RESEARCH ARTICLE SUMMARY

CORAL GENOMICS

Population genetics of the coral *Acropora millepora*: Toward genomic prediction of bleaching

Zachary L. Fuller*, Veronique J. L. Mocellin, Luke A. Morris, Neal Cantin, Jihanne Shepherd, Luke Sarre, Julie Peng, Yi Liao, Joseph Pickrell, Peter Andolfatto, Mikhail Matz†, Line K. Bay*†, Molly Przeworski*†

INTRODUCTION: Coral reefs worldwide are suffering losses at an alarming rate as a result of anthropogenic climate change. Increased seawater temperatures, even only slightly above long-term maxima, can induce bleaching—the breakdown of the symbiotic relationship between coral hosts and their intracellular photosynthetic dinoflagellates from the family Symbiodiniaceae. Because these symbionts provide the majority of energy required by the coral host, prolonged periods of bleaching can eventually lead to the death of the colony. In the face of rapidly increasing temperatures, new conservation strategies are urgently needed to prevent future mass losses of coral cover, and these benefit from an understanding of the genetic basis of bleaching.

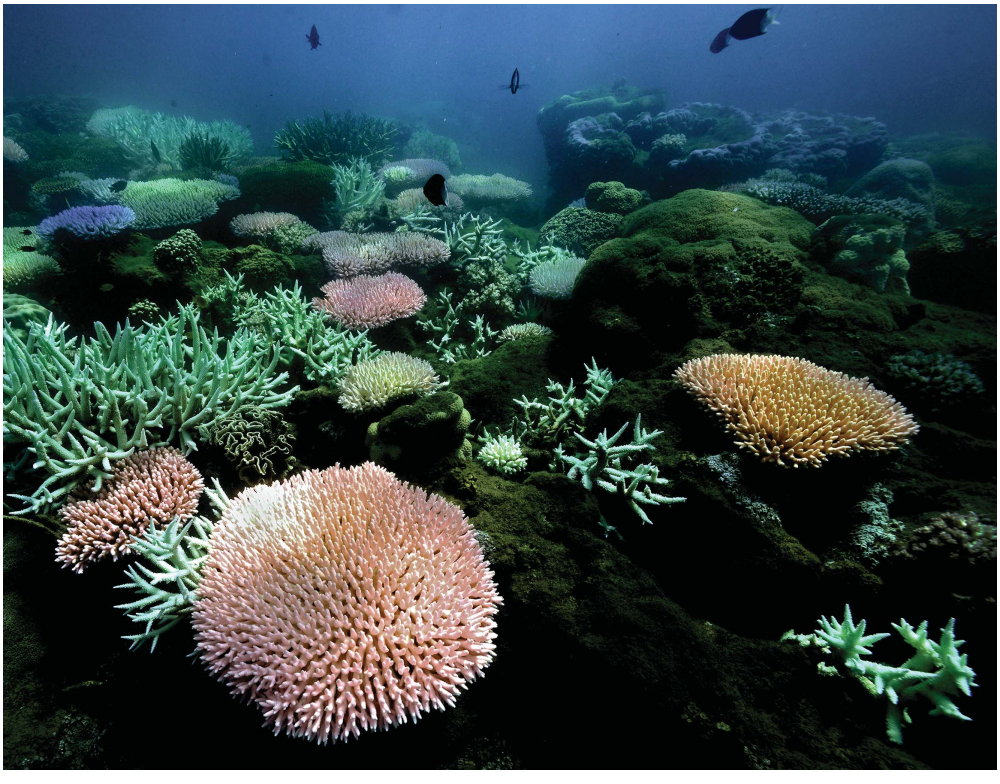
RATIONALE: Bleaching responses vary within and among coral species; in the reef-building coral *Acropora millepora*, a commonly distributed species across the Indo-Pacific, these differences have been shown to be at least partly heritable. In principle, therefore, interindividual differences in bleaching should be predictable from genomic data. Here, we demonstrate the feasibility of using a genomics-based approach to predict individual bleaching responses and suggest ways in which this can inform new strategies for coral conservation.

RESULTS: We first generated a chromosome-scale genome assembly as well as whole-genome sequences for 237 samples collected at 12 reefs distributed across the central Great Barrier Reef during peak bleaching in 2017. We showed that we can reliably impute genotypes in low-coverage sequencing data with a modestly sized reference haplotype panel, demonstrating a cost-effective approach for future large-scale whole-genome sequencing efforts. Very little population structure was detected across the sampled reefs, which was likely the result of the broadcast spawning mode of reproduction in *A. millepora*. Against this genomic background, we detected unusually old variation at the heat-shock co-chaperone *sacsin*, which is consistent with long-term balancing selection acting on this gene. Our genomic sequencing approach simultaneously provides a quantitative measure of bleaching and identifies the composition of symbiont species present within individual coral hosts. Testing more than 6.8 million variants for associations with three different measures of bleaching response, no single site reached genome-wide significance, indicating that variation in bleaching response is not due to common loci of large effect. However, a model that incorporates genetic effects estimated from the genome-wide association data, genomic data on relative symbiont species composition, and environmental variables is predictive of individual bleaching phenotypes.

ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.aba4674

CONCLUSION: Understanding the genetics of heat and bleaching tolerance will be critical to predict coral adaptation and the future of coral reef ecosystems under climate change. This knowledge also supports both conventional management approaches and the development of new interventions. Our work provides insight into the genetic architecture of bleaching response and serves as a proof of principle for the use of genomic approaches in conservation efforts. We show that a model based on environmental factors, genomic data from the symbiont, and genome-wide association data in the coral host can help distinguish individuals most tolerant to bleaching from those that are most susceptible. These results thus build a foundation toward a genomic predictor of bleaching response in *A. millepora* and other coral species. ■



Bleached *A. millepora* colonies on the central Great Barrier Reef. *A. millepora* colonies presenting various severity of bleaching during March 2017 at Feather Reef on the Great Barrier Reef. Colonies with a greater severity of bleaching are those with the most pale colors. Prolonged periods of bleaching can lead to the eventual death of the coral host.

PHOTO: ERIC MASON/AUSTRALIAN INSTITUTE OF MARINE SCIENCE

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zlf2102@columbia.edu (Z.L.F.); mp3284@columbia.edu (M.P.); l.bay@aims.gov.au (L.K.B.)
†These authors contributed equally to this work.
Cite this article as Z. L. Fuller *et al.*, *Science* **369**, eaba4674 (2020). DOI: 10.1126/science.aba4674

RESEARCH ARTICLE

CORAL GENOMICS

Population genetics of the coroal *Acropora millepora*: Toward genomic prediction of bleaching

Zachary L. Fuller[1]*, Veronique J. L. Mocellin[2], Luke A. Morris[2,3,4], Neal Cantin[2], Jihanne Shepherd[1], Luke Sarre[1], Julie Peng[5], Yi Liao[6,7], Joseph Pickrell[8], Peter Andolfatto[1], Mikhail Matz[6]†, Line K. Bay[2]*†, Molly Przeworski[1,9,10]*†

Although reef-building corals are declining worldwide, responses to bleaching vary within and across species and are partly heritable. Toward predicting bleaching response from genomic data, we generated a chromosome-scale genome assembly for the coral *Acropora millepora*. We obtained whole-genome sequences for 237 phenotyped samples collected at 12 reefs along the Great Barrier Reef, among which we inferred little population structure. Scanning the genome for evidence of local adaptation, we detected signatures of long-term balancing selection in the heat-shock co-chaperone *sacsin*. We conducted a genome-wide association study of visual bleaching score for 213 samples, incorporating the polygenic score derived from it into a predictive model for bleaching in the wild. These results set the stage for genomics-based approaches in conservation strategies.

Anthropogenic global warming is transforming ecosystems worldwide, pressing species to rapidly respond to changing environments (*1*–*4*). To what extent they will succeed is unclear; some may undergo range shifts or tolerate climate change through phenotypic plasticity. If there is sufficient heritable variation in traits that confer a fitness advantage in the new environment, others may adapt through evolutionary change (*5*–*9*). Particularly devastating has been the impact of increased seawater temperatures on marine ecosystems—notably on coral reefs, which cover less than 1% of the ocean floor but are home to more than a quarter of its total biodiversity (*10*). Ecological stress brought on by changes in temperature or salinity breaks down the symbiotic relationship between reef-building corals and their intracellular photosynthetic dinoflagellates (family Symbiodiniaceae) in a phenomenon known as bleaching (*11*). Because these symbionts provide most of the energy required to the coral host, prolonged periods of bleaching can lead to the death of the coral. Increased temperatures can also cause direct cellular damage and necrosis to bleached corals (*12*). Temperatures only slightly above long-term maxima can result in corals over vast areas losing their symbionts, decimating reefs worldwide (*13*–*15*).

Despite the vulnerability of corals to warming, there is phenotypic variation in the bleaching and heat stress response both within and among coral species (*16*–*18*). In the coral *Acropora millepora*, a common reef species across the Indo-Pacific, variation in heat tolerance among aposymbiotic larvae has been shown to be partly heritable (*19*). This finding implies both that a response to selection is possible and that bleaching should be partly predicted from genomic data. Simulation studies have suggested the potential for rapid adaptation in *A. millepora* populations (*20*), and gene expression analyses have associated genes related to oxidative, extracellular, transport, and mitochondrial functions with elevated heat tolerance (*19*). However, because previous studies rely on transcriptomics or reduced representation sequencing (*19*–*22*), they cannot distinguish causes from effects of bleaching, provide only a partial view of the genome, and are hampered by the lack of a high-quality, publicly available reference assembly. Thus, we remain far from a genetic understanding of bleaching and thermal tolerance and lack the ability to predict interindividual variation in the responses. In particular, the heritability of a trait does not specify its genetic architecture (*23*), so it remains unclear whether variation in heat tolerance is polygenic or influenced by a small number of common large-effect loci.

