Received: 17 April 2024 | Revised: 28 June 2024 | Accepted: 3 July 2024

DOI: 10.1111/mec.17468

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Chromosomal inversions harbour excess mutational load in the coral, *Acropora kenti*, on the Great Barrier Reef

Jia Zhang[1,2,3] | Nadja M. Schneller[1,2,4] | Matt A. Field[1,2,5] | Cheong Xin Chan[3] | David J. Miller[1,2] | Jan M. Strugnell[2,6,7] | Cynthia Riginos[8,9] | Line Bay[9] | Ira Cooke[1,2]

[1]College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Queensland, Australia

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

[3]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia

[4]School of Biological Sciences, Monash University, Melbourne, Victoria, Australia

[5]Immunogenomics Lab, Garvan Institute of Medical Research, Sydney, New South Wales, Australia

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

[7]College of Science and Engineering, James Cook University, Townsville, Queensland, Australia

[8]School of the Environment, The University of Queensland, Brisbane, Queensland, Australia

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

Correspondence
Ira Cooke, College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Qld, Australia.
Email: ira.cooke@jcu.edu.au

Handling Editor: Maren Wellenreuther

Abstract

The future survival of coral reefs in the Anthropocene depends on the capacity of corals to adapt as oceans warm and extreme weather events become more frequent. Targeted interventions designed to assist evolutionary processes in corals require a comprehensive understanding of the distribution and structure of standing variation, however, efforts to map genomic variation in corals have so far focussed almost exclusively on SNPs, overlooking structural variants that have been shown to drive adaptive processes in other taxa. Here, we show that the reef-building coral, *Acropora kenti*, harbours at least five large, highly polymorphic structural variants, all of which exhibit signatures of strongly suppressed recombination in heterokaryotypes, a feature commonly associated with chromosomal inversions. Based on their high minor allele frequency, uniform distribution across habitats and elevated genetic load, we propose that these inversions in *A. kenti* are likely to be under balancing selection. An excess of SNPs with high impact on protein-coding genes within these loci elevates their importance both as potential targets for adaptive selection and as contributors to genetic decline if coral populations become fragmented or inbred in future.

KEYWORDS
adaptation, conservation genetics, coral, structural variation

1 | INTRODUCTION

Coral reefs are hyperdiverse marine ecosystems that provide crucial ecosystem services to millions of people throughout the tropics. Threats to coral reefs, particularly from ocean warming have placed the evolutionary biology of corals into the spotlight because it is now clear that their long-term survival depends on whether they can adapt to keep pace with climate change (DeFilippo et al., 2022; Logan et al., 2014; Matz et al., 2020; McManus et al., 2021). Decision-making in relation to protected area design (Colton et al., 2022), genetic interventions (Anthony et al., 2017) and reef restoration (DeFilippo et al., 2022) must therefore be informed by a sound understanding of the factors that shape genetic diversity in corals (Baums et al., 2019).

Over the past two decades, adoption of population genomic approaches has greatly improved our understanding of evolutionary processes in corals. Using SNP and microsatellite markers many studies have identified instances of fine-scale population structure and cryptic speciation (Bongaerts et al., 2021; Ladner & Palumbi, 2012; Matias et al., 2022; Rippe et al., 2021; Rose et al., 2021). More recently, the adoption of dense SNP marker sets and whole-genome sequencing has started to reveal the origins and drivers of divergence (Thomas et al., 2022; Zhang et al., 2022) as well as the architecture of key traits such as heat tolerance (Fuller et al., 2020). So far, however, all population genomic work in corals has relied exclusively on SNPs, ignoring structural variants, such as inversions. Inversions are a particularly important form of structural variation because they can be large (multiple Mb in genomic extent) and can strongly suppress recombination between inverted and ancestral arrangements (Kirkpatrick, 2010). These characteristics make them potent evolutionary modifiers that can facilitate local or clinal adaptive processes (Berdan et al., 2023; Kapun et al., 2023; Wellenreuther & Bernatchez, 2018), often have strong phenotypic effects (Joron et al., 2011; Wang et al., 2013) and may capture a large fraction of the standing genetic variation in some species (Harringmeyer & Hoekstra, 2022; Mérot et al., 2021).

The fact that inversions inhibit recombination over large genomic regions provides a mechanism for local adaptation under gene flow (Schluter & Rieseberg, 2022) because an inversion that captures a locally favourable combination of alleles will be protected from recombination with less favourable alleles on the alternate arrangement (Kirkpatrick & Barton, 2006; Tigano & Friesen, 2016). Initially predicted from simple theoretical models (Kirkpatrick & Barton, 2006), this idea is now supported by many studies across diverse taxa (Huang et al., 2020; Meyer et al., 2023; Wang et al., 2013; Wellenreuther & Bernatchez, 2018) in which alternate arrangements of an inversion are found to diverge in frequency between ecotypes. Inversion polymorphisms that become established via such spatially divergent selection have received considerable attention in the literature, perhaps because they can easily be detected as large blocks of sharply elevated F_{ST} between ecotypes (Schaal et al., 2022), and because they often play a role in local adaptation (Faria et al., 2019).

A recent review (Faria et al., 2019) describes these inversions as type I, with the defining characteristic being positive (or indirectly positive) selection for a locally adapted arrangement leading to divergent arrangement frequencies between ecologically or spatially separate populations. There is now increasing recognition that many inversion polymorphisms are maintained by selection that directly favours the heterokaryotype and thus allows an inversion to persist at a stable frequency throughout the population (type II). Several recent studies have shown that accumulation of recessive deleterious mutations within inversions can directly favour the heterokaryotype (Anderson et al., 2005; Ayala et al., 2013; Fabian et al., 2012; Mérot et al., 2021), preventing inversions from reaching fixation and supporting long-term persistence at high frequency throughout the population. Since inversions maintained by this mechanism need not segregate strongly across ecotypes they may be more challenging to identify, and their ecological roles less obvious, however, evidence is emerging in support of the idea that inversions of this type (type II; Faria et al., 2019) may form an important reservoir of genetic variation over and above that of the collinear genome. Recent theoretical work (Berdan et al., 2021) and empirical observations in sunflowers (Huang et al., 2022) suggest that type II inversions may accumulate mutational load at a higher rate than type I inversions because the relative lack of homokaryotypes inhibits purging of deleterious alleles. Since selective forces on inversions may change over time, the reservoir of variation accumulated within type II inversions may eventually become a target for positive selection. This idea is supported by a recent global analysis (Kapun et al., 2023) of the In(3R)Payne inversion in *Drosophila melanogaster* which occurs as a balanced polymorphism in its ancestral African population but has now formed sharp latitudinal clines underlying climate adaptation in North America (Fabian et al., 2012) and Australia (Anderson et al., 2005).

Although large polymorphic inversions have now been observed across a wide variety of taxa (Ayala et al., 2013; Fuentes-Pardo et al., 2023; Harringmeyer & Hoekstra, 2022; Mérot et al., 2021; Todesco et al., 2020), their role in corals, and more broadly in cnidarians, remains an open question. Evidence from comparative genomics has shown that several large inversions have accumulated between species of the genus *Acropora* (Locatelli et al., 2023); however, the prevalence of polymorphic inversions within cnidarian species is unknown and might vary widely between species, as it does in *Drosophila* (Hoffmann & Rieseberg, 2008). There is growing consensus that some important traits such as bleaching tolerance in corals are likely to be controlled by many unlinked genes of small effect (Fuller et al., 2020); however, inversions would complicate this paradigm because they can capture multiple loci that collectively have a large influence on adaptive and speciation processes (Berdan et al., 2023). Since inversions can also harbour recessive deleterious variation, it is possible that they could play a role in genetic decline, for example, if a population becomes fixed for an inversion after a bottleneck, exposing recessive phenotypes via homozygosity. A comprehensive understanding of the contribution of inversions and other structural variants to standing genetic

variation in corals is therefore required in order to project their capacity to adapt to climate change, forecast the consequences of genetic decline and plan genomic interventions.

We set out to identify and characterize inversions and other genome-wide patterns of genetic variation, in the reef-building coral *Acropora kenti*, sampled from five inshore and four offshore reefs along a 500 km section of the central Great Barrier Reef (GBR). *Acropora kenti* (Brook, 1892) (previously referred to as *Acropora tenuis* but now resurrected as an accepted species; Bridge et al., 2023) has been the focus of genetic, developmental and physiological studies for many years (Abrego et al., 2009; Berry et al., 2016; Matias et al., 2022) and is the target of research on potential genomic interventions to mitigate climate change impacts (Quigley & van Oppen, 2022). Our study sites were chosen to include a contrast between inshore and offshore reefs which are spatially separated but likely to experience gene flow. Inshore reefs experience higher turbidity, more variable temperatures and much greater terrestrial influence (nutrients, pollution and freshwater runoff) than offshore (Brodie et al., 2011; Coles & Jokiel, 1992; Furnas, 2003). Evidence from previous studies suggests that there are also differences in dominant algal symbionts harboured by corals at different reefs in this region (Abrego et al., 2009; Matias et al., 2022). While these environmental gradients provide selective pressures that might promote the formation of locally adapted ecotypes, they are also subject to high levels of gene flow which would oppose local adaptation. Our goal was to identify and characterize any polymorphic inversions present within this *A. kenti* population, determine their roles (if any) in promoting local adaptation and as contributors to standing genetic variation.

Using patterns of heterozygosity, linkage disequilibrium and local genetic structure, we identified at least five inversions circulating at high frequency (MAF > 0.17) and ranging in size from 0.2 to 2 Mb. None of these five showed patterns of elevated differentiation between inshore/offshore reefs or between colonies with different dominant symbionts that would demonstrate a role in local adaptation. Instead, our analyses show that all these inversions are highly polymorphic, and harbour an excess of mutations with high impacts on protein-coding genes. This combination of characteristics is most consistent with inversions that are maintained by forms of balancing selection that directly favour the heterokaryotypic state such as associative overdominance (i.e. type II inversions). Our results highlight the fact that structural variants such as inversions may be prevalent in coral populations and play a significant role in structuring standing genetic variation.

