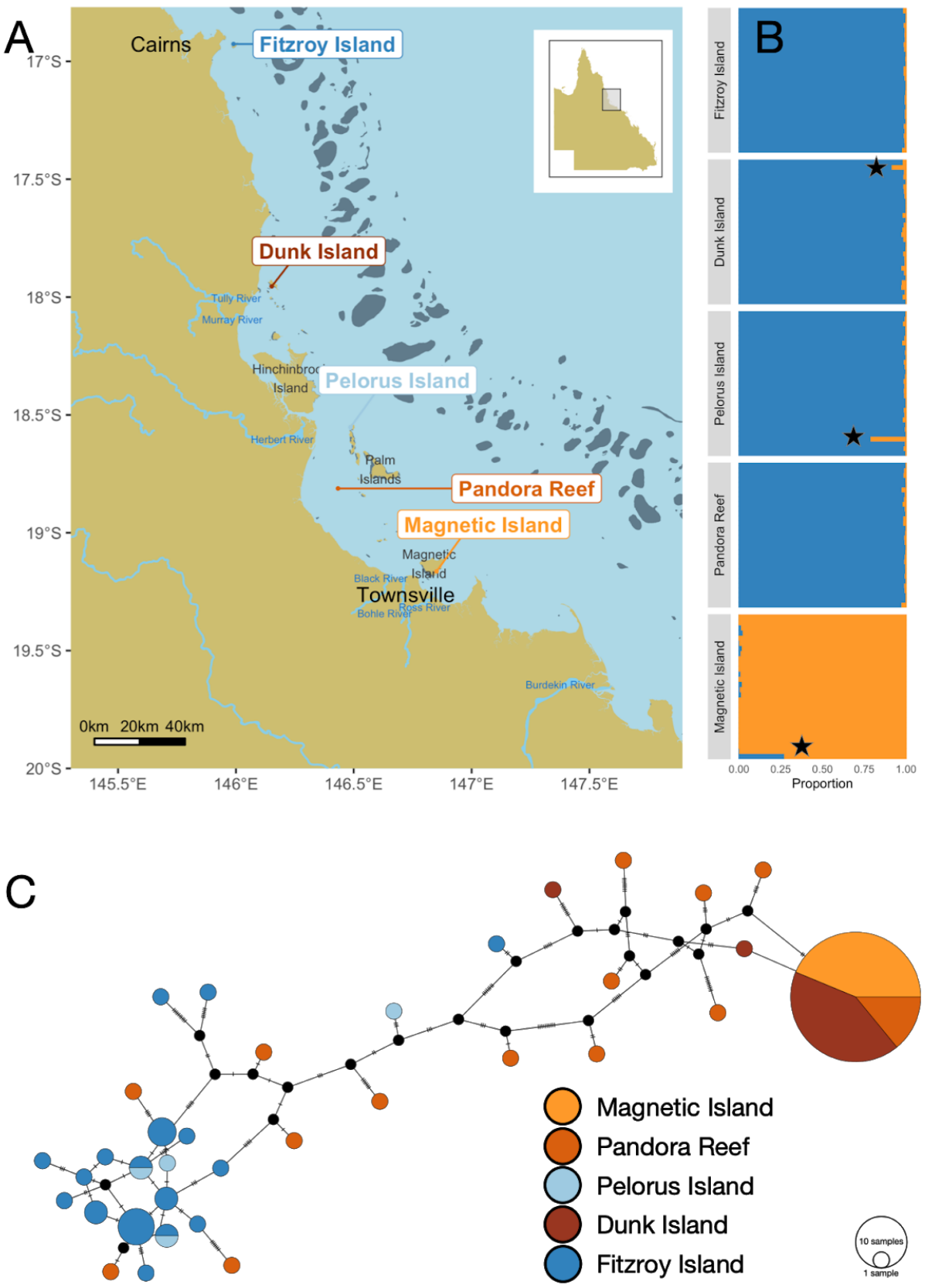

Figure 2: Population structure of coral host and dominant (*Cladocopium*) photosymbionts. Plots A and C use the same colour scheme (bottom right) which shows plume sites in shades of orange and marine sites in shades of blue. **A.** Sampling locations in

the central inshore Great Barrier Reef, Australia. Major river systems are labelled and affect sites northward due to the direction of residual inshore current. **B.** Coral host admixture proportions based on NGSAdmix analysis for each of the five sampling locations. The northern population is shown in dark blue while Magnetic island is in orange. Samples indicated with a black star represent potential hybrids. **C.** Mitochondrial haplotype network based on the dominant strain of the symbiont *Cladocopium* in each sample. Substitutions separating nodes are represented by cross bars.

Demographic history

As essential context to interpreting and controlling for error in selection analyses we first used two complementary approaches, MSMC and ∂a∂i to model demographic history. Combining results from both approaches suggests that the Magnetic Island and northern populations diverged between 600Kya and 270Kya (Figure 3A and 3B). Estimates using ∂a∂i are based on the best fitting model ("isolation_asym_mig"), which allows for a period of isolation followed by secondary contact with asymmetric migration (See Supp Table 2 for full parameter lists of all models). An independent estimate based on divergence of population size trajectories inferred by MSMC suggests a slightly older split with first divergence occurring at around 600Kya becoming clearly distinct by around 300Kya. Both ∂a∂i and MSMC models also suggest that inshore populations of *A. tenuis* have recently declined or have become disconnected from larger parent populations. This is particularly pronounced for Magnetic Island which both modelling approaches estimate as having a higher historical and much smaller contemporary effective population size (Figure 3: also see all alternative ∂a∂i models in Supp Figure S5). Although this bottleneck was consistently observed across all models, MSMC estimates of the timing were much older (~100Kya) than those inferred by the best fitting ∂a∂i model (3Kya). In this instance ∂a∂i estimates are more likely to be reliable since the single diploid genomes used for MSMC (one each from Magnetic Island and Northern populations) provide few recent coalescent events with which to estimate recent demographic change[45,46].

Genome-wide estimates of Tajima's D are sensitive to demographic history and were consistent with the main demographic inferences from ∂a∂i and MSMC (Figure 3C). In particular, a high value for Magnetic Island is consistent with a recent or ongoing bottleneck for that population[47] whereas values close to zero for northern reefs are consistent with little impact from a bottleneck there. Demographic models that included asymmetric migration provided a better fit to the 2D SFS than those without migration or with symmetric migration (Figure3: Supp Table 2). Across all of these models migration coefficients were larger for the southward direction. Although this is against the predominant direction of inshore current flow[48] and predicted movement from biophysical models[49] it is in accordance with studies that have shown southward migration to be the predominant direction of gene flow for *A. tenuis* and *A. millepora* on the GBR[9,49] including for inshore populations. This disconnect between biophysical modelling (which reflects northward inshore currents) and southward gene flow could reflect differences in effective population size. Demographic modelling and genome-wide Tajima's D suggest a strong bottleneck and small effective population size for Magnetic Island, and since migration coefficients in ∂a∂i are measured as a proportion of

individuals at the destination, the small size of the Magnetic Island population could skew outcomes toward southward flow.