[1]Department of Biological Sciences, Columbia University, New York, NY, USA. [2]Australian Institute of Marine Science, Townsville, QLD, Australia. [3]AIMS@JCU, Australian Institute of Marine Science, College of Science and Engineering, James Cook University, Townsville, QLD, Australia. [4]College of Science and Engineering, James Cook University, Townsville, QLD, Australia. [5]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [6]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA. [7]Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, CA, USA. [8]Gencove, New York, NY, USA. [9]Department of Systems Biology, Columbia University, New York, NY, USA. [10]Program for Mathematical Genomics, Columbia University, New York, NY, USA.
***Corresponding author. Email: zlf2102@columbia.edu (Z.L.F.); mp3284@columbia.edu (M.P.); l.bay@aims.gov.au (L.K.B.)**
†These authors contributed equally to this work.

Inferring population structure and demographic history from genome-wide patterns of variation

Heritable traits can be predicted from the results of genome-wide association studies (GWASs) alongside other predictors, an approach widespread in agriculture and increasingly in human genetics (*24*–*26*). We applied a similar approach to conservation biology, focusing on prediction of bleaching in wild populations of *A. millepora*. To this end, we assembled a highly contiguous chromosome-level de novo *A. millepora* genome and generated high-coverage, whole-genome resequencing data for samples collected across 12 reefs on the Great Barrier Reef (GBR).

First, we constructed a chromosome-scale assembly of the diploid *A. millepora* genome using a combination of PacBio and 10X Chromium barcoded Illumina reads generated from DNA extracted from a single adult colony (Fig. 1A, figs. S1 to S6, table S1, and materials and methods). We additionally sequenced two pools of aposymbiotic larvae and mapped these reads to our de novo assembly to remove symbiont contigs. In total, we assembled 475 Mb of sequence with chromosome-scale scaffolds for an N50 of 19.8 Mb (where N50 is the minimum scaffold length needed to cover 50% of the genome) (figs. S1 to S6). Using previously published transcriptome data (*27*, *28*), we predicted 28,188 gene models, which include >96% of core-eukaryotic single-copy orthologs. A contiguous reference genome facilitates complex trait mapping (*29*) and reliable demographic inference (*30*).

To investigate genetic variation associated with bleaching response, we collected branch fragments from a total of 253 individual *A. millepora* colonies during peak bleaching in March 2017 from 12 reefs (Fig. 1B and table S2) distributed along an ~300-km stretch of the central GBR. Ecological conditions were measured by use of 40 environmental and spatial variables [recorded at a resolution of 1 km^2 (*31*)] that varied among reefs, with the strongest axes of variation being longitude and latitude (Fig. 1C). In particular, mean monthly sea surface temperatures differed by more than 0.5°C. The depth of colonies from which fragments were collected varied from 1.2 to 7.2 m, representing the range of depths where *A. millepora* is commonly found on reef flats and upper reef slopes (*32*). For each colony, we recorded the bleaching level visually using the six-point coral health chart to the nearest half increment (*33*) then later estimated photosynthetic pigment concentration and symbiont cell density (Fig. 1D). Individual colonies sampled presented a range of bleaching phenotypes, both within and between reefs (fig. S7).