2 | RESULTS

2.1 | Population structure and symbiont diversity

To facilitate detection and characterization of inversions in *A. kenti*, we used the ANGSD framework to call 3.8 million biallelic SNPs from shallow (2–5×) whole-genome sequencing data across 208 genetically distinct colonies sampled from nine reefs in the central Great Barrier Reef (GBR) (Figure 1a). Full details of SNP calling and quality control are given in methods. Analyses of population structure and admixture with PCAngsd showed that eight of these reef populations formed a single genetic cluster that was distinct from Magnetic Island (Figure 1b). The strong distinction of Magnetic Island has been observed in previous whole-genome sequencing (WGS) studies (Cooke et al., 2020; Matias et al., 2022) and is thought to reflect divergence around 500 kya–1 Mya. Despite this divergence, we observed four highly admixed individuals, indicative of recent crosses between Magnetic Island and other reefs (Figure 1c). The eight non-Magnetic Island reefs in our study include one location (Pelorus Island) that overlaps with locations dominated by Cluster 1A identified by Matias et al (Matias et al., 2022), implying that this lineage occurs across the full length of the GBR and can be found in both inshore and offshore locations (Matias et al., 2022).

Among the eight non-Magnetic Island reefs, we found no clear evidence of genetic structure between reefs or between inshore versus offshore locations. This was evident in a PCA based on non-Magnetic Island data (Figure S4) and a tree inferred from the identity-by-state (IBS) distance matrix between all pairs of samples (Figure S5). Since inshore and offshore samples were sequenced in separate batches, such lack of structure suggests that batch effects (if present) are likely to have minimal influence on population genetic inferences. Nevertheless, we found higher variability in individual heterozygosity estimates for inshore samples than for offshore (Figure S7), suggesting that some minor batch effects may be present despite stringent data quality filters and a common sequencing platform used for all samples. To minimize uncertainties arising from possible sequencing batch effects, our results focus on patterns of genetic structure that occur within batches, or which are genomically localized, and therefore robust to genome-wide differences in sequencing data.

Taxonomic analysis of raw sequencing reads with Kraken (Figure S8) showed that with few exceptions all colonies harboured *Cladocopium* as the dominant symbiont. Further investigation of symbiont reads mapping to the *Cladocopium proliferum* (*syn. goreaui*) (Butler et al., 2023) genome revealed two distinct mitochondrial haplotypes (Figure 2; Figure S9) and two major clusters based on the d2s kmer-based distance metric (Figure S10). The geographic distribution of samples harbouring the two mitochondrial haplotypes suggests that they correspond to *Cladocopium* types, C1 and C2, previously identified using SSCP polymorphisms (Abrego et al., 2009; Quigley et al., 2020) in *A. kenti* at Magnetic Island (C1; 22/28 of MI samples) and Pelorus Island (C2; 18/30 of PI samples). While the C1 haplotype was strongly associated with three inshore reefs (MI, PR and DI), the C2 haplotype was more broadly distributed across inshore and offshore locations.

2.2 | Identification of inversion loci and genotyping of individuals for inversion karyotypes

Absent or weak (see below) population structure among the eight non-Magnetic Island reefs allowed us to identify putative

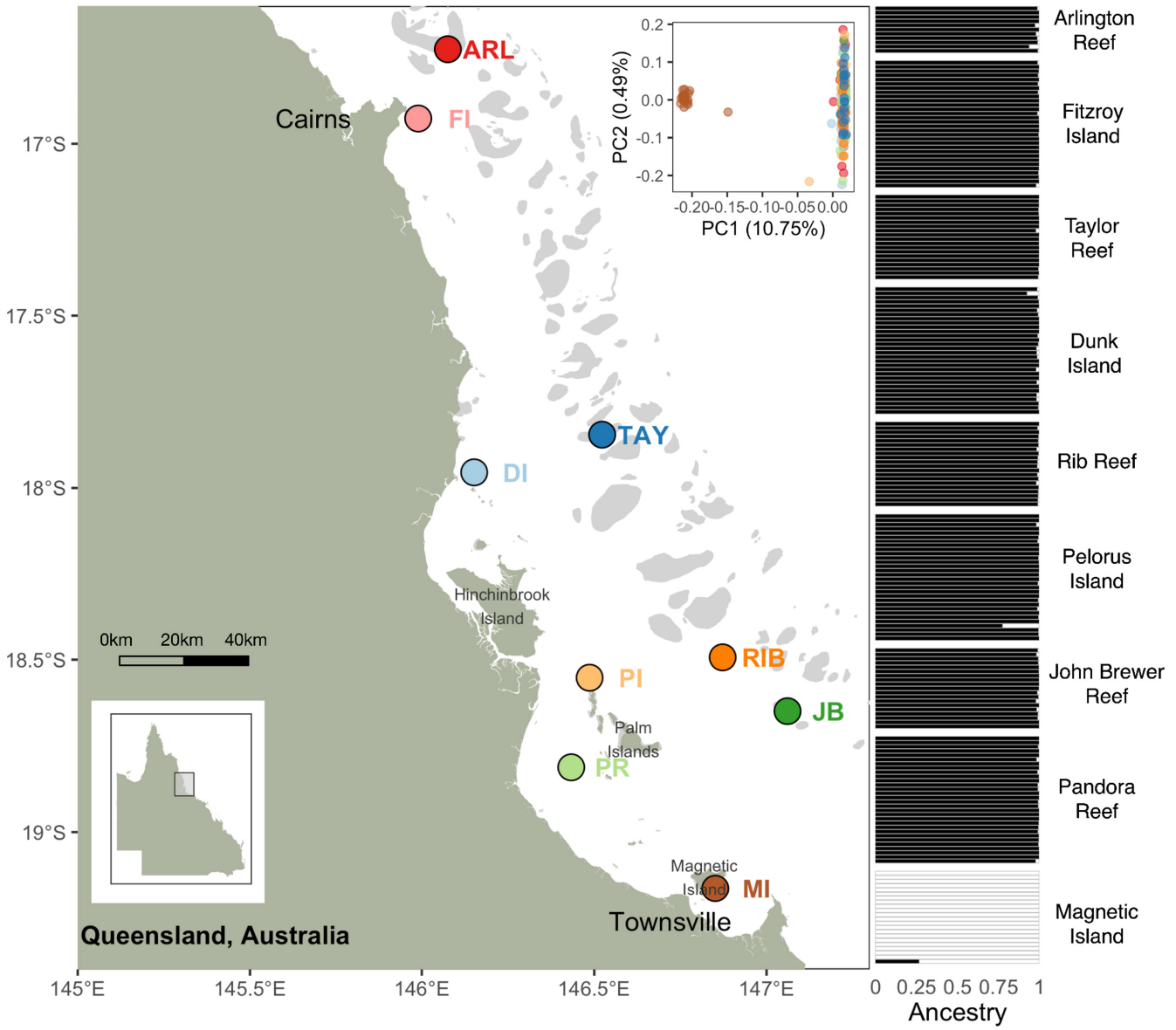

FIGURE 1 Population structure and sampling locations within the central Great Barrier Reef. Inset shows a principal components analysis with two main clusters (Magnetic Island: Left, Northern Reefs: Right) for which the ancestry proportions are shown for each individual. Offshore reefs sequenced in this study are as follows: Arlington Reef (ARL), Taylor Reef (TAY), Rib Reef (RIB) and John Brewer Reef (JB). Inshore reefs (sequenced in Cooke et al. 2020) are as follows: Fitzroy Island (FI), Dunk Island (DI), Palm Islands (PI), Pandora Reef (PR) and Magnetic Island (MI).

inversion polymorphisms using a scan for local genetic structure based on principal component analysis (Meisner et al., 2021). This scan (performed using PCAngsd on genotype likelihoods) revealed five genomic regions ranging in size from 200 kb to 2 Mb exhibiting exceptionally strong population structure compared with the genomic background (Figure 3a). A similar scan for the Magnetic Island population failed to reveal any clear peaks in the strength of local population structure (Figure S11); however, this probably reflects a lack of statistical power rather than absence of inversions at Magnetic Island (Figure S11; Section 4). All five genomic regions with strong local population structure in the non-Magnetic Island population also exhibited patterns of heterozygosity and linkage disequilibrium indicative of strongly suppressed recombination in heterokaryotypes, a feature often considered to be diagnostic for inversions (Harringmeyer & Hoekstra, 2022; Mérot et al., 2021), although chromosomal fusions and translocations may generate similar signatures (Mérot et al., 2020). Firstly, visual inspection of population structure within each region revealed three major clusters along PC1 (Figure 3b; Figure S12), as expected based on the three possible genotypes of an inversion polymorphism (Harringmeyer & Hoekstra, 2022; Huang et al., 2020). Heterozygosity within each locus was highest in corals assigned to the central cluster (corresponding to heterokaryotypes) (Figure 3c; Figure S12), which is an expected consequence of sharply reduced recombination between inverted alleles. Finally, linkage disequilibrium was elevated within the inverted region