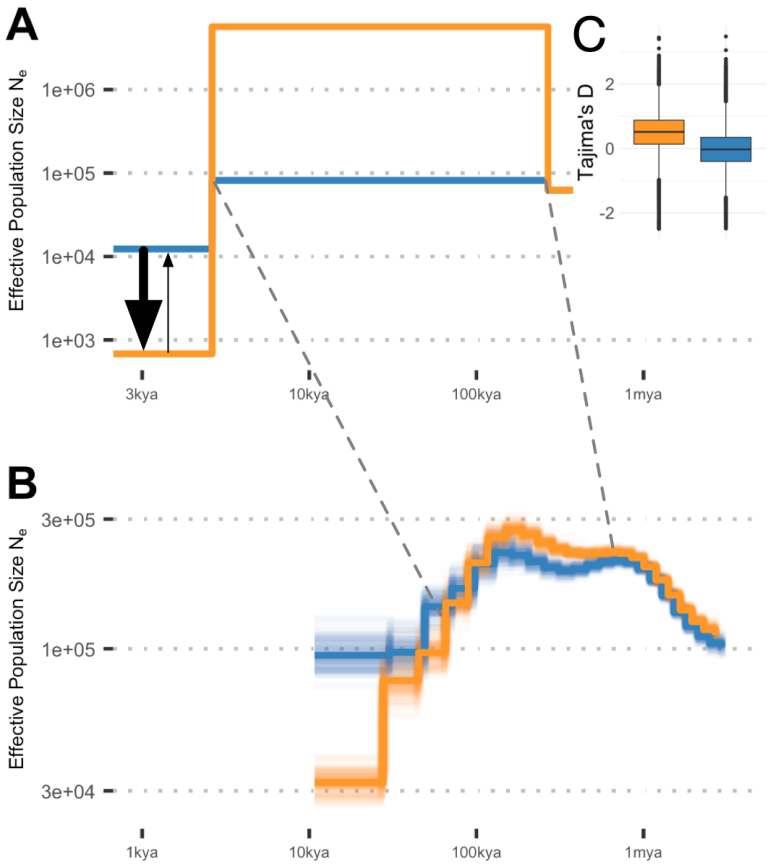

Figure 3: Population genomic statistics and models of demographic history. Orange represents Magnetic Island and blue represents northern populations in all component figures. Dashed lines connect shared events in both models, the point of divergence between populations, and onset of recent bottleneck. **A** Changes in effective population size and estimated gene flow for the best fitting ∂a∂i model. Relative gene flow is shown using arrows with larger flow from north to south indicated by arrow size. **B** Changes in effective population size estimated using MSMC from deep sequencing of a single colony from Magnetic Island (MI) and a colony from Fitzroy Island (FI). **C** (Inset top right) Genome wide estimates of Tajima's D calculated using PCAngst.

Photosymbiont differences between reefs

The availability of complete genomes from five major genera within the Symbiodiniaceae allowed us to identify (using Kraken) between 100k and 300k reads of symbiont (Symbiodiniaceae [4]) origin in each sample (Supp Figure S4). For almost all samples the dominant symbiont genus was *Cladocopium* (formerly Clade C[4]) with a single sample showing a significant proportion of *Durusdinium* (formerly Clade D[4]). Although the number of symbiont reads was insufficient to perform genotyping on nuclear loci it provided sufficient

depth (at least 5x) to determine the dominant mitochondrial haplotype in each sample. The resulting haplotype network appeared to be structured according to location, with the majority of samples from “plume” (Magnetic Island, Dunk Island and Pandora Reef) sharing a single haplotype. Samples from “marine” locations (Fitzroy Island, Pelorus Island) did not include this haplotype but instead were highly diverse, forming a sparse network of divergent types. A test for genetic structuring using AMOVA confirmed that differentiation by reef was significant ($p<1e-5$) but it was not possible to unambiguously link this with the water quality designations from our study design (plume vs marine) ($p=0.1$). Notably, the structuring of photosymbionts by reef followed a distinctly different biogeographical pattern to hosts. Although the results point toward long-term frequency of plume exposure as one potential driver of these differences we cannot rule out other factors such as thermal history or acute differences in plume exposure. Nevertheless, clear structuring by reef suggests adaptation to local conditions.

Signatures of selection

A major challenge when identifying genomic sites under selection is the fact that demographic history and migration can produce confounding signals [50,51]. To estimate false positive rates from such processes SweepFinder 2 was used in combination with coalescent simulations under a range of demographic models (All ∂a∂i and MSMC models) to generate neutral data from which the composite likelihood ratio (CLR) statistic could be calculated. Using this data as a control we were able to calculate an empirical false discovery rate for a given CLR threshold and neutral model (see methods). For northern reefs we found that irrespective of the neutral model used it was always possible to find a CLR threshold that would control false positives to less than 10% (Supp Figure S6). This was not the case for Magnetic Island however, where its extreme demographic history (strong bottleneck) combined with small sample size meant that distinguishing neutral from selective processes was highly model dependent and always required a higher CLR threshold. We therefore focused our attention on the northern reefs where a CLR threshold of 100 was sufficient to control the false discovery rate to less than 10% under the best fitting model and almost all other demographic models. In support of their status as sites under positive selection, the 98 loci passing this threshold also tended to have extremely low values of Tajima’s D (Figure 4) and were often associated with regions of high differentiation (high pairwise Fst) between Magnetic Island and northern reefs (Figure 4). Although SweepFinder 2 is designed to detect classical hard sweeps caused by positive selection, a similar signature can also be produced through persistent background selection[50,52]. Since background selection involves removal of deleterious alleles it should act most strongly on genes with highly conserved functions and is therefore more likely to act similarly in both Magnetic Island and northern populations. Based on this reasoning we removed 20 loci with low Fst values (<0.1) to maintain our focus on genes under positive selection. Finally, we removed a further 7 loci with positive values of Tajima’s D since these could represent genes under balancing selection[47].

Our final list of 71 putative selective sweeps were spread throughout the genome (Figure 5) and overlapped with 61 genes of which 39 could be annotated based on homology or the

presence of conserved domains. A complete list of sweep loci along with associated genes and their functional groupings is provided in supplementary information.

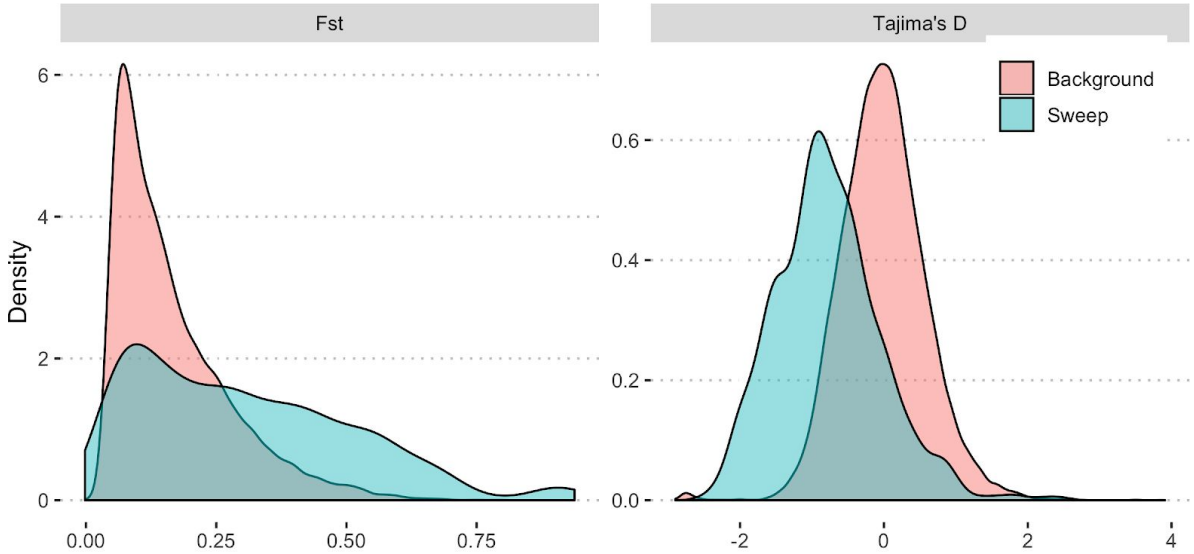

Figure 4: Distribution of genome-wide Fst and Tajima's D for all loci (red) and regions with CLR > 100 (blue) in the northern reefs population .

One of the characteristics of selective sweeps is that the signature of selection decays rapidly after fixation has occurred[53]. Even under ideal conditions (no demographic change, classical hard sweep) the CLR has very little power to detect sweeps that originated more than 0.5Ne generations ago and is best able to detect more recent sweeps (0.05-0.2Ne generations ago)[50]. Our best estimate of contemporary Ne for the northern population comes from *∂a∂i* simulations (Figure 3) and is around 10000. If this is correct it puts an upper bound on sweep ages at 25 Kya with peak detectability between 2500 and 10Kya. This timing suggests that many of the sweep loci identified here could represent signatures of selection that arose as a consequence of dramatic coastal change, including creation of the GBR shelf and inshore environment at the start of the Holocene. Even more recent changes, such as increases in sediment and nutrient flux since European settlement [54,55] , could also produce signatures of selection but our data lacks the resolution to make inferences about this. Key adaptive challenges involved in occupation of the GBR shelf at the start of the Holocene include higher turbidity [56], regular influx of riverine sediment and nutrients[19,57] as well as occasional periods of reduced salinity during severe flooding events [17,58,59]. Our results suggest that significant adaptation to these conditions has been complex, involving selection on a wide range of genes as well as inter-genic loci. Understanding the functional importance of these adaptations is challenging since a significant proportion of loci under selection involve genes that lack homologues in well-studied taxa. This is reflected in the results of Gene Ontology enrichment testing where we searched for GO terms over-represented in sweep genes using topGO[60]. A single, relatively generic term (GO:0005509 calcium ion binding; p = 4.5e-5) associated with EGF domain-containing genes and Skeletal Organic Matrix Proteins (SOMPs) was found to be enriched. To complement this formal analysis we also searched for sweep genes among recent gene

expression and comparative genomic studies in corals and other symbiotic cnidarians. This analysis (detailed below) revealed that many genes associated with sweeps were implicated in roles related to the challenges of the inshore environment, (Figure 5).