From this collection of samples, we resequenced 48 whole genomes of diploid

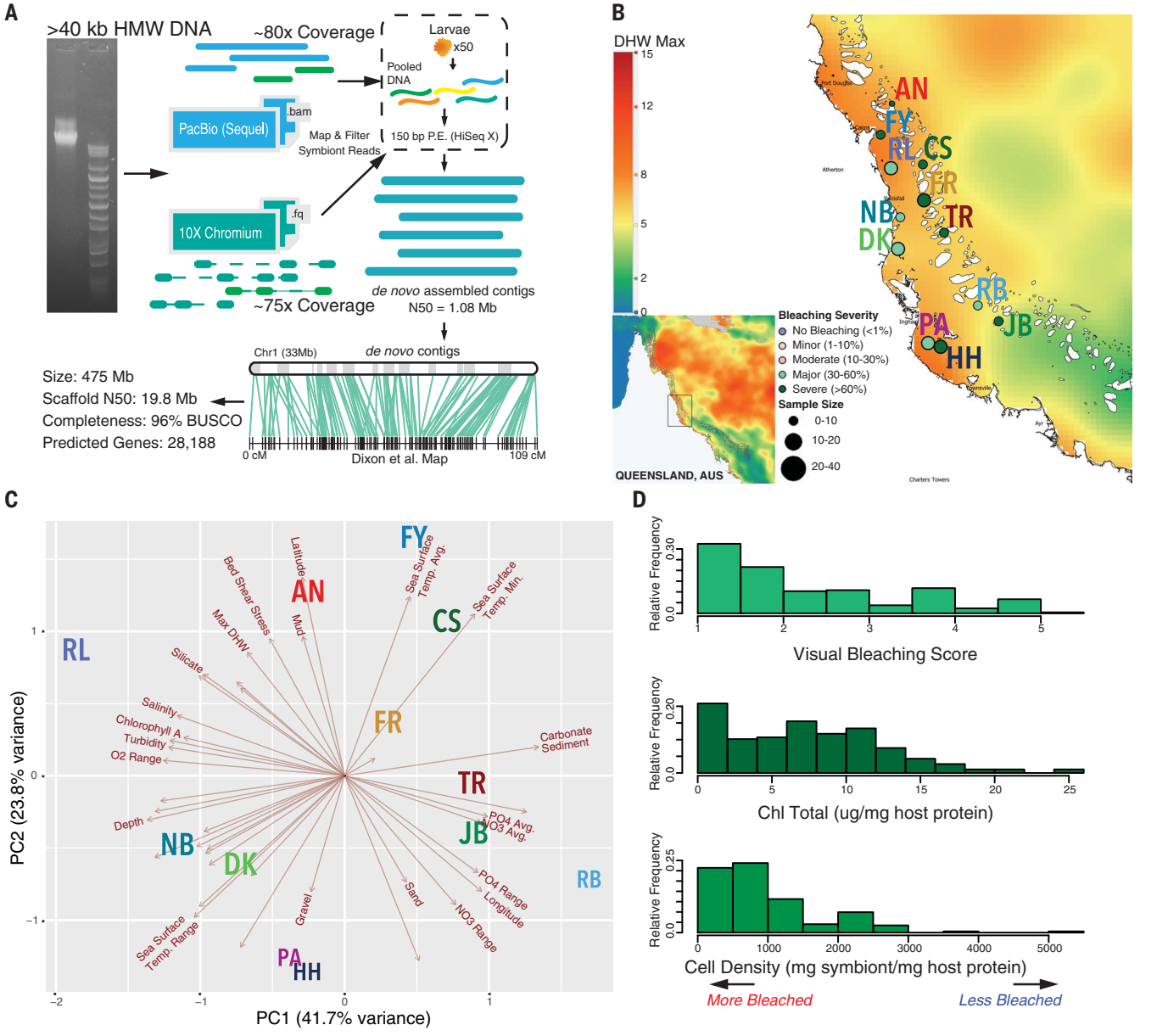


Fig. 1. Genome assembly and sample collection for *A. millepora*. (**A**) A de novo assembly for the *A. millepora* genome was constructed by using a hybrid sequencing approach. The alignment of reads generated from two pools of aposymbiotic larvae was used to filter out symbiont contigs (fig. S3). Assembled coral contigs were aligned to previously published linkage maps (*19*, *80*) to create a chromosome-scale assembly. (**B**) A total of 253 individuals were collected from across 12 reefs on the GBR in 2017. The size of the circles represents the number of individuals collected at each site, and the circles are colored according to the bleaching severity at each reef. The maximum degree heating weeks (DHW) is shown across the region from which samples were collected (and across the GBR in the inset). The spatial layer of DHW represents interpolated maximum values from the National Oceanic and Atmospheric Administration Coral Reef Watch (CRW) v3.1 satellite product at a resolution of 5 km. Each reef label is colored arbitrarily but consistent with labels presented in other figures. AN, Arlington Reef; FY, Fitzroy Reef; RL, Russell Island; CS, Coates Reef; FR, Feather Reef; NB, North Barnard Islands; TR, Taylor Reef; DK, Dunk Island; RB, Rib Reef; PA, Pandora Reef; JB, John Brewer Reef; HH, Havannah Island. (**C**) A PCA was performed for 40 environmental and spatial variables for each reef, with abbreviations in bold representing their location. The component loads for each environmental variable projected on the first two principal components are depicted with arrows. (**D**) The distribution of visual scores, chlorophyll abundance standardized by host coral protein content, and standardized symbiont cell densities among the collected individuals for which phenotype measurement was possible. A visual score of 1 indicates severe bleaching, whereas a score of 6 represents full pigmentation.

individuals (four from each of the 12 reefs) at high coverage (>20x for all, 36 samples >100x) (fig. S8 and table S3) to examine the extent of population structure, characterize the demographic history, and scan for signals of adaptation. We mapped reads from all individuals to our de novo genome assembly and called approximately 6.8 million biallelic single-nucleotide polymorphisms (SNPs) after applying stringent quality filters. We validated these filtered SNP calls by independently amplifying and sequencing 10 intergenic regions in eight samples at high coverage (>1400x) (fig. S10 and table S4), observing an overall genotype concordance of 99.6% between regions sequenced

by the two approaches. After an initial principal components analysis (PCA) (fig. S9) and test for outliers, we removed four individuals that were likely the result of sample misidentification at the time of collection (or conceivably represent rare diverged lineages) (fig. S16), bringing the total number of high-coverage resequenced genomes to 44. These 44 genomes are all approximately equidistantly related (fig. S11).