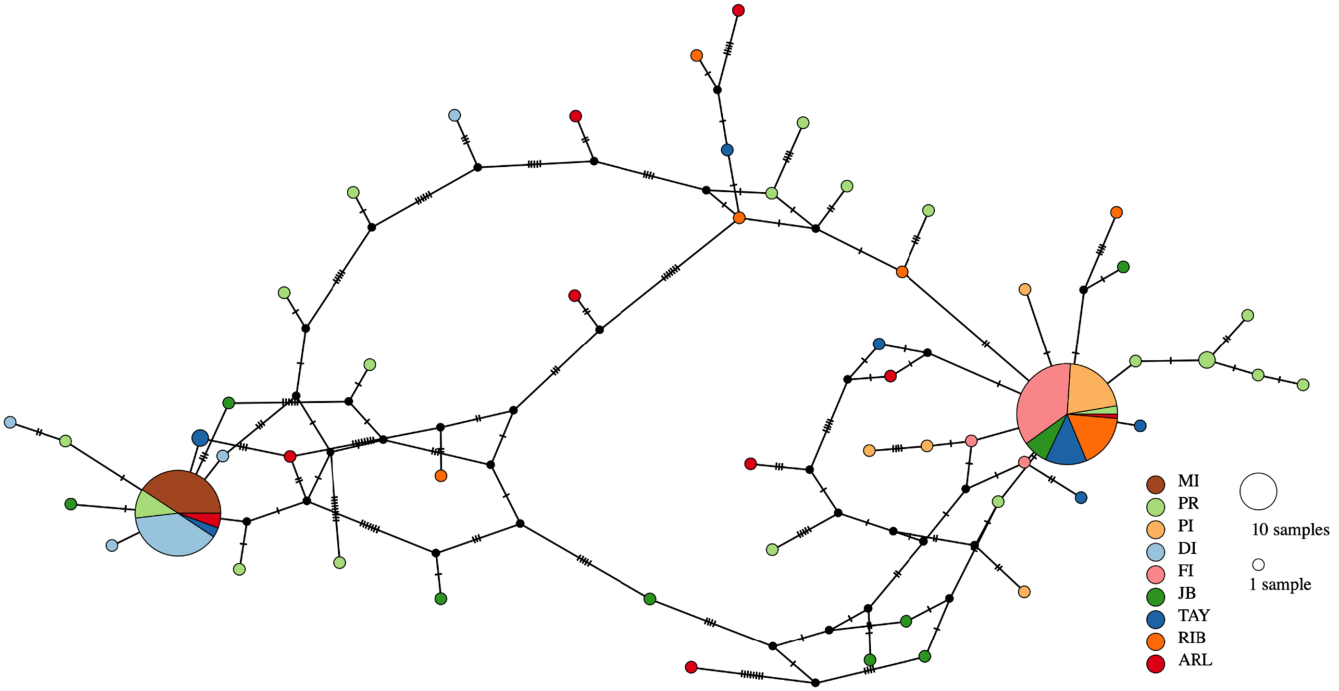

FIGURE 2 Symbiont mitochondrial haplotype network based on consensus sequences of reads mapping to the *Cladocopium proliferum* (*syn. goreaui*) mitochondrial genome within coral whole-genome sequencing data. Individual haplotypes are shown as separate nodes with node size reflecting the number of samples. Edges connect related nodes with cross bars indicating the number of different sites. Sample location abbreviations and colours are the same as in Figure 1.

for heterokaryotypes (Figure 3d) but less so for homokaryotypes. This LD effect was clearly observable for putative inversions L1 and L2 but not for inversions L3–L5 likely due to their smaller sizes (0.2–0.5 Mb) and occurrence in less-contiguous regions of the assembled genome (Figure S13).

In addition to genetic evidence for inversions presented in Figure 3, we used the structural variant detection software Manta (Chen et al., 2016) to obtain direct read-based evidence for inversion events for two samples for which we had sequencing data at greater depth (~20×). One of these deeper sequenced samples (FI-1-3) was from the non-Magnetic Island population, and the other from Magnetic Island (MI-1-4). PCA-based genotyping (e.g. Figure 3b) for the non-Magnetic Island individual indicated that it was heterozygous for the L1, L3 and L4 inversions; however, of these only the L1 inversion was contained within a single scaffold of our assembly and therefore suitable for analysis with Manta. In this individual, short-read data showed clear support for a 1.2 Mb inversion event (called by Manta and manually verified; Figure S14) that closely matched the L1 region identified via local genetic structure (Figure 3). Although it was not possible to genotype the Magnetic Island individual (MI-1-4) using local PCA due to the small sample size at that location, short-read data confirmed that it was also heterozygous for the L1 inversion (Figure S14).

All five inversion loci detectable via our local PCA-based method were present at high frequency (0.17–0.34) and were polymorphic across all reef sites (Figure 4a). None deviated significantly from genotype proportions expected under Hardy–Weinberg equilibrium ($p > .19$) as might be expected under strong spatially divergent selection (excess of homokaryotypes), or if recessive mutations resulted in a lethal homokaryotype (excess of heterokaryotypes). To investigate mutational load, we used SnpEff (Cingolani et al., 2012) to predict the severity of impact of SNPs on protein-coding genes. SnpEff predicts coding effects such as start/stop codon gains or losses, frameshifts and amino acid substitutions, and classifies these into categories ranging from low (e.g. synonymous variant) to high impact (e.g. frameshift and premature stop codon). Ignoring intergenic regions to avoid confounding effects from gene density, we compared the predicted impact of SNPs within inversions and contrasted this with an equal number of SNPs from a random selection of 100×50 kb regions scattered throughout the genome. We found that the predicted impact of SNPs within inversions was shifted towards higher values compared with the genomic background (Figure 4b). Variants in the highest impact category are predicted to cause major disruptions to the protein sequence and are therefore likely to be associated with deleterious effects. These highly disruptive variants were more abundant in inversions across all allele frequency classes.

2.3 | Signatures of selection are not associated with inversion loci

Since inversions can facilitate divergent selection under gene flow (Berdan et al., 2023; Kirkpatrick & Barton, 2006), we identified two pairs of contrasting environmental conditions against which such

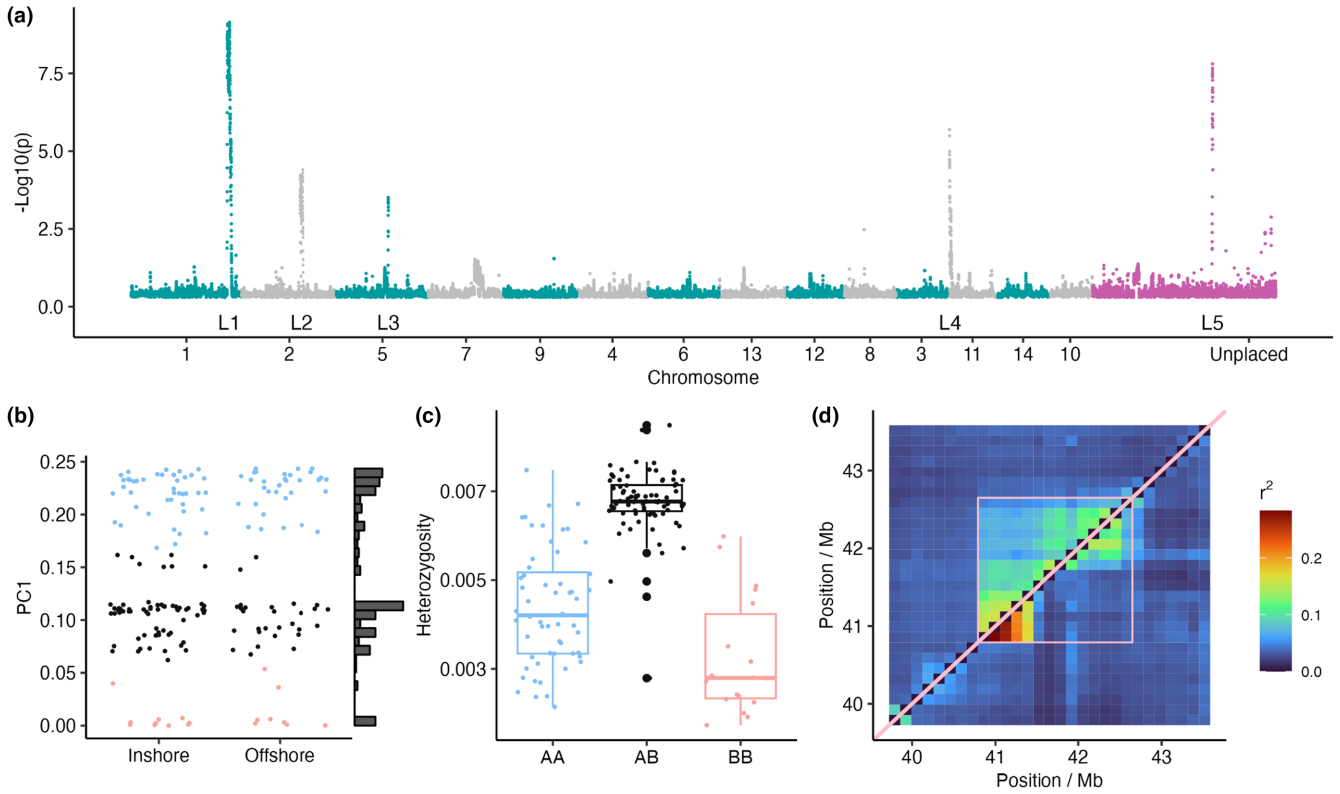

FIGURE 3 Location and characteristics of highly polymorphic inversions in *Acropora kenti*. (a) Manhattan plot showing *p*-values derived from Galinsky statistics indicative of local population structure with five putative inversions (L1–L5) visible as strong signals compared with the background. Genomic coordinates reflect placement of *A. kenti* scaffolds via alignment to the *Acropora millepora* chromosome-level genome. Alternating green and grey points indicate chromosome membership while pink points are on unplaced scaffolds. Plots b–d show the hallmarks of an inversion for L1 (similar plots for all other inversions are provided in Figures S12 and S13). (b) Strong local population structure with three clusters along PC1 (explains 34.6% of variance). (c) Individual heterozygosity within the L1 region with individual corals genotyped for the L1 inversion according to their corresponding cluster in b. (d) Pairwise linkage ($r^{2\prime}$ statistic) across a 3 Mb region centred on the L1 inversion. Top diagonal represents 93 heterozygous individuals and bottom diagonal 76 individuals homozygous for the major arrangement. Each grid square shows the average value for all relevant SNPs. Pink box delineates inversion bounds inferred from PCAngsd (part a).

divergence would be expected to occur in our study; (DeFilippo et al., 2022) inshore versus offshore sites which differ strongly in nutrient, thermal and turbidity regimes (Brodie et al., 2011; Coles & Jokiel, 1992; Furnas, 2003), and (McManus et al., 2021) individuals clearly harbouring C1- or C2-dominant symbiont communities (most common haplotypes shown in Figure 2). Analyses with AMOVA failed to identify any significant associations between allele frequencies of any of the five inversions and reef, shore (inshore vs. offshore) or dominant symbiont (Table S5).

To identify sites potentially under divergent selection in the non-Magnetic Island population independent of inversions, we performed genome-wide scans of pairwise F_{ST} between inshore and offshore, and between C1- and C2-dominant colonies. We found that none of the five inversions overlapped with these highly divergent regions (F_{ST} *z*-score>6) and average F_{ST} within all inversions was within 3 SD of the mean in all cases. In addition, levels of absolute divergence (D_{xy}) were not elevated in these highly differentiated regions, indicating that they are not regions of locally restricted gene flow (islands of speciation; Cruickshank & Hahn, 2014) or inversions that we did not detect via our PCA-based method. Further analysis of highly differentiated regions in the non-Magnetic Island population revealed that they did not generally have the characteristics of strong recent selective sweeps (reduced Tajima's *D* and reduced D_{xy}; Figure 5b,d).