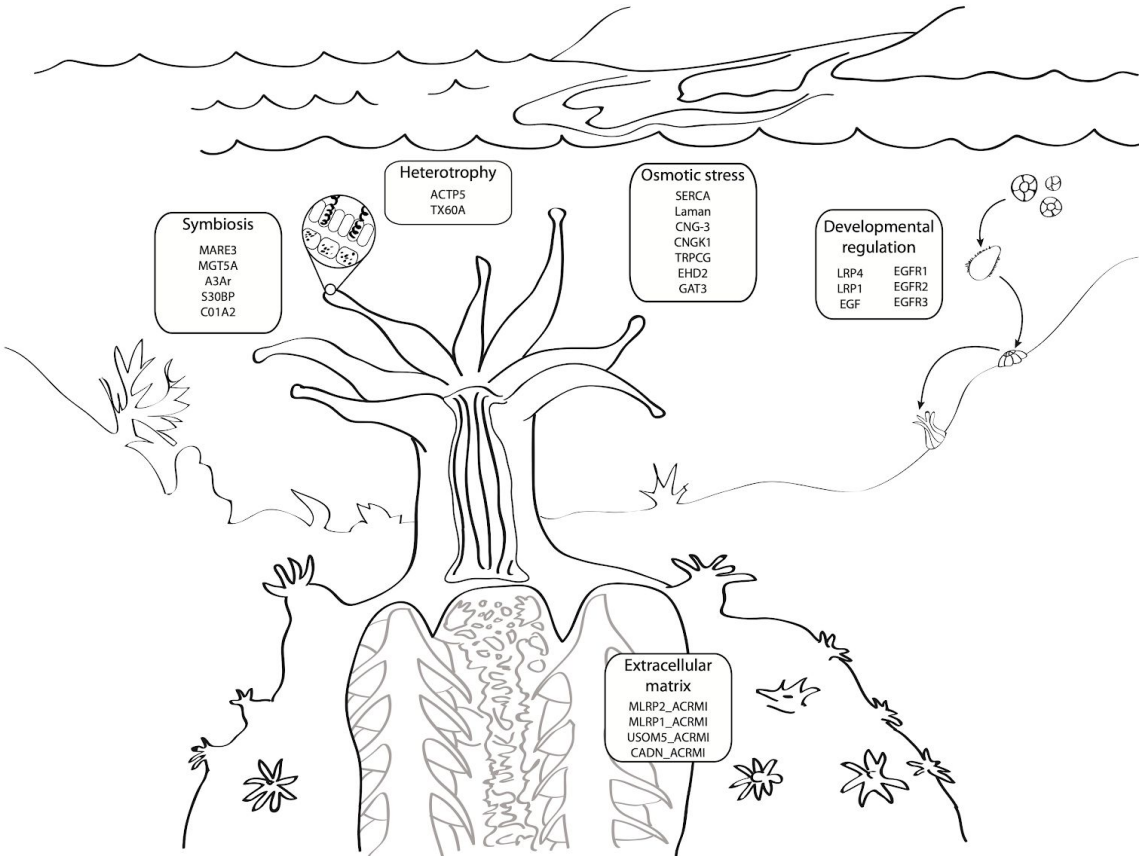

Figure 5: Genes associated with sites under positive selection arranged according to their putative role in adaptations to an inshore environment.

An acute physiological challenge that is particularly important in inshore environments[18] (compared with offshore) is the increased prevalence of stressful [61,62] and potentially lethal [63,64] hypo-saline conditions. A recent RNASeq study found that genes involved in solute transport, amino-acid metabolism and protein recycling were major components of the response to osmotic stress in *Acropora millepora* [65]. Our results suggest that inshore corals have experienced selection associated with genomic regions close to many of these genes. This includes homologs of the calcium and potassium ion transporters SERCA, CNG-3, CNGK1, TRPCG and GAT-3, EHD2 which regulates vesicle transport and Lysosomal alpha-mannosidase (Laman) which is involved in glycoprotein recycling. Notably three of these (SERCA, GAT-3, Laman) were very close homologs (97, 82, 97 percent identity respectively) to proteins differentially expressed in response to osmotic stress in *A. millepora* [65].

Structuring of symbiont genotypes according to local conditions on individual reefs and possibly to water quality (Figure 2C) suggests that the ability to efficiently interact with specific symbionts may be a key requirement for adaptation of the coral holobiont. In support of this we observed 5 genes associated with selective sweeps in the host genome

that were implicated in the establishment and/or maintenance of symbiosis. All of these were homologous to genes observed to be differentially expressed in symbiotic vs aposymbiotic states in *Acropora digitifera* larvae[66], adult corallimorpharians [67] or *Aiptasia*[68], an anemone model for symbiosis. This included genes involved in cytoskeletal remodelling required for symbiosome formation (MARE3, CO1A2), Apoptotic regulators (A3Ar, S30BP). In addition, a larger group of genes under positive selection included those with dual roles in symbiosis and developmental regulation such as low density lipoprotein receptors[68] and a variety of EGF domain containing proteins.

While EGF domain containing proteins are common in animal genomes a specific class of these containing large tandem arrays of calcium binding type EGF domains appeared to be under strong selection in *A. tenuis*. Two of these were associated with exceptionally strong sweep signals adjacent to each other on a 400Kb region that contained a further five EGFRs (Figure 5). Although the large tandem arrays of EGF domains in these proteins resemble those in TGF-beta binding proteins Fibrillin and LTBP they lacked the TB domains that characterise this class [69]. Proteins with various numbers of EGF-like domains have been reported as part of the skeletal organic matrix of both *A. millepora* [70] and *A. digitifera* [71], but developmental gene expression data suggests that they may have a general role in cell adhesion, which then leads to their incorporation in the skeletal organic matrix [71]. Another potential role for these proteins is in the establishment and maintenance of symbiosis with separate studies in *Hydra* [72] and a corallimorpharians [67] both reporting increased expression in a symbiont infected state compared to an aposymbiotic control.

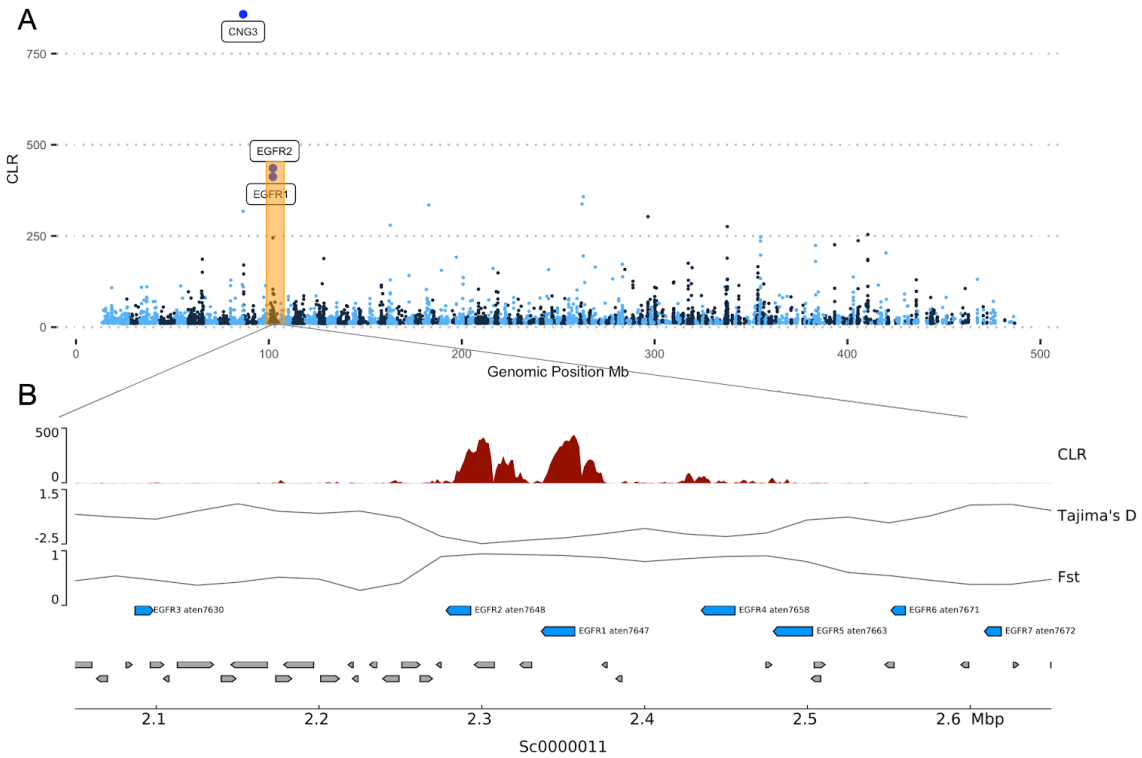

Figure 5: Sites under selection. A Manhattan plot showing genome wide distribution of the SweepFinder 2 CLR statistic. B Detailed view showing a region of the *A. tenuis* genome with evidence of strong selection on two genes encoding EGF repeats (EGFR1, EGFR2). Other EGF repeat containing genes surrounding the sweep are numbered and shown in blue. Other genes are shown in grey.