From this set of 44 unrelated individuals, we estimated the decay of pairwise linkage disequilibrium (LD), r^2, as a function of physical distance genome-wide. We observed that on average, r^2 falls below 0.05 after ~15 kb (Fig. 2A). We also estimated the nucleotide diversity (as π, the mean pairwise difference per base pair) in nonoverlapping intergenic regions of 1 kb, which yielded an average of π = 0.363% (Fig. 2B). Assuming a mutation rate of 4×10^{-9} per base pair per generation (*20*, *34*), this value of diversity corresponds to a long-term effective population size (N_e) of around 2.26×10^5. Applying the pairwise sequential Markovian coalescent (PSMC) (*35*) to each genome, we inferred how N_e has changed over the past ~10^6 generations (Fig. 2C). All samples display a highly similar demographic trajectory, with the inferred N_e showing a steady decline. Qualitatively similar results were obtained when demographic histories for the same samples were inferred jointly with the multiple sequentially Markovian coalescent (MSMC) (fig. S12) (*36*). The increase in estimated N_e back in time could reflect greater census sizes in the past or greater population structure in the ancestral population from which all current lineages similarly derive (*35*). A pattern of decline in estimated N_e has also been observed in other corals, including a number of *Acropora* and *Orbicella* species (*20*, *37*, *38*).