Despite a strong background level of divergence reflecting historical separation (Cooke et al., 2020), we identified two exceptionally differentiated regions between Magnetic Island and non-Magnetic Island populations (Figure 5e). Tajima's *D* in both these highly differentiated regions was sharply reduced (Figure 5f) compared with the genomic background indicating that divergence in these regions is associated with strong linked selection (selective sweeps). One of these regions overlaps with a selective sweep that was previously identified from inshore samples and contains a Tandem repeat of EGFR genes (Cooke et al., 2020).

Inversions captured 214 genes with diverse functions, however, we found that there was a strong enrichment ($p = 8e\text{-}5$; Fisher exact test in topGO) for genes involved in DNA binding (GO:0003677) due to the presence of 14 genes with this GO term across the L1, L2 and L5 inversions. Of these, 4 were involved in transcriptional regulation and 3 were ATP-dependent helicases (1 on L1 and 2 on L2), a protein family that plays crucial roles in recombination and DNA repair.

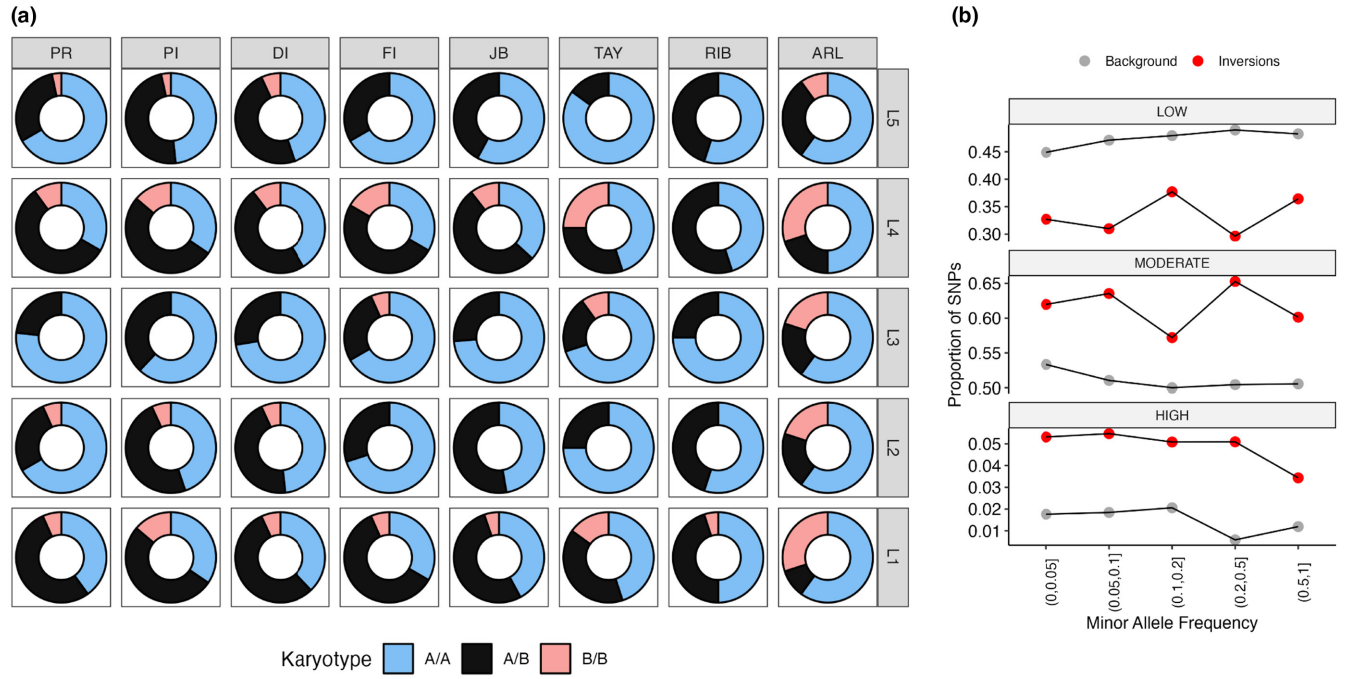

FIGURE 4 (a) Distribution of karyotype frequencies for inversions across all non-Magnetic Island sites. The letter A is used to denote the most frequent arrangement. (b) Mutational load of inversions in *Acropora kenti*. Proportions of SNPs rated as low, moderate and high impact by SnpEff are shown across all allele frequency classes within inversions (red) and in the background collinear genome (grey).

3 | DISCUSSION

Our results clearly show that at least five chromosomal rearrangements, most likely inversions, exist as common polymorphisms within *Acropora kenti* on the central Great Barrier Reef. Elevated mutational load within these inversions and uniformly high frequencies of the minor arrangement among reefs (Figure 4) suggest they are under balancing selection. Inversions such as these (lacking strong spatial or ecological structure) are relatively challenging to detect because they do not produce signatures of elevated divergence (large blocks of high F_{ST}) and can only be identified through direct read-based methods or local PCA (e.g. Figure 3). Moreover, datasets that include overall population structure (in the collinear genome) might fail to detect inversions through local PCA analysis as signals from inversions would be difficult to distinguish from background population structure. Our results also suggest that identifying inversions through local PCA analysis requires a large sample size as we were unable to detect any signatures of inversions at Magnetic Island (28 samples) despite direct read-based evidence from one individual (MI-1-4) that at least the L1 inversion is present in this population. These factors suggest that inversion polymorphisms of the type identified here may be present in many taxa, including corals, but have largely been overlooked until now due to a historical focus on signatures of selection (Cooke et al., 2020; Thomas et al., 2022) and population structure (Bongaerts et al., 2021; Shinzato et al., 2015) in coral genomic studies.

Our observation that mutational load (Figure 4b) was higher within inversions compared with the background genome is consistent with similar findings in butterflies (Jay et al., 2021) and sunflowers (Huang et al., 2022), as well as theoretical work (Berdan et al., 2021), showing that reduced effective population size within inversions contributes to higher rates of genetic drift. An important consequence of accumulated mutational load within inversions is that if harmful mutations are completely or partially recessive, then the heterokaryotype will experience elevated fitness relative to the homokaryotypes. This can potentially lead to a situation where a balanced polymorphism is maintained by associative overdominance (Berdan et al., 2021, 2023; Pamilo & Pálsson, 1998). The clearest empirical examples of inversions that persist due to associative overdominance come from systems in which deleterious effects are so strong that one or both homokaryotypes have lower viability, leading to detectable deviations from Hardy–Weinberg equilibrium (HWE) (Jay et al., 2021; Lindtke et al., 2017). Since none of the five inversions identified in this paper deviated significantly from HWE (Figure 4a), they are unlikely to be associated with recessive lethal effects. However, given the large number of high-impact SNPs found within inversions by SnpEff, some moderate-fitness impacts are expected, and if these SNPs are at least partially recessive, they will contribute to heterokaryotype advantage. Note that a moderate-fitness differential between hetero- and homokaryotypes may not have generated detectable deviations from HWE as detecting these requires much larger sample sizes than used in our study (Lachance, 2009).

Even if the elevated mutational load observed within inversions in *A. kenti* contributes to associative overdominance, it is unlikely that this is the only factor contributing to establishment and persistence of these polymorphisms. Recent theoretical work has shown that in species with large population sizes like *A. kenti* ($Ne \sim 1 \times 10^5$; Cooke et al., 2020), associative overdominance alone is unlikely to allow new inversions to become established but could