Heterotrophy has been observed to occur at higher rates for some species in inshore reefs [73,74] and experimentally in response to increased turbidity[75]. In addition, experiments on *Acropora millepora* and *Pocillopora damicornis* showed increased heterotrophic feeding efficiency in colonies from inshore reefs compared with conspecifics sourced from offshore [74]. Since cnidarians including corals rely on toxins injected into prey to assist with heterotrophic feeding [76] an important aspect of this inshore adaptation may involve modifications to the toxin repertoire. In support of this we observed strong selection on two homologs of anemone pore-forming toxins, DELTA actitoxin (aten_0.1.m1.22362) and Delta thalatoxin (aten_0.1.m1.29056). Delta thalatoxin is the major toxin component isolated from nematocysts of the Okinawan sea anemone (*Actineria villosa*)[77] and both toxins have strong haemolytic activity [77,78] which is a feature of coral specific toxins identified so far [79].

Proteins involved in developmental regulation and formation of the skeletal organic matrix constituted a large fraction of genes under selection. In particular this included four of the 36 SOMPs originally identified through proteomics analysis of *Acropora millepora* skeletal materials[70]. This enrichment for proteins that might affect growth and morphology is consistent with work that has shown strong differences in skeletal growth rates and densities for *A. tenuis* along a water quality gradient[31]. A natural extension of this, which has been demonstrated in other species,[80] is that these differences should also exist between inshore

and offshore environments. Our results suggest that skeletal organic matrix proteins may play an important role in the mechanism of this morphological adaptation.

Differential selection between marine and plume environments

Although all four northern reefs are from a single population, strong local selection under low gene flow can lead to differentiation at specific loci[23] which could manifest as independent soft-sweeps in different locations. To search for evidence of local adaptation due to differences in water quality we performed SweepFinder 2 analyses separately on each of the four northern reefs and identified sweeps that were present exclusively in plume or marine habitats. To accommodate a smaller sample size and ongoing gene flow between these reefs we used a slightly relaxed CLR threshold of 50 (FDR < 15%). This revealed seven sweeps that were exclusive to plume sites and eight exclusive to marine (Supp Info). Several genes in both locations had functions related to transcription, probably reflecting differential expression of as-yet unknown genes. In addition, genes associated with sweeps in plume included a Mucin-like protein (MLP_ACRMI) which forms part of the skeletal organic matrix in *Acropora*[70]. Local adaptations related to skeletal organic matrix proteins in plume locations is particularly interesting in light of transplant experiments which showed dramatic skeleton loss when corals from a plume location (Magnetic Island) were transplanted to a marine one (Pelorus Island)[24]. Another interesting protein under selection in plume locations was a homolog of NLRC3 which is an important modulator of Toll-like receptors involved in establishment of symbiosis[81] and which might therefore be involved in maintaining the high symbiont specificity seen in these reefs. Sweeps exclusive to marine included a particularly strong sweep associated with a gene encoding a glutaredoxin domain containing protein, GRCR2. Proteins in this family are upregulated during heat stress in *Acropora palmata*[82] and have been found to be under positive selection in *Acropora*[83].

Discussion

The survival of inshore coral reefs under climate change depends not-only on their ability to adapt to increased thermal stress but also to deal with potential changes in precipitation and drastic changes in water quality that can accompany sea level rise [30,56] as a result of coastal erosion and resuspension of terrigenous sediments. Our results show that *A. tenuis* holobionts from the inshore GBR have been significantly shaped by the inshore environment and fine-scale local differences between reefs on the GBR shelf. The most striking differences at a local scale were found in the photosymbiont association which seemed to be highly specific (single dominant mitochondrial haplotype) at some reefs and much more generic (multiple divergent types) at others. While our study points toward water quality as a driver, additional sampling locations coupled with more detailed long-term environmental data are required to understand the ecological basis for this genetic pattern.

The availability of a high quality reference genome for *A. tenuis* and the use of whole genome resequencing was crucial to our ability to confidently detect genomic regions under selection and to infer demographic history for the host. Although this approach has rarely been used so far in coral population genetics (although see[84]), the moderate size of many coral genomes (~500Mb) and continually falling sequencing costs means that whole genome

sequencing is now cost competitive with reduced representation approaches for population genomics. Our study demonstrates that powerful inferences are possible based purely on allele frequency information available from shallow sequencing but this approach is not without its limitations. In particular, while we were able to set an approximate upper bound on the age of selective sweeps we were not able to provide the precise dating required to differentiate between the ages of different adaptive loci. Such increased resolution could potentially be achieved using deeper sequencing to enable a haplotype-based approach.

Methods

Reference genome sequencing, assembly and annotation

All genome sequencing was based on DNA extracted from sperm collected from a single *A. tenuis* colony collected on 10th of December 2014 from Orpheus Island. An initial round of sequencing was performed on an Illumina HiSeq 2500 instrument using sheared and size selected DNA (~ 450bp) to generate 88.6 million 250bp paired end reads. Further sequencing was then performed on a PacBio RSI instrument by Ramaciotti Center for genomics using 62 SMART Cells resulting in a total of 50.6 Gb of long read data (~108x genome coverage) with read N50 of 15Kb. In addition, to facilitate gene annotation, RNA was extracted from tissues collected from the same colony and sequenced on an illumina HiSeq 2500 instrument. All short-read data is available under ENA accession PRJEB23295.

A draft reference genome was assembled purely from PacBio reads using the FALCON (version 0.7.3) assembler [85] followed by separation of haplotigs from primary assembly with FALCON-unzip and error correction with Quiver [86]. The resulting primary assembly was highly contiguous (N50 ~1320 Kb) but had a larger total size (629Mb) than the haploid size of 470Mb estimated by kmer analysis of Illumina data with sga.preqc [87]. Mapping of transcript sequences to the draft genome assembly revealed the presence of numerous alternate haplotypes in the primary assembly which most likely accounts for the inflated assembly size. To address this issue we used Haplomerger 2 to identify and merge alternate haplotypes. Finally, small redundant scaffolds less than 2kb in length were identified by blast and removed. This resulted in a final merged assembly with a primary size of 487Mb and N50 of 2837 Kb. As any unmerged haplotypes that might remain should be located entirely on the shortest contigs we used only the largest contigs (>1Mb) for population genomic analyses. These large contigs accounted for 81% of the estimated genome size.

Protein-coding gene annotation was performed as previously described [88]. Briefly, de novo and genome-guided transcriptome assembly was carried out using Trinity [89], followed by PSyTrans (https://github.com/sylvainforet/psytrans) to remove Symbiodinium transcripts. Transcripts were then assembled to the genome assembly using PASA [90], from which a set of likely ORFs were generated. Based on their protein coding ability and completeness, these ORFs were carefully assessed to produce a high confident and non-redundant training gene set. This was used to train AUGUSTUS [91] and SNAP [92], and the resulting parameters were employed by the corresponding program in MAKER2 pipeline [93], from which an ab initio

gene model was predicted. Finally, putative transposable elements in the gene model were excluded based on transposonPSI (http://transposonpsi.sourceforge.net) and hhblits [94] search to transposon databases.

Functional annotation of gene models was performed as follows. For each gene in the *A. tenuis* genome, homologues with high quality functional annotations were identified using BLAST searches (E-value threshold 1x10-5) of predicted proteins (BLASTp) and predicted transcripts (BLASTx) against the SwissProt (June 2018) database. Genes were also annotated by searching for conserved domains using InterProScan[95] (version 5.36-75). In addition to finding conserved domains, InterProScan results were also used to infer Gene Ontology based purely on the domains present.