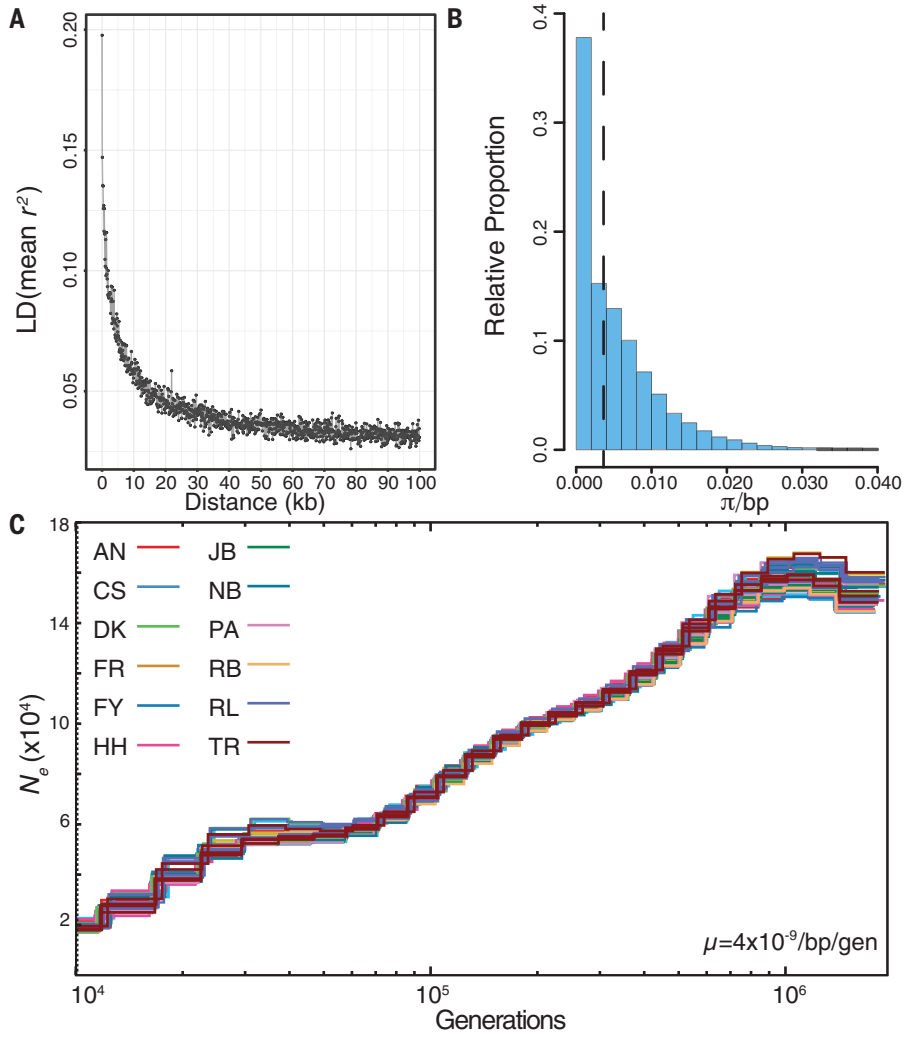


Fig. 2. Properties of genetic variation and inferred demographic history in sampled *A. millepora*. (**A**) The decay of LD, measured as the squared genotypic correlation coefficient (r^2), was estimated for a randomly chosen 1% of SNPs by using all sequenced samples. Each point represents the mean r^2 in bins of 100 bp. (**B**) Nucleotide diversity was measured by the average pairwise differences per base pair (π) (*86*) for SNPs in 1-kb windows, using intergenic regions genome-wide. (**C**) Effective population sizes (N_e) inferred from 44 sequenced *A. millepora* genomes by using PSMC (*35*), assuming a mutation rate μ of 4×10^{-9} per base pair per generation (*20*, *34*). Although an error in the estimate of μ will lead to a rescaling of both axes, it will not affect the qualitative conclusion of a decline in N_e toward the present. The *x* axis is scaled in terms of generations. This approach provides very little information about the recent past (here, less than ~10^4 generations).

The concordant long-term demographic histories inferred across all individuals suggest high connectivity among the 12 sampled reefs. Accordingly, there is no detectable relationship between geographic distance and fixation index (F_{ST}) between pairs of sampled reefs (Fig. 3A). A PCA of SNPs in approximate linkage equilibrium further reveals no obvious clusters (Fig. 3B). In particular, although environmental conditions differ between inshore and offshore reefs (Fig. 1C), F_{ST} between these samples is approximately 0 (-2.76×10^{-4}). More generally, no significant relationships were detected between either of the first two eigenvectors and any of the 40 environmental or spatial variables for each reef, indicating that any population structure that is present is at most very weakly associated with our measured environmental variables. The lack of population structure can be visualized by estimating and plotting relative effective migration rates. Using the program EEMS (*39*), we inferred homogenous migration rates and no obvious barriers to gene flow among sampled reefs (Fig. 3C). This pattern of gene flow across large geographic distances on the central GBR is presumably a result of its broadcast spawning mode of reproduction (*10*, *17*, *19*, *20*, *40*–*42*). These findings are consistent with previous inferences based on the SNP allele frequency spectrum in *A. millepora* (*20*) as well as analyses of microsatellites in *Acropora tenuis* (*43*). However, in other Indo-Pacific Acroporids, such as *Acropora cytherea* and *Acropora hyacinthus*, high genetic divergence, possibly representing cryptic species, has been observed among corals over smaller geographic scales (*44*). Thus, not all broadcast spawning coral species show low population structure at comparable geographic scales.

A scan for adaptation points to a heat-shock protein co-chaperone

The near absence of genetic differentiation across reefs indicates that there is little population structure over hundreds of kilometers. Yet, environmental conditions differ among reefs, notably in terms of thermal regimes (Fig. 1B). Because adults only reproduce with those individuals in their vicinity whereas larvae can disperse over much larger ranges (*16*, *20*), larvae may experience strong selection each generation, as they settle on a heterogeneous

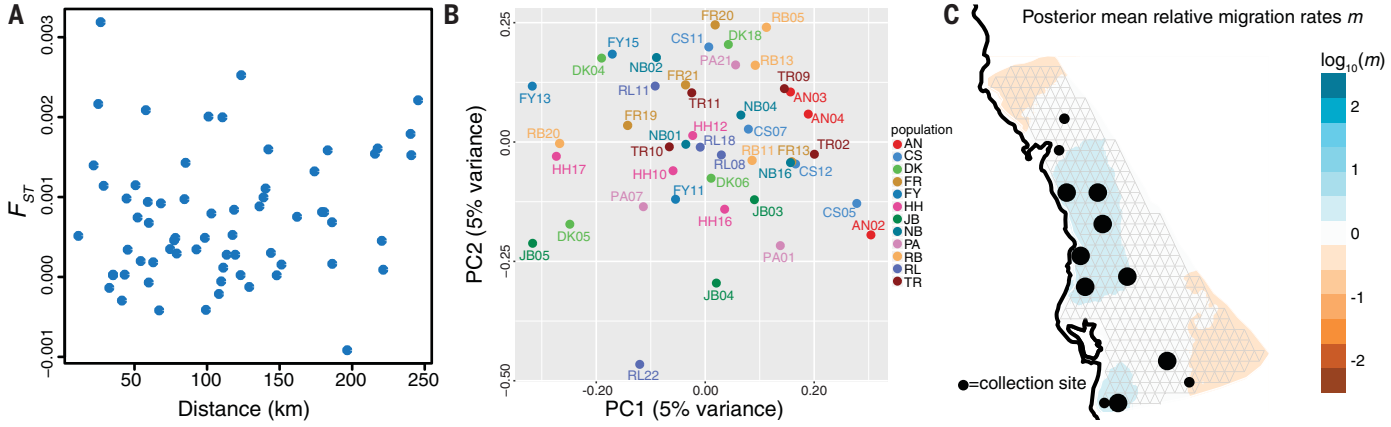


Fig. 3. Characterizing population structure and gene flow across 12 reefs. (**A**) There is no discernable relationship between geographic distance and genetic differentiation (measured as F_{ST}) across pairwise comparisons of sampled reefs. (**B**) A PCA by using LD-pruned genome-wide SNPs for the 44 resequenced high-coverage genomes. Each point is color-coded by the reef from which the individual was sampled. (**C**) EEMS (*39*) was used to estimate and visualize the relative effective migration surface by using LD-pruned and common (MAF > 5%) genome-wide SNPs for the 44 resequenced high-coverage genomes. The size of the dots reflects the number of individuals sequenced from each reef.