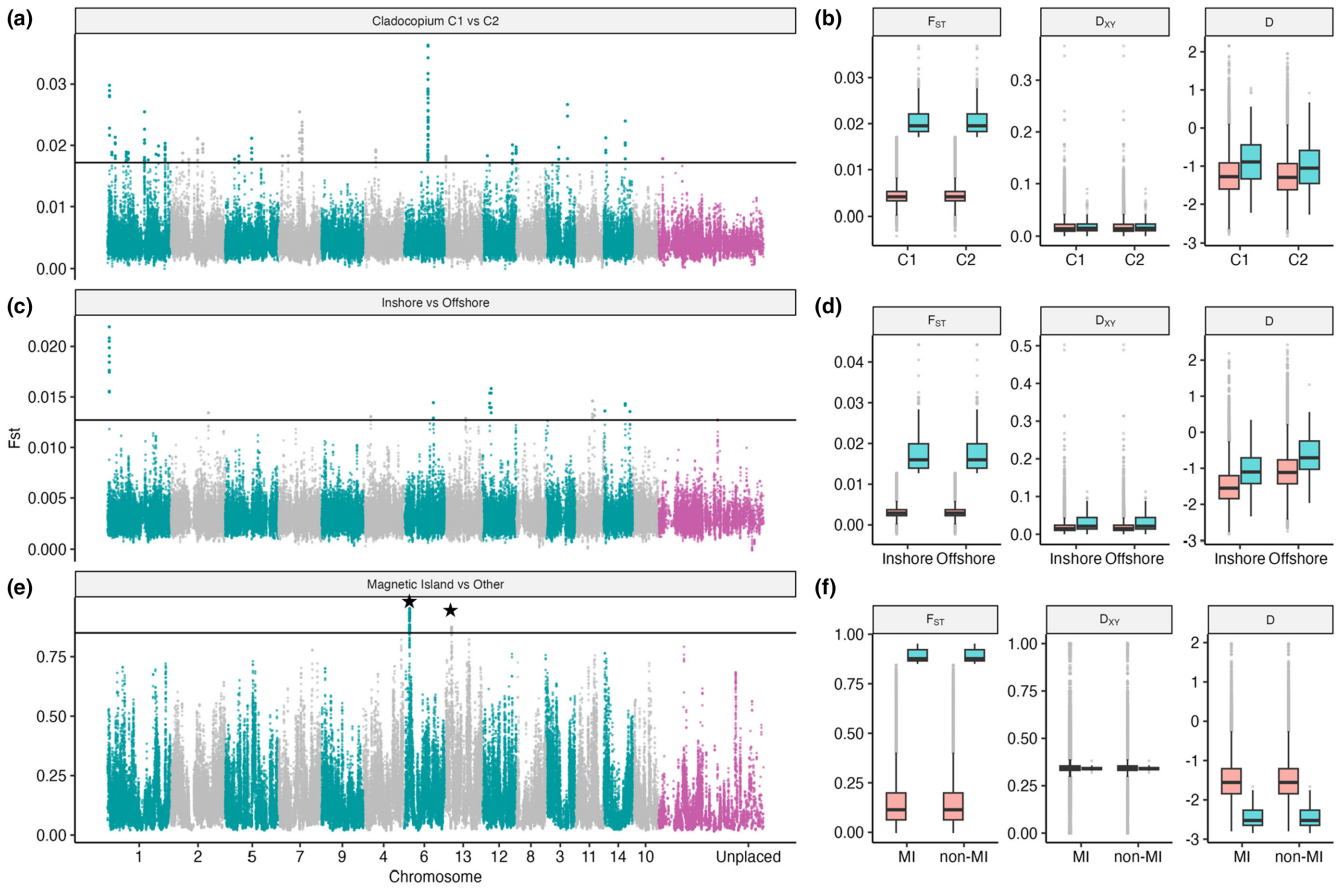

FIGURE 5 Genome-wide scan for highly differentiated regions between colonies dominated by C1 or C2 symbionts (a, b), inshore and offshore reefs (c, d) and Magnetic Island and non-Magnetic Island (e, f). Manhattan plots (a, c, e) show F_{ST} calculated within 20 kb windows with the horizontal line delineating extreme (*z*-score > 6) values. Boxplots (b, d, f) to the right of each Manhattan plot show corresponding diversity statistics, F_{ST}, D_{xy} and Tajima's *D* calculated in extreme F_{ST} regions (blue) and contrasted with the genomic background (pink). Regions identified as putative selective sweeps are indicated with stars in e.

contribute to their persistence (Berdan et al., 2022) for long time periods. While large N_e may contribute to persistence of inversions via associative overdominance, it opposes their initial establishment (Berdan et al., 2022), which suggests that some other factor is likely to have conferred a selective advantage on the inversion early in its evolution. Such a selective advantage can arise through a wide range of possible mechanisms (reviewed in Berdan et al., 2023), including those related to capture of multiple beneficial alleles on the inverted arrangement or direct fitness effects due to breakpoint or orientation effects. Regardless of the mechanism for establishment, newly inverted arrangements accumulate deleterious variants rapidly due to drift and this can eventually result in overdominance. Once established, an inversion polymorphism could then persist through overdominance opposing fixation of the fitter arrangement.

One possibility that we have not yet explicitly considered is that the inversions in *A. kenti* are not maintained by overdominance at all but represent the products of divergent selection along environmental gradients that went unmeasured in our study. Under this hypothetical scenario, inversions become established and are maintained through adaptive divergence between arrangements favouring homokaryotypes in different environments. If this occurred at a fine spatial scale within reefs and if heterokaryotypes remained viable, we might expect to see similar patterns of arrangement frequency as observed in Figure 3a. Similarly, the abundance of heterokaryotypes would contribute to mutational load within inversions consistent with Figure 3b. Despite this, we consider the divergent selection hypothesis to be unlikely for three reasons. Firstly, because recent simulation results have shown that although gene flow can promote the establishment of inversions, this becomes less likely when gene flow is extremely high (Schaal et al., 2022) as it would be for *A. kenti* at the within-reef scale. Secondly, sampling at all reefs in our study would need to have collected roughly equal numbers of both putative ecotypes despite having no knowledge of their existence. Strong sampling bias would likely result in detectable deviations from HWE. Finally, a high-fitness burden due to the large number of heterokaryotypes would strongly favour the emergence of assortative mating, for example, through offset spawning times or fertilization barriers. There is so far no evidence that any such mechanism exists in *A. kenti*.

Elevated mutational load at inversions in *A. kenti* may have important conservation genetic implications. In particular, the presence of recessive harmful loci in linkage increases the chances that population bottlenecks could result in inbreeding depression. This could be mitigated in conservation management programmes such

as coral aquaculture or assisted gene flow by screening colonies for spawning to ensure a diversity of inversion karyotypes. The potential role of inversions in local or clinal adaptation should also be considered. Although our study was unable to identify any ecological or spatial factors associated with adaptive selection at inversions, future studies at different spatial scales or that measure different ecological variables should keep this possibility in mind. This is important because inversions have often been shown to underpin local adaptation, and genetic interventions such as assisted gene flow should seek to preserve locally adaptive variation.

Structural variants including inversions are often large and have a major impact on evolutionary processes (Mérot et al., 2020), yet our ability to detect and study them remains limited. Our results highlight the fact that structural variants such as inversions are present in corals and can have important impacts on genetic diversity and fitness. As technologies for structural variant detection improve, it is therefore important that SVs are considered alongside SNPs as a key component of genetic variation.

4 | METHODS

4.1 | Sample collection and sequencing

Offshore samples were collected in March 2017 from four locations in the central GBR under Great Barrier Reef Marine Park Authority collection permit G16/38488.1. A. Reefs for offshore samples were selected to approximately match the latitudes of four inshore locations for which sequencing data were available from a previous study (Cooke et al., 2020). A fifth inshore location (Magnetic Island) was also included, but since this is known to harbour a genetically diverged population, no matching offshore reef was considered necessary. This resulted in a total of 228 samples, including 80 from offshore locations (Arlington Reef (ARL) $n=20$, Taylor Reef (TAY) $n=20$, Rib Reef (RIB) $n=20$ and John Brewer Reef (JB) $n=20$), and 148 from inshore reefs, including Magnetic Island (Fitzroy Island (FI) $n=30$, Dunk Island (DI) $n=30$, Palm Islands (PI) $n=30$, Pandora reef (PR) $n=30$, Magnetic Island (MI) $n=28$).

All samples were sequenced using the same sequencing protocol (100 bp paired end) and platform (Illumina HiSeq). Sequencing depth was generally shallow (2–5×) for most samples but two (FI-1-3, MI-1-4) had much deeper coverage (>20×). Sequencing coverage for offshore samples was slightly higher on average (4–5×) than inshore (2–3×). Mapping and coverage statistics for all samples are summarized in Figure S1 and Table S1.

4.2 | Data pre-processing and mapping

We followed the GATk germline variant calling best practices workflow to generate mapped bam files from raw reads for each sample. Reads passing quality checks from each sample and lane were converted to unmapped bam format (uBAM) files. Adapters were marked using MarkIlluminaAdapters (Picard) before mapping to the reference genome assembly using bwa (v0.7.17-r1188). After mapping, PCR and optical duplicates were marked using MarkDuplicates (Picard). Two samples (FI-1-3, MI-1-4) sequenced at much higher coverage (28×, 26×) were down-sampled using sambamba (v0.8.2) to achieve a target depth of approximately 3×.

4.3 | Removal of clones and misidentified samples

To ensure no misidentified samples were present in our data, we first reconstructed a mitochondrial genome for each sample by aligning raw reads to the reference mitogenome sequence of *Acropora kenti* (GenBank accession AF338425) and then extracting the most common base at each position using the -doFasta 2 option in ANGSD. These sequences of our samples were then used as queries to search the NCBI non-redundant nucleotide sequence database (nt June 23 2022) using megablast (v0.8.2) (Morgulis et al., 2008) with the option to output a maximum of five best matches with taxonomy information. While the best match for most samples was the mitogenome of *A. kenti*, nine samples from Arlington Reef and one sample from John Brewer Reef matched *Acropora echinata* or *Acropora florida* (Table S2). Next, we inferred a phylogenetic tree using IQ-TREE (v1.6.4) (Nguyen et al., 2015) based on the alignments (mafft v7.394) (Katoh & Standley, 2013) of mitogenome sequences of all samples together with the reference mitochondrial sequences of *A. kenti* (AF338425) and *A. echinata* (LC201841.1). The resulting tree (Figure S2) revealed that nine samples from Arlington Reef and one from John Brewer Reef formed a distinct monophyletic clade together with the *A. echinata* mitogenome. Since these same samples also had particularly low mapping rates and genome coverage (Figure S1), it is highly likely that they were misidentified in the field. We, therefore, excluded these samples from all further analyses.

Clones and closely related samples were identified by first estimating pairwise relatedness statistics with ngsRelate v2 (Hanghøj et al., 2019) (https://github.com/ANGSD/NgsRelate). Pairs of samples with outlying relatedness were then identified using R1 versus R0, and R1 versus KING-robust kinship plots (Waples et al., 2019). This revealed eight pairs of closely related (expected kinship of 0.125) samples all of which were from Magnetic Island (Figure S3). Retaining samples with higher depth where possible, we then removed six samples to ensure that no close relatives were present in further analyses.

After removal of misidentified samples and close relatives, our final sample set contained 212 in total, including 22 from MI, 30 each from inshore reefs PR, PI, FI and DI, 20 each from offshore reefs RIB and TAY, 19 from JB and 11 from ARL.

4.4 | Variant calling and genotype likelihood calculations

To account for the uncertainty of genotypes of each site due to low (2–5×) per-sample sequencing coverage, we used ANGSD to

estimate genotype likelihoods. These genotype likelihoods formed the basis of all population genomic analyses and unless otherwise specified were generated as follows. ANGSD was run using the genotype likelihood (GL) model from GATK (−gl 2), inferring major and minor alleles from GL data (−doMajorMinor 1) and estimating allele frequencies from GL data (−doMaf 1). Read data were filtered to include only bases with a quality score of at least 30 (−minQ 30) and reads with a mapping quality score of at least 30 (−minMapQ). SNPs were filtered to remove rare alleles (MAF >0.05), keeping only sites with p value $< 10^{-6}$ (-SNP_pval 1e-6) and only sites with data for at least 100 individuals.

Analysis with ANGSD was restricted to a 258 Mb subset of the genome to avoid duplicated, low complexity and poorly assembled regions as follows. First, GENMAP (v1.2.0) (Pockrandt et al., 2020) was used to estimate the mappability of each site. Mappability scores were computed with *k*-mers with no more than two mismatches (-K 50 -E 2), and sites were retained if they had a mappability score equal to 1, which suggests they can be uniquely mapped. Second, we used mdust (v2006.10.17; default parameters) to identify and exclude low-complexity regions in the genome. Third, we excluded any sites from short (<1 Mb) scaffolds to reduce the influence of artefacts at the ends of fragmented reference sequences. Finally, since regions with very high or very low mapping depth are often associated with ambiguous mapping due to repeats, we removed sites with global depth across all samples higher or lower than the range's 1% percentile (minimum 17×, maximum 1102×). The proportion of genome sequence remaining after each of these steps is provided in Table S3.