Orthology Analysis

Orthofinder[96] was used to search for orthologous groups of genes shared between gene sets from the following cnidarian genomes, *Acropora digitifera* (Genbank GCF_000222465.1), *Acropora tenuis* (http://aten.reefgenomics.org/), *Acropora millepora* [37], *Fungia sp* (http://ffun.reefgenomics.org/), *Galaxea fascicularis* (http://gfas.reefgenomics.org/), *Nematostella vectensis* (Genbank GCF_000209225.1) and *Stylophora pistilata (*Genbank GCF_002571385.1).

Sampling Design and Rationale for Population Genomics

Water quality is a significant factor influencing the health of inshore coral reefs[55,97] and is subject to intensive monitoring on the GBR[57]. To capture adaptation in the *A. tenuis* holobiont related to water quality we chose reef sites at contrasting ends of two water quality gradients both of which are well characterised in terms of exposure to riverine flood plumes[57,98]. The southern gradient is primarily influenced by secondary plume waters from the Burdekin river which affects reefs to a diminishing extent heading northwards. Predicted numbers of wet season days exposed are Magnetic Island >67%, Pandora Reef 33-67% and Pelorus Island <33%[57]. The northern gradient captures the wet tropics region where multiple rivers influence water quality. Here the plume location, Dunk Island is highly exposed (>67% of wet season days) to plume waters from the Tully and to a lesser extent the Herbert rivers[59,99] while Fitzroy Island receives flood waters from the Johnstone and Russell-Mulgrave river catchments much less often (lower end of 33-67% of wet season days).

Sample collection, sequencing and variant calling

Samples were collected in early 2015 from a total of 148 adult corals taken from five inshore locations on the central Great Barrier Reef (see Figure 2) including Magnetic Island and four northern locations, Fitzroy Island, Dunk Island, Pandora Reef and Pelorus Island. Samples were taken by cutting a branch and then placing it in ethanol for storage. Most samples were collected on a voyage of the RV Cape Ferguson in February (Fitzroy Island 20/2/15, Dunk Island 21/2/15, Pandora Reef 22/2/15 and Magnetic Island (Geoffrey Bay) 23/2/15).

Half the samples from Pelorus Island were collected on 22/2/15 with the remaining 15 samples collected from nearby Orpheus Island (Pioneer Bay) on 24/4/15. Sequencing libraries were prepared separately for each sample and then used to generate four pools for sequencing across four flow cells of a HiSeq 2500 instrument (Ramaciotti Center for Genomics) to obtain a total of 1.2 billion 100bp paired end reads. With the exception of two samples that were chosen for deep sequencing (one from Fitzroy Island and one from Magnetic Island) libraries were pooled evenly so as to obtain approximately 3x coverage per sample.

Raw reads for all samples were pre-processed and mapped against the *A. tenuis* genome using the GATK best practices workflow as follows. Reads passing quality checks were converted to unmapped bam files with sample, lane and flowcell tags added as appropriate. Adapters were marked using Picard (v2.2.1; http://broadinstitute.github.io/**picard**), mapping performed using bwa mem (v0.7.13; [100]) and PCR duplicates marked using Picard (v2.2.1). Variants were then called using Freebayes (v1.0.2-16; [101]) with haplotype calling turned off and ignoring reads with mapping quality less than 30 or sites with read quality less than 20. The two high coverage (20x) samples were down sampled to 3x using samtools to ensure parity with other samples prior to variant calling.

In order to ensure that only high quality variant sites and high quality genotype calls were used for population genomic analyses we performed the following basic filtering steps on raw variant calls from Freebayes. In cases where additional filtering was required it is described under the relevant analysis section. Sites with Phred scaled site quality scores less than 30 were removed. Only biallelic sites were retained. Sites where the mean read depth per individual was less than one third or greater than twice the genome wide average after mapping of 2.7x were removed. Genotypes with Phred scaled genotype quality score less than 20 were set to missing and sites with more than 50% missing genotypes were removed. Sites with highly unbalanced allele read counts (Freebayes allele balance Phred scaled probability > 20) were removed. The mdust program was used to mark low complexity regions and sites within these were removed. SNP sites within 10 base pairs of an indel were removed. Indel sites with length greater than 50 base pairs were removed.

Population structure

To determine whether samples could be divided into genetically distinct clusters we used the program PCAngst (0.973)[102] which has been designed specifically to infer population structure and admixture from low coverage data. As input to PCAngst we used genotype likelihoods at all variant sites after quality filtering (see above), and converted these to Beagle format with ANGSD (0.920)[103]. Individual samples were plotted as a function of the top two principal components based on an eigendecomposition of the genotype relationship matrix calculated by PCAngst (Figure S2). This revealed two distinct clusters corresponding to the Magnetic Island and northern populations which is consistent with the optimal cluster number of 2, also inferred by PCAngst. Admixture proportions for each sample are shown in Figure 2 and show nearly complete assignment to either northern or Magnetic Island populations for most samples. Three samples (MI-1-1, PI-1-16, DI-2-4) showed mixed ancestry and were potential North/Magnetic Island hybrids while PCA analysis revealed two

other samples (MI-2-9, MI-1-16) from Magnetic Island that did not appear to belong to either population cluster. We excluded all potential hybrids as well as these two outlying individuals from selective sweep and population demographic analyses.

Demographic history with PSMC'

Changes in effective population size through time were inferred based on the genome wide distribution of heterozygous sites in two individual colonies sequenced to high coverage, (Fitzroy Island FI-1-3, 18x; Magnetic Island MI-1-4 20x) using PSMC' implemented in the software MSMC (v2.1.2)[45]. In order to avoid biases due to differences in callability of heterozygous sites across the genome we only used these two high coverage samples and also used the snpable (http://lh3lh3.users.sourceforge.net/snpable.shtml) program to mask all regions of the genome where it was not possible to obtain unambiguous read mappings. Variant calling was performed for deep sequenced samples using samtools (1.7) and bcftools (1.9) as these generate outputs in the format required by the bamCaller python script included as part of the msmc-tools package (https://github.com/stschiff/msmc-tools). This script supplements the callability mask (see above) with an additional mask to avoid regions where read coverage is too low (less than half the genome-wide mean) or too high (double the mean) indicating a problematic region such as a collapsed repeat. As a final quality control measure only scaffolds larger than 1Mb (~80% of the genome) were included in MSMC analyses as excessive fragmentation has been shown to bias results [104].

A distribution of MSMC estimates was obtained by generating 100 bootstrap runs, each of which involved recombining a random subsampling of 500kb sized chunks to produce 20 large scaffolds of length 10Mb. Raw outputs from MSMC were converted to real values using a mutation rate of $1.86e^{-8}$ events per base per generation and a generation time of 5 years which is consistent with the fast growth rate and relatively high turn-over of Acroporid corals.

The neutral mutation rate used in the calculation above was estimated by exploiting the equivalence between mutation and substitution rates for well-separated lineages (Kimura 1983). The well studied congeneric species, *Acropora digitifera* was chosen for this comparison because its divergence time to *A. tenuis* has been estimated at 15 million years [32], a value that is sufficient to allow near complete lineage separation but small enough that the number of multiply mutated sites should be small. First, 5583 one to one orthologs shared by *A. digitifera*, *A. tenuis* and *A. millepora* were identified using ProteinOrtho (v5.16b; [105]). Amino acid sequences for *A. tenuis* and *A. digitifera* for these orthologs were then aligned with MAFFT (7.407; [106]) and converted to codon alignments using pal2nal (v14; [107]). Synonymous substitution rates, dS were estimated for each ortholog pair using codeml from the PAML suite (4.9i; [108]). Since dS is a ratio, its values follow a log-normal distribution and we therefore, chose the central value as the mean of log(dS). Finally, the neutral mutation rate, μ was estimated using a divergence time (T) of 15.5Mya, a generation time (G) of 5 years and the formula $\mu = G*dS/(2T)$.

Demographic history with ∂a∂i

Historical population sizes inferred under PSMC' are made under the assumption of a single population without migration and have limited power to resolve changes in the recent past. In order to address these issues, we used the ∂a∂i framework [109] to fit a variety of more complex demographic models (Supplementary Table S2) to the joint site frequency spectrum between Magnetic Island and northern populations. A total of six demographic models were chosen to explore the timing and completeness of separation, population size changes and possible secondary contact between the northern and Magnetic Island populations. To explore the relationship between model fit and complexity we included models ranging from a very simple island model without migration (3 parameters) to a complex model allowing for population sizes before and after population split and asymmetric migration (10 parameters).