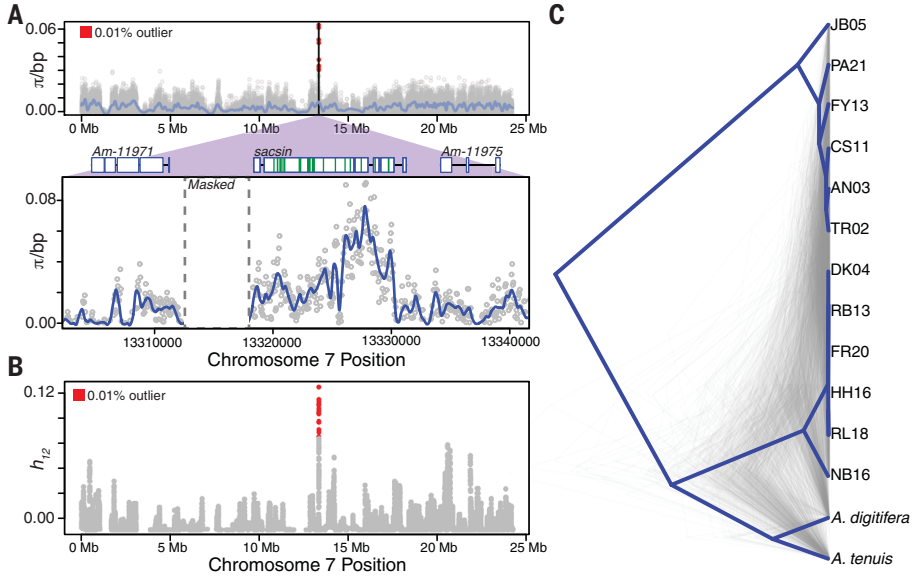


Fig. 4. Genomic scans for local adaptation detect a signal at *sacsin*. (**A**) (Top) Values of pairwise nucleotide diversity (π) (*86*) estimated in sliding windows of 1 kb and a step-size of 500 bp across chromosome 7, with points colored in red indicating regions in the top 0.01% of values genome-wide. A loess-smoothed trend line of π across windows is indicated in purple. The peak with the most extreme values falls in the *sacsin* gene region. (Bottom) A close-up view of π estimated across a region of chromosome 7 surrounding the *sacsin* gene, which includes all of the outlier windows indicated in red at top. π per base pair was calculated in sliding windows of 100 bp (gray dots); the loess-smoothed line is shown in blue. The predicted gene structure of *sacsin* is indicated above, as well as the flanking upstream and downstream genes. The 29 nonsynonymous differences fixed between samples of the two most common haplotypes in *sacsin* are denoted with green lines. There are also 20 fixed synonymous differences. Gray vertical dashed lines delimit a region that was masked from variant calling because of predicted repetitive elements. (**B**) The h_{12} summary statistic (*47*), which measures the frequency of the two most common haplotypes, is plotted across chromosome 7 (supplementary materials, materials and methods). Red dots represent h_{12} values in the top 0.01% genome-wide. The peak with the most extreme values falls in the *sacsin* gene region. (**C**) A gene tree for the central 1-kb region in *sacsin* is shown in dark blue; it was constructed for one randomly chosen haplotype from a randomly selected individual from each of the 12 sampled reefs. The gene tree is rooted to aligned sequences from the reference genomes for *A. digitifera* (*87*) and *A. tenuis* (draft assembly from www.reefgenomics.org). Shown in gray are gene trees for 1000 randomly sampled 1-kb regions from across the genome for the same individuals. Each gene tree was inferred by using maximum likelihood implemented in *dnaml* (*88*).

reefscape. Models suggest that this scenario of strong, spatially varying selection in the face of continual migration can lead to the maintenance of locally beneficial alleles over long periods of time (*45*, *46*). We therefore searched for signals of adaptation in the 44 resequenced genomes by scanning for genomic regions displaying the high levels of diversity that would result from such long-lived balancing selection. Genome-wide, the most extreme values (π > 5.0%, measured in 1-kb windows) (Fig. 4A) fell over a region containing a single gene, which is orthologous to *sacsin* in *Acropora digitifera*. The same genic region also has a highly elevated value of the h_{12} statistic (*47*), which measures the frequencies of the two most common haplotypes (Fig. 4B). Confirming that this pattern of relatedness is unusual, a PCA for the region stands out by comparison with the rest of the genome [as assessed with *lostruct* (*48*)] (fig. S13). These summary statistics point to an unusually deep genealogy with long internal branches at *sacsin*; the divergence of two haplotypes even predates the species split with *A. tenuis* and *A. digitifera* (Fig. 4C and fig. S15). There was no significant elevation in coverage surrounding *sacsin* for any sample to suggest the presence of a paralog (fig. S17). To further ensure that these findings were not artifactual, we confirmed the presence of the two diverged