4.5 | Population structure and admixture

We used PCAngsd (v1.10) (Meisner et al., 2021) to explore population structure and calculate admixture coefficients for all individuals. As input to PCAngsd, we use genotype likelihoods calculated with ANGSD (see above) across the entire dataset (212 samples). The output covariance matrix estimated based on individual allele frequency was then used to compute eigenvalues and eigenvectors using the R package eigen and to generate PCA plots. PCAngsd was also used to automatically infer the best number of clusters (K = 2, −admix_auto 10,000) and perform admixture analysis. This revealed four highly admixed individuals (MI-1-1_S10, PI-1-16_S16, DI-2-4_S17 and ARL_15_S69) that were excluded from further analyses. We also used NGAdmix to explore admixture with alternative numbers of clusters (K = 2, K = 3). The results for K = 2 were qualitatively similar to PCAngsd and results for K = 3 showed little support for a third cluster. Results for NGSAdmix with K = 2 are shown in Figure 1.

To confirm that structure within the non-Magnetic Island population was not obscured by the presence of Magnetic Island samples, we reran ANGSD and PCAngsd excluding all Magnetic Island samples and admixed samples. A PCA plot based on this analysis (Figure S4) showed no visible structure between inshore and offshore.

To complement PCAngsd analyses, we built sample trees based on pairwise genetic distances measured using the identity-by-state (-doIBS 1) calculation in ANGSD which randomly samples a single read from each position from each sample within filtered reference sites. The R function hclust was then used to generate a UPGMA tree from the IBS distance matrix and this was visualized with ggtree (Yu et al., 2017) (Figure S5).

4.6 | Genome-wide estimates of genetic diversity within and between reefs

To calculate reef-specific diversity and divergence statistics, we first used the realSFS program within ANGSD to estimate one-dimensional (1D) folded site frequency spectra (SFS) for each of the nine reef locations separately, and two-dimensional folded SFS (2D) for each reef pair. Before estimating the SFS with realSFS, a two-step procedure was first implemented to generate a saf (site allele frequency likelihood) file followed by an optimization of the saf file using ANGSD (−dosaf 1). Pairwise nucleotide diversity (π), Watterson's θ and Tajima's D were estimated from the 1D-SFS of each reef using the thetaStat function within ANGSD with a sliding window size of 10 kb and s step size of 4 kb. Global estimates of F_{ST} for each pair of reefs were computed directly from the 2D-SFS using the Reich estimator implemented in realSFS. A bootstrapped UPGMA tree based on pairwise F_{ST} was also generated using the R package ape (Figure S6).

4.7 | Individual heterozygosity

The heterozygosity for each sample was estimated in ANGSD as the proportion of heterozygous sites in the 1D-SFS of each individual. A saf file was generated for each sample using ANGSD and used to estimate the 1D-SFS with the realSFS. The heterozygosity rate is calculated by dividing the number of variant sites by the total number of sites in R. This calculation was performed for each sample based on all available reads and then again for each sample after down-sampling to 2× coverage to determine whether differences in coverage could explain consistent differences between inshore and offshore samples (Figure S7). The same procedure for calculating individual heterozygosity was also used within inversions by restricting the analysis to reads overlapping each inversion.

4.8 | Analysis of symbiont reads

To identify the dominant genus of Symbiodiniaceae within each sample, we used the moqc pipeline (https://github.com/marine-omics/moqc), which performs a taxonomic classification of raw reads with the program KrakenUniq (Breitwieser et al., 2018). The database for

KrakenUniq included a representative full genome for each of the five major coral-associating genera as well as sequences from common contaminants and the genome of the coral host. Full details of all sequences used for database construction are provided as part of the moqc documentation.

Since KrakenUniq profiles for almost all samples showed *Cladocopium* as the dominant symbiont, we mapped deduplicated reads to the *Cladocopium* mitochondrial genome (downloaded from http://symbs.reefgenomics.org/). We then used the doHaploCall function within ANGSD to obtain the consensus base at each position and print only positions where there is at least one variant allele (minMinor 1) and exclude positions where more than 10 individuals have an ambiguous base (maxMis 10). This resulted in alignment with 145 variable sites that we then cleaned further with goalign (Lemoine & Gascuel, 2021) to remove sequences from 29 samples that had more than 4% ambiguous bases. This alignment was used to generate a haplotype network with PopArt (Figure 2) and was also used to generate a maximum-likelihood tree with IQ-Tree (Nguyen et al., 2015). IQ-Tree was run using ModelFinder to automatically detect the best evolutionary model with 1000 ultrafast bootstraps. Visual inspection of this tree was used to identify individual samples belonging to the two most common haplogroups (Figure S9) referred to in the text as C1 and C2.

To verify that distinctions between C1 and C2 symbiont harbouring colonies are not due to idiosyncrasies of mitochondrial genomes or difficulties calling consensus sequences from low coverage, we used d2ssect (https://github.com/bakeronit/d2ssect) to calculate pairwise distances between samples based on shared kmers within reads of symbiont origin using D2S statistic (Song et al., 2014). All deduplicated reads that mapped to the Cladocopium genome were used for d2ssect analysis which generates a distance matrix based on d2s statistics. The cmdscale function in R was then used to reduce this distance matrix to two dimensions for plotting. The resulting plot, coloured by symbiont mito-haplogroup, is shown in Figure S10.

4.9 | Identifying and genotyping inversion loci

Inversion loci were initially identified using PCAngsd (Meisner et al., 2021) as peaks in local genetic structure (Figure 3). PCAngsd was run on all non-hybrid non-Magnetic Island individuals ($n = 187$) with the '-selection' option which calculates variant weights based on the first principal component (Galinsky statistic). Results were converted into pseudo-chromosome coordinates using RagTag (Alonge et al., 2019) (see below) and Galinsky statistics were converted to p-values using the pchisq function in R. Results for individual variants were then smoothed by calculating the average of −log10(p) within 100 kb sliding windows with a 10 kb step using bedtools (Quinlan & Hall, 2010). Inversion boundaries were then calculated by finding the start and end points of regions where −log10(p) exceeded 3 ($p < .001$).

An attempt was made to replicate the inversion finding process described above for the 21 Magnetic Island individuals (excluding hybrids, clones and close-kin) but this did not yield any peaks exceeding our p-value threshold. To demonstrate that this lack of signal was most likely due to low sample size, we performed the same analysis on a random subset of 21 individuals from the non-Magnetic Island population. The results of this analysis are shown in Figure S11.

To infer inversion genotypes, we first extracted genotype likelihood data specific to each inversion from the overall (genome-wide) genotype likelihoods calculated with ANGSD. A separate PCAngsd analysis was then run on each of these genotype likelihood files and genotypes were inferred by partitioning samples into three clusters along the first principal component (Figure S12). The central cluster was always assumed to comprise heterozygotes (A/B) and clusters at the extremes were arbitrarily assigned to one of the two potential homozygous genotypes (A/A or B/B). Clusters were inferred using k-means clustering in R with k = 3. Visual inspection of PC1 values and genotype assignments clearly indicates that not all samples could be unambiguously assigned to a single cluster, and it is therefore likely that our genotype assignments are not error free. Nevertheless, we found that genotype proportions for all inversions did not deviate significantly from those expected under Hardy–Weinberg equilibrium (Table S4).

4.10 | Heterozygosity, linkage disequilibrium and pairwise sample distances within inversions

To calculate individual heterozygosity within each inversion, we used ANGSD and realSFS to generate a folded allele frequency spectrum for each sample and then calculated heterozygosity as the proportion of heterozygotes to invariant sites.

Linkage disequilibrium was calculated for genomic regions including each inversion plus up to 1 Mb of flanking sequence at each end. Linkage disequilibrium statistics were then calculated using ngsLD (Fox et al., 2019) on genotype likelihood data for SNPs within each of these regions with individuals grouped into genotypes inferred via PCA (see section above). To reduce computation time and output file size, a random sample of 1% of SNPs was used for inversions L1–L4, and a sample of 50% of SNPs was used for L5. Only SNPs with data for at least 25 individuals were included. Outputs from ngsLD were then further processed in R to calculate average values of the EM r^2 statistic for all SNPs in a regular 30 × 30 grid over each interval.

To investigate the age of inversions relative to the split between Magnetic Island and non-Magnetic Island populations, we calculated genetic distances between pairs of individuals at inversion loci using ngsDist (Vieira et al., 2016) https://github.com/fgvieira/ngsDist. Starting with ANGSD genotype likelihoods calculated on all 212 correctly identified unrelated individuals, we extracted data for SNPs contained within each inversion and also for a set of 100, 50 kb regions randomly sampled from the genome. We then used ngsDist to calculate a pairwise genetic distance matrix for each of these data subsets, using the '--pairwise_del' option to ensure that only sites with data for both individuals were used. We then used the 'hclust' package in R to generate a UPGMA tree based on pairwise distance matrices for each of these datasets. These trees are shown in Figure S14.

4.11 | Variant severity in inversions

Since ANGSD variant calling focuses entirely on SNPs, we used the bcftools mpileup variant caller to call SNPs and small indels for the purposes of variant severity analysis. As input to bcftools, we used the same read alignment files that were used for ANGSD and we used the -q 20 and -Q 20 flags to ignore low-quality (Phred < 20) basecalls and alignments (MAPQ < 20). We then filtered variant calls to retain only biallelic variants with at least a minor allele count of 2 and with quality scores greater than 30. Variants were also removed if they had more than 50% missing genotypes or a highly skewed allele balance (FS < 20). The resulting vcf file was then split into a component that overlapped with inversions and a background component that overlapped with 100 randomly selected 50 kb regions. We then ran snpEff (Cingolani et al., 2012) (version 5.2a) on each of these files to predict variant impacts, restricting the analysis to genic regions to avoid biases due to possible differences in gene density between regions. Variants were then tabulated according to the top 3 severity levels (low, medium and high) and allele frequency. In general, we found that variant counts decreased as a function of allele frequency, however, a relatively large number of variants with frequencies close to 1 (>0.99) were present. We removed these variants as they may represent assembly or sequencing errors in the reference.