In order to prepare inputs for ∂a∂i we created a python script, vcf2dadi (https://github.com/iracooke/atenuis_wgs_pub/blob/master/bin/vcf2dadi.py) to calculate allele frequencies at each site based on observed read counts for each allele encoded in the Freebayes generated vcf file. The script rejects sites with fewer than 50% called genotypes and calculates allele frequencies by using the ratio of total observations for each allele to infer the number of alleles of each type present among called samples. This method takes advantage of the high overall read depth across all samples to accurately impute allele frequencies for each population without using potentially biased low coverage genotype calls. Prior to running the vcf2dadi script further filtering was performed on the vcf file to remove sites with minor allele count less than 2, and to thin sites to a physical separation distance of at least 1000bp. This produced a final dataset consisting of 266k SNPs in ∂a∂i format. These were imported to ∂a∂i and converted to a folded 2D site frequency spectrum and projected down from the original sample size of 56 x 240 alleles to 45 x 180. This resulted in a final dataset that included 15.5k segregating sites.

All ∂a∂i models were optimised in a hierarchical manner to minimise the possibility of obtaining best-fit parameters corresponding to a local optimum. All models used generic starting parameters to begin ten independent runs, where each run involved four successive rounds with the best fit parameters from the previous round undergoing perturbation prior to being used as input for the next round.

Signatures of selection

A genome-wide search for selective sweeps was performed using SweepFinder 2 [50,110] which calculates a test statistic for each site in the genome that reflects the likelihood of observing SNP data under the assumption of a sweep compared with a null model where no sweep has occurred. In order to prepare inputs for SweepFinder 2 we used the same vcf2dadi script described under the section "demographic history with ∂a∂i", with an option to generate outputs in SweepFinder format. The site frequencies generated by this script were then used to generate a combined frequency spectrum across the entire genome which SweepFinder 2 uses as an empirical neutral model. Each population was then scanned for sweeps by calculating the SweepFinder test statistic on a genome-wide grid every 1000bp. To avoid treating adjacent loci as independent sweeps we then ran another python script sf2gff.py (https://github.com/iracooke/atenuis_wgs_pub/blob/master/bin/sf2gff.py) which identifies contiguous sweep regions and converts them to gff format. In this script a sweep

region is deemed to start when the sweepfinder statistic rises above a value of 10 and ends when it falls below this value. Within the region the maximum value of the test statistic is taken to represent the strength of the sweep. This resulted in 5315 putative sweeps which were then subject to significance testing (see below).

Since the SweepFinder test statistic is a composite likelihood it requires an appropriate null model to determine a significance threshold. In order to address this problem and also to control for potential false positives due to demographic factors we used the program, ms [111] to generate simulated allele frequency data for all demographic scenarios modelled using ∂a∂i as well as the demographic history inferred by MSMC analyses. Neutral model data simulated with ms was subjected to SweepFinder 2 analysis to calculate a null distribution of the composite likelihood ratio statistic (CLR) for each demographic model. At a given threshold value, T, of the CLR it is then possible to calculate the false discovery rate as the ratio of the number of loci with (CLR>T) in the neutral model versus the Magnetic Island or northern population data respectively. For the northern population a threshold CLR value of 100 or greater resulted in an FDR of less than 10% under any of the demographic models except for 'sym_mig_size' which was not the best fitting model and predicted a deep bottleneck followed by population expansion. For Magnetic Island no such threshold could be found indicating that the influence of neutral factors was too large to allow genuine sweeps to be detected. For this reason we did not proceed further with SweepFinder analyses using the Magnetic Island data.

For the northern population all sweeps with a CLR value greater than a conservative significance threshold of 100 were identified and Bedtools (2.7.1[112]) was used to find genes that lay within 1000bp of these.

Folded vs unfolded frequency spectra

We used the folded site frequency spectrum (SFS) for both dadi and SweepFinder 2 analyses. Although both methods can benefit from provision of an unfolded SFS this requires an accurate determination of the ancestral state for each allele. This is a particular challenge for *A. tenuis* because other *Acropora* species that could potentially be used as an outgroup (*Acropora digitifera*, *Acropora millepora*) are separated by a relatively large evolutionary distance (15 Million years [32]). With such distantly related outgroup taxa there is a high likelihood that the outgroup sequence will not represent the ancestral state, but will instead capture independently fixed mutations in each lineage. We therefore decided to conduct all analyses with folded SFS in this study, which reduces power, but avoids the risk of false inferences due to incorrect ancestral assignment.

Population genetic statistics, Fst, Tajima's D

In order to avoid biases due to inaccurate genotype calls in low coverage data we calculated Fst and Tajima's D using ANGSD [103] (version 0.938) which calculates these parameters without calling genotypes [113]. All calculations were performed based on the 14 million variant sites identified after quality filtering on Freebayes calls (see section on variant calling).

Mitochondrial Genomes

Host consensus mitochondrial genome sequences were obtained for each sample as follows. Reads were mapped against the published *Acropora tenuis* mitochondrial genome sequence[114] (genbank accession AF338425) using bwa mem (v0.7.17). Mapped mitochondrial reads for all samples constituted at least 25x coverage and were used to call a consensus sequence for each sample using bcftools (v1.9). Full length consensus mitochondrial genome sequences for all samples were used to build a TCS [115] type haplotype network and visualised using the software, PopArt [116] (v1.7)

Mitochondrial genome sequences for photosymbionts were obtained in a similar way. Reads previously identified as not mapping to the host genome were mapped against the *Cladocopium goreaui* mitochondrial sequences available from the reefgenomics website (http://symbs.reefgenomics.org/download/). Mapped reads were then used to call a consensus sequence using bcftools in the same manner as for host mitogenomes. Overall mitogenome coverage was much lower and more variable for symbiont genomes than for host genomes reflecting a larger mitogenome sequence (55kb) and reduced number of reads. We therefore removed samples with fewer than 2500 reads (~5x coverage) to avoid obtaining consensus calls from very low coverage. The resulting mitogenome sequences from 107 remaining samples were loaded into Geneious Prime (2019.2.3) and only sites with no ambiguous bases for any sample were retained. The resulting 6170bp segment was used to construct a TC haplotype network shown in Figure 2 using PopArt [116] (v1.7).

Acknowledgements

We would like to thank Dr. Anthony Bellantuono, Dr. Camila Granados-Cifuentes and Ms. Katherine Dougan for early access to the *Durusdinium trenchii* genome.

This project was supported by a Queensland Government DSITIA Accelerate Partnerships award to the University of Queensland on behalf of the Australian Institute of Marine Science, the Australian National University, Bioplatforms Australia, the Great Barrier Reef Foundation, the Great Barrier Reef Marine Park Authority, and James Cook University (2014).

This research/project was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

The data used in this project were funded by the Great Barrier Reef Foundation's Resilient Coral Reefs Successfully Adapting to Climate Change research and development program in collaboration with the Australian Government, Bioplatforms Australia through the National Collaborative Research Infrastructure Strategy (NCRIS), Rio Tinto and a family foundation. The authors also acknowledge the work done by the Reef Future Genomics (ReFuGe) 2020 Consortium organised by the Great Barrier Reef Foundation.

Supplementary Information

Detailed methods including shell scripts used to run command-line analyses and R code used to generate plots is available from https://github.com/iracooke/atenuis_wgs_pub

https://docs.google.com/document/d/1EQqTLdvePd5iJTfdy88UVoY8aOG-ssX23HNv7AvGh0g/edit?usp=sharing

1. Holbrook, N. J. *et al.* A global assessment of marine heatwaves and their drivers. *Nat. Commun.* **10**, 2624 (2019).
2. Hughes, T. P. *et al.* Coral reefs in the Anthropocene. *Nature* **546**, 82–90 (2017).
3. Doney, S. C. *et al.* Climate change impacts on marine ecosystems. *Ann. Rev. Mar. Sci.* **4**, 11–37 (2012).
4. LaJeunesse, T. C. *et al.* Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts. *Curr. Biol.* **28**, 2570–2580.e6 (2018).
5. Heron, S. F., Maynard, J. A., van Hooidonk, R. & Eakin, C. M. Warming Trends and Bleaching Stress of the World's Coral Reefs 1985-2012. *Sci. Rep.* **6**, 38402 (2016).
6. Hughes, T. P. *et al.* Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017).
7. Robbins, S. J. *et al.* A genomic view of the reef-building coral Porites lutea and its microbial symbionts. *Nat Microbiol* **4**, 2090–2100 (2019).
8. Pollock, F. J. *et al.* Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat. Commun.* **9**, 4921 (2018).
9. Matz, M. V., Treml, E. A., Aglyamova, G. V. & Bay, L. K. Potential and limits for rapid genetic adaptation to warming in a Great Barrier Reef coral. *PLoS Genet.* **14**, e1007220 (2018).
10. Hughes, T. P. *et al.* Climate Change, Human Impacts, and the Resilience of Coral Reefs. *Science* **301**, (2003).