4.12 | Significance of ecological variables in structuring genetic variance in inversions

We used analysis of molecular variance (AMOVA) (Excoffier et al., 1992) to test the significance of various factors in structuring genetic variance at inversion loci. For this analysis, we used inversion genotypes inferred using k-means clustering along PC1 (see above) and did not consider SNPs within inversion haplotypes. Analyses for each inversion were conducted separately with two models ~shore/reef and ~symbiont/reef, where shore codes for inshore and offshore reef locations. Only colonies that could be clearly assigned to either C1- or C2- dominant symbionts were used for the symbiont analysis. AMOVAs were calculated using the amova function in the poppr package and statistical significance of variance components was tested using the randtest.amova function with 999 random permutations. A summary of *p*-values and phi statistics for all tests is provided in Table S5.

4.13 | Calculating sliding-window population genetic statistics

We used ANGSD to estimate genome-wide patterns of pairwise F_{ST}, genetic diversity and Tajima's *D*. F_{ST} and D_{xy} were calculated for three pairwise contrasts between samples: (1) inshore and offshore reefs excluding Magnetic Island, (2) non-Magnetic Island and Magnetic Island and (3) samples that could be unambiguously assigned as harbouring C1 versus C2 symbionts.

For each sample grouping, realSFS was used to estimate the 1D SFS and then the 2D SFS for each pair of groups. Pairwise F_{ST} was then calculated in sliding windows using (realSFS fst stats2 -type 1) with a window size of 20 kb and step of 4 kb. For each sample, grouping theta statistics were estimated using thetaStat within ANGSD (thetaStat do_stat -type 1) using the same sliding windows for F_{ST} scans. To avoid false signals resulting from window-based statistics dominated by very little data in the window, we excluded windows where the number of available sites (passing quality checks) was less than 10% of all sites.

We used a Perl script getDxy.pl (https://github.com/mfumagalli/ngsPopGen/blob/9ee3a6d5e733c1e248e81bfc21514b0527da967b/scripts/getDxy.pl) provided by the ngsPopGen toolset to calculate the D_{xy} for every site in the mafs files generated by ANGSD; non-bi-allelic sites were removed in the calculation. Per-site D_{xy} values were then grouped into sliding windows from F_{ST} estimates and the average value was assigned as the value for each window using Bedtools intersect and groupby.

4.14 | GO term enrichment of genes captured by inversions

To investigate genes captured by inversions, we used bedtools to find all 214 genes that overlapped with inversion coordinates. Functional annotations for these genes and all genes in the *A. kenti* genome were obtained from supplementary material in (Cooke et al., 2020). Functional enrichment of genes within inversions was tested using topGO with the weight01 algorithm which reduces the likelihood of false-positive enrichment in higher-level terms. Statistical significance of enrichment was assessed using Fisher's exact test and terms with $p < .01$ are listed in Table S6. For the most significantly enriched term (GO:0003677), supporting genes are listed in Table S7.

4.15 | Building a pseudo-chromosome reference

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

4.16 | Read-based evidence for inversions

We ran manta (Chen et al., 2016) with default parameters to call structural variants from one deeply sequenced individual at Fitzroy Island and one at Magnetic Island. Inversions were extracted and overlapped

using a consensus approach previously described (pubmed: 31844586) and visualized using IGV (pubmed: 22517427) for manual validation.

AUTHOR CONTRIBUTIONS

JZ performed the majority of analyses with assistance from IC and advice from CXC, DM and JS. NS and MF performed analyses focussed on confirming the presence of inversions using read-based evidence. NS performed statistical analysis of inversion karyotypes. CXC and JZ performed analysis of symbiont data. LB, CR and IC designed the study, adding offshore samples to previous work that focussed purely on inshore locations. JZ and IC wrote the manuscript with contributions from all authors.

ACKNOWLEDGEMENT

Open access publishing facilitated by James Cook University, as part of the Wiley - James Cook University agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

This work was supported by a grant from the Great Barrier Reef Foundation to cover sequencing costs associated with offshore samples. The authors would like to acknowledge support from the Australian Institute of Marine Science for collection and processing of samples.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Detailed methods including code and additional data are available on GitHub https://github.com/bakeronit/acropora_kenti_wgs. Raw sequencing data are available under the NCBI Bioprojects PRJEB37470 and PRJEB73886. The Acropora kenti reference genome and gene models used for all analyses are available at https://aten.reefgenomics.org/.

ORCID

Cheong Xin Chan https://orcid.org/0000-0002-3729-8176
David J. Miller https://orcid.org/0000-0003-0291-9531
Jan M. Strugnell https://orcid.org/0000-0003-2994-637X
Cynthia Riginos https://orcid.org/0000-0002-5485-4197
Ira Cooke https://orcid.org/0000-0001-6520-1397

REFERENCES

Abrego, D., Van Oppen, M. J. H., & Willis, B. L. (2009). Onset of algal endosymbiont specificity varies among closely related species of Acropora corals during early ontogeny. *Molecular Ecology*, *18*, 3532–3543.

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology*, *20*, 224.

Anderson, A. R., Hoffmann, A. A., McKechnie, S. W., Umina, P. A., & Weeks, A. R. (2005). The latitudinal cline in the In(3R)Payne inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Molecular Ecology*, *14*, 851–858.

Anthony, K., Bay, L. K., Costanza, R., Firn, J., Gunn, J., Harrison, P., Heyward, A., Lundgren, P., Mead, D., Moore, T., Mumby, P. J., van Oppen, M. J. H., Robertson, J., Runge, M. C., Suggett, D. J., Schaffelke, B., Wachenfeld, D., & Walshe, T. (2017). New interventions are needed to save coral reefs. *Nature Ecology & Evolution*, *1*, 1420–1422.

Ayala, D., Guerrero, R. F., & Kirkpatrick, M. (2013). Reproductive isolation and local adaptation quantified for a chromosome inversion in a malaria mosquito. *Evolution*, *67*, 946–958.

Baums, I. B., Baker, A. C., Davies, S. W., Grottoli, A. G., Kenkel, C. D., Kitchen, S. A., Kuffner, I. B., LaJeunesse, T. C., Matz, M. V., Miller, M. W., Parkinson, J. E., & Shantz, A. A. (2019). Considerations for maximizing the adaptive potential of restored coral populations in the western Atlantic. *Ecological Applications*, *29*, e01978.

Berdan, E. L., Barton, N. H., Butlin, R., Charlesworth, B., Faria, R., Fragata, I., Gilbert, K. J., Jay, P., Kapun, M., Lotterhos, K. E., Mérot, C., Durmaz Mitchell, E., Pascual, M., Peichel, C. L., Rafajlović, M., Westram, A. M., Schaeffer, S. W., Johannesson, K., & Flatt, T. (2023). How chromosomal inversions reorient the evolutionary process. *Journal of Evolutionary Biology*, *36*, 1761–1782. https://doi.org/10.1111/jeb.14242

Berdan, E. L., Blanckaert, A., Butlin, R. K., & Bank, C. (2021). Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genetics*, *17*, e1009411.

Berdan, E. L., Blanckaert, A., Butlin, R. K., Flatt, T., Slotte, T., & Wielstra, B. (2022). Mutation accumulation opposes polymorphism: Supergenes and the curious case of balanced lethals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*, 20210199.

Berry, K. L. E., Hoogenboom, M. O., Flores, F., & Negri, A. P. (2016). Simulated coal spill causes mortality and growth inhibition in tropical marine organisms. *Scientific Reports*, *6*, 25894.

Bongaerts, P., Cooke, I. R., Ying, H., Wels, D., den Haan, S., Hernandez-Agreda, A., Brunner, C. A., Dove, S., Englebert, N., Eyal, G., Forêt, S., Grinblat, M., Hay, K. B., Harii, S., Hayward, D. C., Lin, Y., Mihaljević, M., Moya, A., Muir, P., … Hoegh-Guldberg, O. (2021). Morphological stasis masks ecologically divergent coral species on tropical reefs. *Current Biology*, *31*, 2286–2298.e8.

Breitwieser, F. P., Baker, D. N., & Salzberg, S. L. (2018). KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. *Genome Biology*, *19*, 198.

Bridge, T. C. L., Cowman, P. F., Quattrini, A. M., Bonito, V. E., Sinniger, F., Harii, S., Head, C. E. I., Hung, J. Y., Halafihi, T., Rongo, T., & Baird, A. H. (2023). A tenuis relationship: Traditional taxonomy obscures systematics and biogeography of the "*Acropora tenuis*" (Scleractinia: Acroporidae) species complex. *Zoological Journal of the Linnean Society*, zlad062. https://doi.org/10.1093/zoolinnean/zlad062

Brodie, J. E., Devlin, M., Haynes, D., & Waterhouse, J. (2011). Assessment of the eutrophication status of the great barrier reef lagoon (Australia). *Biogeochemistry*, *106*, 281–302.

Butler, C. C., Turnham, K. E., Lewis, A. M., Nitschke, M. R., Warner, M. E., Kemp, D. W., Hoegh-Guldberg, O., Fitt, W. K., van Oppen, M. J. H., & LaJeunesse, T. C. (2023). Formal recognition of host-generalist species of dinoflagellate (Cladocopium, Symbiodiniaceae) mutualistic with Indo-Pacific reef corals. *Journal of Phycology*, *59*, 698–711.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S., & Saunders, C. T. (2016). Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, *32*, 1220–1222.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff:

SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*, 80–92.

Coles, S. L., & Jokiel, P. L. (1992). Effects of salinity on coral reefs. In D. W. Connell & D. W. Hawker (Eds.), *Pollution in tropical aquatic systems* (pp. 147–166). CRC Press.

Colton, M. A., McManus, L. C., Schindler, D. E., Mumby, P. J., Palumbi, S. R., Webster, M. M., Essington, T. E., Fox, H. E., Forrest, D. L., Schill, S. R., Pollock, F. J., DeFilippo, L. B., Tekwa, E. W., Walsworth, T. E., & Pinsky, M. L. (2022). Coral conservation in a warming world must harness evolutionary adaptation. *Nature Ecology & Evolution*, *6*, 1405–1407.

Cooke, I., Ying, H., Forêt, S., Bongaerts, P., Strugnell, J. M., Simakov, O., Zhang, J., Field, M. A., Rodriguez-Lanetty, M., Bell, S. C., Bourne, D. G., van Oppen, M. J., Ragan, M. A., & Miller, D. J. (2020). Genomic signatures in the coral holobiont reveal host adaptations driven by Holocene climate change and reef specific symbionts. *Science Advances*, *6*, eabc6318.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157.

DeFilippo, L. B., McManus, L. C., Schindler, D. E., Pinsky, M. L., Colton, M. A., Fox, H. E., Tekwa, E. W., Palumbi, S. R., Essington, T. E., & Webster, M. M. (2022). Assessing the potential for demographic restoration and assisted evolution to build climate resilience in coral reefs. *Ecological Applications*, *32*, e2650.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*, 479–491.

Fabian, D. K., Kapun, M., Nolte, V., Kofler, R., Schmidt, P. S., Schlötterer, C., & Flatt, T. (2012). Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology*, *21*, 4748–4769.

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in Ecology & Evolution*, *34*, 239–248.

Fox, E. A., Wright, A. E., Fumagalli, M., & Vieira, F. G. (2019). ngsLD: Evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*, *35*, 3855–3856.

Fuentes-Pardo, A. P., Farrell, E. D., Pettersson, M. E., Sprehn, C. G., & Andersson, L. (2023). The genomic basis and environmental correlates of local adaptation in the Atlantic horse mackerel (*Trachurus trachurus*). *Evolutionary Applications*, *16*, 1201–1219.

Fuller, Z. L., Mocellin, V. J. L., Morris, L. A., Cantin, N., Shepherd, J., Sarre, L., Peng, J., Liao, Y., Pickrell, J., Andolfatto, P., Matz, M., Bay, L. K., & Przeworski, M. (2020). Population genetics of the coral Acropora millepora: Toward genomic prediction of bleaching. *Science*, *369*, eaba4674.

Furnas, M. J. (2003). *Catchments and Corals*: *Terrestrial Runoff to the Great Barrier Reef* (Australian Institute of Marine Science).

Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A., & Korneliussen, T. S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience*, *8*(5). https://doi.org/10.1093/gigascience/giz034

Harringmeyer, O. S., & Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution*, *6*, 1965–1979. https://doi.org/10.1038/s41559-022-01890-0

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, *39*, 21–42.

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, *29*, 2535–2549. https://doi.org/10.1101/829622

Huang, K., Ostevik, K. L., Elphinstone, C., Todesco, M., Bercovich, N., Owens, G. L., & Rieseberg, L. H. (2022). Mutation load in sunflower inversions is negatively correlated with inversion heterozygosity. *Molecular Biology and Evolution*, *39*, msac101.

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics*, *53*, 288–293.

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Whibley, A., Becuwe, M., Baxter, S. W., Ferguson, L., Wilkinson, P. A., Salazar, C., Davidson, C., Clark, R., Quail, M. A., Beasley, H., Glithero, R., Lloyd, C., Sims, S., … Ffrench-Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, *477*, 203–206.

Kapun, M., Mitchell, E. D., Kawecki, T. J., Schmidt, P., & Flatt, T. (2023). An ancestral balanced inversion polymorphism confers global adaptation. *Molecular Biology and Evolution*, *40*, msad118.

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, *8*, e1000501.

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*, 419–434.

Lachance, J. (2009). Detecting selection-induced departures from Hardy-Weinberg proportions. *Genetics*, *Selection*, *Evolution*, *41*, 15.

Ladner, J. T., & Palumbi, S. R. (2012). Extensive sympatry, cryptic diversity and introgression throughout the geographic distribution of two coral species complexes. *Molecular Ecology*, *21*, 2224–2238.

Lemoine, F., & Gascuel, O. (2021). Gotree/Goalign: Toolkit and go API to facilitate the development of phylogenetic workflows. *NAR Genomics and Bioinformatics*, *3*, lqab075.

Lindtke, D., Lucek, K., Soria-Carrasco, V., Villoutreix, R., Farkas, T. E., Riesch, R., Dennis, S. R., Gompert, Z., & Nosil, P. (2017). Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology*, *26*, 6189–6205.

Locatelli, N. S., Kitchen, S. A., Stankiewicz, K. H., Cornelia Osborne, C., Dellaert, Z., Elder, H., Kamel, B., Koch, H. R., Fogarty, N. D., & Baums, I. B. (2023). 1 Genome assemblies and genetic maps highlight chromosome-scale macrosynteny in Atlantic 2 acroporids. https://doi.org/10.1101/2023.12.22.573044

Logan, C. A., Dunne, J. P., Eakin, C. M., & Donner, S. D. (2014). Incorporating adaptive responses into future projections of coral bleaching. *Global Change Biology*, *20*, 125–139.

Matias, A. M. A., Popovic, I., Thia, J. A., Cooke, I. R., Torda, G., Lukoschek, V., Bay, L. K., Kim, S. W., & Riginos, C. (2022). Cryptic diversity and spatial genetic variation in the coral *Acropora tenuis* and its endosymbionts across the great barrier reef. *Evolutionary Applications*, *16*, 293–310. https://doi.org/10.1111/eva.13435

Matz, M. V., Treml, E. A., & Haller, B. C. (2020). Estimating the potential for coral adaptation to global warming across the indo-West Pacific. *Global Change Biology*, *26*, 3473–3481.

McManus, L. C., Forrest, D. L., Tekwa, E. W., Schindler, D. E., Colton, M. A., Webster, M. M., Essington, T. E., Palumbi, S. R., Mumby, P. J., & Pinsky, M. L. (2021). Evolution and connectivity influence the persistence and recovery of coral reefs under climate change in the Caribbean, Southwest Pacific, and coral triangle. *Global Change Biology*, *27*, 4307–4321.

Meisner, J., Albrechtsen, A., & Hanghøj, K. (2021). Detecting selection in low-coverage high-throughput sequencing data using principal component analysis. *BMC Bioinformatics*, *22*, 470.

Mérot, C., Berdan, E. L., Cayuela, H., Djambazian, H., Ferchaud, A.-L., Laporte, M., Normandeau, E., Ragoussis, J., Wellenreuther, M., & Bernatchez, L. (2021). Locally adaptive inversions modulate genetic

variation at different geographic scales in a seaweed fly. *Molecular Biology and Evolution*, *38*, 3953–3971.

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, *35*, 561–572.

Meyer, L., Barry, P., Riquet, F., Foote, A., Sarkissian, C. D., Cunha, R., Arbiol, C., Cerqueira, F., Desmarais, E., Bordes, A., Bierne, N., Guinand, B., & Gagnaire, P.-A. (2023). Divergence and gene flow history at two large chromosomal inversions involved in long-snouted seahorse ecotype formation. *bioRxiv*. https://doi.org/10.1101/2023.07.04.547634

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics*, *24*(16), 1757–1764. https://doi.org/10.1093/bioinformatics/btn322

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274.

Pamilo, P., & Pálsson, S. (1998). Associative overdominance, heterozygosity and fitness. *Heredity*, *81*(Pt 4), 381–389.

Pockrandt, C., Alzamel, M., Iliopoulos, C. S., & Reinert, K. (2020). GenMap: ultra-fast computation of genome mappability. *Bioinformatics*, *36*(12), 3687–3692. https://doi.org/10.1093/bioinformatics/btaa222

Quigley, K. M., Alvarez Roa, C., Torda, G., Bourne, D. G., & Willis, B. L. (2020). Co-dynamics of Symbiodiniaceae and bacterial populations during the first year of symbiosis with *Acropora tenuis* juveniles. *Microbiology*, *9*, e959.

Quigley, K. M., & van Oppen, M. J. H. (2022). Predictive models for the selection of thermally tolerant corals based on offspring survival. *Nature Communications*, *13*, 1543.

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

Rippe, J. P., Dixon, G., Fuller, Z. L., Liao, Y., & Matz, M. (2021). Environmental specialization and cryptic genetic divergence in two massive coral species from the Florida keys reef tract. *Molecular Ecology*, *30*, 3468–3484.

Rose, N. H., Bay, R. A., Morikawa, M. K., Thomas, L., Sheets, E. A., & Palumbi, S. R. (2021). Genomic analysis of distinct bleaching tolerances among cryptic coral species. *Proceedings of the Biological Sciences*, *288*, 20210678.

Schaal, S. M., Haller, B. C., & Lotterhos, K. E. (2022). Inversion invasions: When the genetic basis of local adaptation is concentrated within inversions in the face of gene flow. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*, 20210200.

Schluter, D., & Rieseberg, L. H. (2022). Three problems in the genetics of speciation by selection. *Proceedings of the National Academy of Sciences of the United States of America*, *119*, e2122153119.

Shinzato, C., Mungpakdee, S., Arakaki, N., & Satoh, N. (2015). Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu archipelago. *Scientific Reports*, *5*, 18211.

Song, K., Ren, J., Reinert, G., Deng, M., Waterman, M. S., & Sun, F. (2014). New developments of alignment-free sequence comparison: Measures, statistics and next-generation sequencing. *Briefings in Bioinformatics*, *15*, 343–353.

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

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, *25*, 2144–2164.

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, *584*, 602–607.

Vieira, F. G., Lassalle, F., Korneliussen, T. S., & Fumagalli, M. (2016). Improving the estimation of genetic distances from next-generation sequencing data. *Biological Journal of the Linnean Society*, *117*, 139–149.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, *493*, 664–668.

Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, *28*, 35–48.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, *33*, 427–440.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*, 28–36.

Zhang, J., Richards, Z. T., Adam, A. A. S., Chan, C. X., Shinzato, C., Gilmour, J., Thomas, L., Strugnell, J. M., Miller, D. J., & Cooke, I. (2022). Evolutionary responses of a reef-building coral to climate change at the end of the last glacial maximum. *Molecular Biology and Evolution*, *39*, msac201. https://doi.org/10.1093/molbev/msac201

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhang, J., Schneller, N. M., Field, M. A., Chan, C. X., Miller, D. J., Strugnell, J. M., Riginos, C., Bay, L., & Cooke, I. (2024). Chromosomal inversions harbour excess mutational load in the coral, *Acropora kenti*, on the Great Barrier Reef. *Molecular Ecology*, *33*, e17468. https://doi.org/10.1111/mec.17468