11. Anthony, K. *et al.* New interventions are needed to save coral reefs. *Nat Ecol Evol* **1**, 1420–1422 (2017).
12. Dixon, G. B. *et al.* CORAL REEFS. Genomic determinants of coral heat tolerance across latitudes. *Science* **348**, 1460–1462 (2015).
13. Schoepf, V. *et al.* Stress-resistant corals may not acclimatize to ocean warming but maintain heat tolerance under cooler temperatures. *Nat. Commun.* **10**, 4031 (2019).
14. Morikawa, M. K. & Palumbi, S. R. Using naturally occurring climate resilient corals to construct bleaching-resistant nurseries. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10586–10591 (2019).
15. Kenkel, C. D., Setta, S. P. & Matz, M. V. Heritable differences in fitness-related traits among populations of the mustard hill coral, Porites astreoides. *Heredity* **115**, 509–516 (2015).
16. Richards, Z. T., Garcia, R. A., Wallace, C. C., Rosser, N. L. & Muir, P. R. A diverse assemblage of reef corals thriving in a dynamic intertidal reef setting (Bonaparte Archipelago, Kimberley, Australia). *PLoS One* **10**, e0117791 (2015).
17. Furnas, M. J. *Catchments and Corals: Terrestrial Runoff to the Great Barrier Reef*. (Australian Institute of Marine Science, 2003).
18. Coles, S. L. & Jokiel, P. L. Effects of salinity on coral reefs. in *Pollution in Tropical Aquatic Systems* (eds. Connell, D. W. & Hawker, D. W.) 147–166 (CRC Press, Florida, 1992).
19. Brodie, J. E., Devlin, M., Haynes, D. & Waterhouse, J. Assessment of the eutrophication status of the Great Barrier Reef lagoon (Australia). *Biogeochemistry* **106**, 281–302 (2011).
20. Fisher, R., Bessell-Browne, P. & Jones, R. Synergistic and antagonistic impacts of suspended sediments and thermal stress on corals. *Nat. Commun.* **10**, 2346 (2019).
21. Wooldridge, S. A. Water quality and coral bleaching thresholds: formalising the linkage

for the inshore reefs of the Great Barrier Reef, Australia. *Mar. Pollut. Bull.* **58**, 745–751 (2009).

22. Kenkel, C. D. & Matz, M. V. Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nat Ecol Evol* **1**, 14 (2016).
23. Bay, R. A. & Palumbi, S. R. Multilocus adaptation associated with heat resistance in reef-building corals. *Curr. Biol.* **24**, 2952–2956 (2014).
24. Rocker, M. M. & Brandl, S. J. Transplantation of corals into a new environment results in substantial skeletal loss in Acropora tenuis. *Mar. Biodivers.* **45**, 321–326 (2015).
25. Humanes, A., Ricardo, G. F., Willis, B. L., Fabricius, K. E. & Negri, A. P. Cumulative effects of suspended sediments, organic nutrients and temperature stress on early life history stages of the coral Acropora tenuis. *Sci. Rep.* **7**, 44101 (2017).
26. Humanes, A., Noonan, S. H. C., Willis, B. L., Fabricius, K. E. & Negri, A. P. Cumulative Effects of Nutrient Enrichment and Elevated Temperature Compromise the Early Life History Stages of the Coral Acropora tenuis. *PLoS One* **11**, e0161616 (2016).
27. Ulstrup, K. E. & Van Oppen, M. J. H. Geographic and habitat partitioning of genetically distinct zooxanthellae (Symbiodinium) in Acropora corals on the Great Barrier Reef. *Mol. Ecol.* **12**, 3477–3484 (2003).
28. Quigley, K. M., Bay, L. K. & van Oppen, M. J. H. The active spread of adaptive variation for reef resilience. *Ecol. Evol.* **9**, 11122–11135 (2019).
29. Chakravarti, L. J., Beltran, V. H. & van Oppen, M. J. H. Rapid thermal adaptation in photosymbionts of reef-building corals. *Glob. Chang. Biol.* **23**, 4675–4688 (2017).
30. Hinestrosa, G., Webster, J. M. & Beaman, R. J. Spatio-temporal patterns in the postglacial flooding of the Great Barrier Reef shelf, Australia. *Cont. Shelf Res.* **173**, 13–26 (2019).
31. Rocker, M. M., Francis, D. S., Fabricius, K. E., Willis, B. L. & Bay, L. K. Variation in the health and biochemical condition of the coral Acropora tenuis along two water quality

gradients on the Great Barrier Reef, Australia. *Mar. Pollut. Bull.* **119**, 106–119 (2017).

32. Mao, Y., Economo, E. P. & Satoh, N. The Roles of Introgression and Climate Change in the Rise to Dominance of Acropora Corals. *Curr. Biol.* **28**, 3373–3382.e5 (2018).
33. Hernandez, R. D. *et al.* Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
34. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C. & Clark, A. G. Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* **8**, 857–868 (2007).
35. Simpson, J. T. Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* **30**, 1228–1235 (2014).
36. Shinzato, C. *et al.* Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* **476**, 320–323 (2011).
37. Ying, H. *et al.* The Whole-Genome Sequence of the Coral Acropora millepora. *Genome Biol. Evol.* **11**, 1374–1379 (2019).
38. Fuller, Z. L. *et al.* Population genetics of the coral Acropora millepora: Towards a genomic predictor of bleaching. *bioRxiv* 867754 (2019) doi:10.1101/867754.
39. Zapata, F. *et al.* Phylogenomic Analyses Support Traditional Relationships within Cnidaria. *PLoS One* **10**, e0139068 (2015).
40. Devlin, M. J. *et al.* Water Quality and River Plume Monitoring in the Great Barrier Reef: An Overview of Methods Based on Ocean Colour Satellite Data. *Remote Sensing* **7**, 12909–12941 (2015).
41. Rosser, N. L. Demographic history and asynchronous spawning shape genetic differentiation among populations of the hard coral Acropora tenuis in Western Australia. *Mol. Phylogenet. Evol.* **98**, 89–96 (2016).
42. Hellberg, M. E. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evol. Biol.* **6**, 24 (2006).
43. Shearer, T. L., Van Oppen, M. J. H., Romano, S. L. & Wörheide, G. Slow mitochondrial

DNA sequence evolution in the Anthozoa (Cnidaria). *Mol. Ecol.* **11**, 2475–2487 (2002).

44. van Oppen, M. J., Willis, B. L. & Miller, D. J. Atypically low rate of cytochrome b evolution in the scleractinian coral genus Acropora. *Proc. Biol. Sci.* **266**, 179–183 (1999).
45. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
46. Patton, A. H. *et al.* Contemporary Demographic Reconstruction Methods Are Robust to Genome Assembly Quality: A Case Study in Tasmanian Devils. *Mol. Biol. Evol.* **36**, 2906–2921 (2019).
47. Maruyama, T. & Fuerst, P. A. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**, 675–689 (1985).
48. Thomas, C. Modelling marine connectivity in the Great Barrier Reef and exploring its ecological implications. (Ecole Polytechnique de Louvain, 2015).
49. Riginos, C. *et al.* Asymmetric dispersal is a critical element of concordance between biophysical dispersal models and spatial genetic structure in Great Barrier Reef corals. *Diversity and Distributions* **0**, (2019).
50. Huber, C. D., DeGiorgio, M., Hellmann, I. & Nielsen, R. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol. Ecol.* **25**, 142–156 (2016).
51. Stephan, W. Genetic hitchhiking versus background selection: the controversy and its implications. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 1245–1253 (2010).
52. Pavlidis, P. & Alachiotis, N. A survey of methods and tools to detect recent and strong positive selection. *J. Biol. Res.* **24**, 7 (2017).
53. Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002).

54. Roff, G. *et al.* Palaeoecological evidence of a historical collapse of corals at Pelorus Island, inshore Great Barrier Reef, following European settlement. *Proc. Biol. Sci.* **280**, 20122100 (2013).

55. De'ath, G., Fabricius, K. E., Sweatman, H. & Puotinen, M. The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17995–17999 (2012).

56. Webster, J. M. *et al.* Response of the Great Barrier Reef to sea-level and environmental changes over the past 30,000 years. *Nat. Geosci.* **11**, 426–432 (2018).

57. Thompson, A. *et al. Marine Monitoring Program. Annual Report for Inshore water quality and coral reef monitoring 2013-2014*. http://elibrary.gbrmpa.gov.au/jspui/bitstream/11017/2975/5/MMP_AIMS_Inshore_WQ_Coral_Monitoring_2013-2014%20v2.pdf (2014).

58. Schroeder, T. *et al.* Inter-annual variability of wet season freshwater plume extent into the Great Barrier Reef lagoon based on satellite coastal ocean colour observations. *Mar. Pollut. Bull.* **65**, 210–223 (2012).

59. Devlin, M. & Schaffelke, B. Spatial extent of riverine flood plumes and exposure of marine ecosystems in the Tully coastal region, Great Barrier Reef. *Mar. Freshwater Res.* **60**, 1109–1122 (2009).

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

61. Kerswell, A. P. & Jones, R. J. Effects of hypo-osmosis on the coral Stylophora pistillata: nature and cause of 'low-salinity bleaching'. *Marine Ecology Progress Series* **253**, 145–154 (2003).

62. Ferrier-Pagès, C., Gattuso, J. P. & Jaubert, J. Effect of small variations in salinity on the rates of photosynthesis and respiration of the zooxanthellate coral Stylophora pistillata.

Marine Ecology Progress Series **181**, 309–314 (1999).

63. Jokiel, P. L., Hunter, C. L., Taguchi, S. & Watara P, L. Ecological impact of a fresh-water 'reef kill' in Kaneohe Bay, Oahu, Hawaii. *Coral Reefs* **12**, 177184 (1993).
64. Van Woesik, R., De vantier, L. M. & Glazebrook, J. S. Effects of Cyclone 'Joy' on nearshore coral communities of the Great Barrier Reef. *Marine Ecology Progress Series* **128**, 261–270 (1995).
65. Aguilar, C. *et al.* Transcriptomic analysis reveals protein homeostasis breakdown in the coral Acropora millepora during hypo-saline stress. *BMC Genomics* **20**, 148 (2019).
66. Mohamed, A. R. *et al.* The transcriptomic response of the coral Acropora digitifera to a competent Symbiodinium strain: the symbiosome as an arrested early phagosome. *Mol. Ecol.* **25**, 3127–3141 (2016).
67. Lin, M.-F., Takahashi, S., Forêt, S., Davy, S. K. & Miller, D. J. Transcriptomic analyses highlight the likely metabolic consequences of colonization of a cnidarian host by native or non-native Symbiodinium species. *Biol. Open* **8**, (2019).
68. Oakley, C. A. *et al.* Symbiosis induces widespread changes in the proteome of the model cnidarian Aiptasia. *Cell. Microbiol.* **18**, 1009–1023 (2016).
69. Robertson, I., Jensen, S. & Handford, P. TB domain proteins: evolutionary insights into the multifaceted roles of fibrillins and LTBPs. *Biochem. J* **433**, 263–276 (2011).
70. Ramos-Silva, P. *et al.* The skeletal proteome of the coral Acropora millepora: the evolution of calcification by co-option and domain shuffling. *Mol. Biol. Evol.* **30**, 2099–2112 (2013).
71. Takeuchi, T., Yamada, L., Shinzato, C., Sawada, H. & Satoh, N. Stepwise Evolution of Coral Biomineralization Revealed with Genome-Wide Proteomics and Transcriptomics. *PLoS One* **11**, e0156424 (2016).
72. Hamada, M. *et al.* Metabolic co-dependence drives the evolutionarily ancient Hydra-Chlorella symbiosis. *Elife* **7**, (2018).

73. Fox, M. D. *et al.* Gradients in Primary Production Predict Trophic Strategies of Mixotrophic Corals across Spatial Scales. *Curr. Biol.* **28**, 3355–3363.e4 (2018).

74. Anthony, K. R. N. Enhanced particle-feeding capacity of corals on turbid reefs (Great Barrier Reef, Australia). *Coral Reefs* **19**, 59–67 (2000).

75. Anthony, K. R. & Fabricius, K. E. Shifting roles of heterotrophy and autotrophy in coral energetics under varying turbidity. *J. Exp. Mar. Bio. Ecol.* **252**, 221–253 (2000).

76. Schmidt, C. A., Daly, N. L. & Wilson, D. T. Coral Venom Toxins. *Frontiers in Ecology and Evolution* **7**, 320 (2019).

77. Oshiro, N. *et al.* A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone Actineria villosa. *Toxicon* **43**, 225–228 (2004).

78. Pungerčar, J., Anderluh, G., Maček, P., Franc, G. & Štrukelj, B. Sequence analysis of the cDNA encoding the precursor of equinatoxin V, a newly discovered hemolysin from the sea anemone Actinia equina1The sequence data presented here have been submitted to the EMBL/GenBank Databases under the accession number U51900.1. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1341**, 105–107 (1997).

79. Ben-Ari, H., Paz, M. & Sher, D. The chemical armament of reef-building corals: inter- and intra-specific variation and the identification of an unusual actinoporin in Stylophora pistilata. *Sci. Rep.* **8**, 251 (2018).

80. Lough, J. M. & Barnes, D. J. Comparisons of skeletal density variations in Porites from the central Great Barrier Reef. *J. Exp. Mar. Bio. Ecol.* **155**, 1–25 (1992).

81. Mohamed, A. R. *et al.* Deciphering the nature of the coral-Chromera association. *ISME J.* **12**, 776–790 (2018).

82. DeSalvo, M. K., Sunagawa, S., Voolstra, C. R. & Medina, M. Transcriptomic responses to heat stress and bleaching in the elkhorn coral Acropora palmata. *Mar. Ecol. Prog.*

Ser. **402**, 97–113 (2010).

83. Voolstra, C. R. *et al.* Rapid evolution of coral proteins responsible for interaction with the environment. *PLoS One* **6**, e20392 (2011).
84. Shinzato, C., Mungpakdee, S., Arakaki, N. & Satoh, N. Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago. *Sci. Rep.* **5**, 18211 (2015).
85. Chin, C.-S. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
86. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
87. Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
88. Ying, H. *et al.* Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. *Genome Biol.* **19**, 175 (2018).
89. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
90. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
91. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–9 (2006).
92. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
93. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
94. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2011).

95. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

96. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

97. Fabricius, K., De'ath, G., McCook, L., Turak, E. & Williams, D. M. Changes in algal, coral and fish assemblages along water quality gradients on the inshore Great Barrier Reef. *Mar. Pollut. Bull.* **51**, 384–398 (2005).

98. Álvarez-Romero, J. G. *et al.* A novel approach to model exposure of coastal-marine ecosystems to riverine flood plumes based on remote sensing techniques. *J. Environ. Manage.* **119**, 194–207 (2013).

99. Angly, F. E. *et al.* Marine microbial communities of the Great Barrier Reef lagoon are influenced by riverine floodwaters and seasonal weather events. *PeerJ* **4**, e1511 (2016).

100. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

101. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]* (2012).

102. Meisner, J. & Albrechtsen, A. Inferring Population Structure and Admixture Proportions in Low Depth Next-Generation Sequencing Data. *BiorXiv* (2018) doi:10.1101/302463.

103. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

104. Gower, G. *et al.* 1 Population size history from short genomic scaffolds: how short 2 is too short? doi:10.1101/382036.

105. Lechner, M. *et al.* Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).

106. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

107. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–12 (2006).

108. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

109. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genet.* **5**, e1000695 (2009).

110. Nielsen, R. *et al.* Genomic scans for selective sweeps using SNP data. *Genome Res.* **15**, 1566–1575 (2005).

111. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).

112. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

113. Fumagalli, M. *et al.* Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* **195**, 979–992 (2013).

114. van Oppen, M. J. H. *et al.* The mitochondrial genome of Acropora tenuis (Cnidaria; Scleractinia) contains a large group I intron and a candidate control region. *J. Mol. Evol.* **55**, 1–13 (2002).

115. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659 (2000).

116. Leigh, J. W. & Bryant, D. popart : full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).