haplotypes by independently sequencing and assembling overlapping 2-kb amplicons spanning the entire gene for a subset of four samples, which include homozygotes for both alternative haplotypes and a heterozygote (fig. S18 and tables S6 and S7). In a de novo assembly of sequences from both haplotypes across the coding region, each amplicon aligned in the expected order (fig. S19).

Sacsin is a co-chaperone for the heat-shock protein Hsp70 and contains multiple regions of homology to *Hsp90* (*49*, *50*). *Sacsin* has been shown to be significantly up-regulated when exposed to elevated temperatures in the laboratory in multiple coral species from the genus *Acropora* and *Pocillipora* (*51*–*54*), indicating that variation in the gene is plausibly associated with heat response in natural populations. Our analyses reveal that variation in *sacsin* has been maintained for an atypically long time (millions of generations), indicating that the gene is a target of balancing selection and hence of functional importance in the wild as well as in the laboratory. The source of balancing selection is unclear, however; there is no discernable association of the haplotype frequencies with genetic or current environmental principal components ($P > 0.15$) (table S5), although there may be an association with the dominant symbiont type ($P = 0.059$).

A GWAS for bleaching

Using the 44 high-coverage genomes, we used read-backed phasing (*55*) to construct a reference haplotype panel and impute genotypes in low-pass sequences of additional samples. We generated a mean of ~1.5x coverage whole-genome data (*56*) for 193 individuals after applying filters based on quality, read depth, and the proportion of missing data (fig. S21). Of these samples, 34 had also been sequenced at high coverage, allowing us to assess the accuracy of our imputation approach. In the 34 samples, the overall correlation between sequenced and imputed genotypes was >94% for SNPs with minor allele frequency >0.05 and genotype probability ≥0.95 (fig. S22), an imputation accuracy similar to what is observed in human data that use reference panels of similar size (*57*, *58*).

Combining the low-coverage genomes with the 44 high-coverage ones, we obtained genotype calls at ~6.8 million sites in a total of 237 total samples. After excluding outliers that were likely species misidentifications or possibly rare diverged lineages (*59*), individuals with high levels of relatedness, and those with missing phenotype data (figs. S23 to S24), our final sample size was 213. Consistent with our previous results, in the larger sample we again observed minimal population structure, evidence for high gene flow, and a peak of elevated genetic diversity at *sacsin* (fig. S25). Although the low-population structure is helpful for a GWAS, it makes precise estimates of SNP heritability unattainable without substantially larger sample sizes (*60*, *61*).

The sequencing also yielded data from the symbiont Symbiodiniaceae, of which multiple species are known to associate with *A. millepora* (*62*, *63*). Because intracellular symbiont DNA was extracted and sequenced simultaneously with the coral host, we used our genomic sequencing approach to characterize the symbiont species present in each sample (*11*, *59*). The abundance of symbiont reads in each sample was significantly different between quartiles of visual scores ($P = 2.6 \times 10^{-7}$ by means of a Kruskal-Wallis test) (Fig. 5A) as well as correlated with total chlorophyll content and symbiont cell density (fig. S27). We further determined the composition of symbiont types for each sample by calculating the relative proportion of reads that map to available draft genomes. In our samples, individuals dominated by *Durusdinium* had significantly higher visual scores on average ($P = 2.46 \times 10^{-7}$ by means of a Mann-Whitney *U* test) (Fig. 5B), which is consistent with previous findings that *Durusdinium* (formerly *Symbiodinium* Clade D) is associated with substantially higher thermal tolerance (but often lower growth rates) across diverse coral genera in the wild (*11*, *64*–*66*). Our results demonstrate that a low-coverage sequencing approach can be used to simultaneously obtain a quantitative measure of bleaching and determine the relative composition of symbiont types.

Given these data from corals and their symbionts, we used a linear mixed model (LMM) to test for additive effects of SNPs (with minor allele frequencies >5%) on the quantile-normalized visual score (*67*, *68*), including as covariates the top four principal components of environmental variables (which together account for >85% of the total environmental variance), the first two genetic PCs, batch effects (the collection date and sequencing batch), the collection depth (*69*), and the proportion of *Durusdinium* reads relative to all symbiont reads. We also conducted similar GWASs for total chlorophyll content and symbiont cell density (fig. S28). The three measures of bleaching are highly correlated (Spearman's correlation coefficient among pairs of phenotypes >0.6) (fig. S26); because we are missing phenotype data for chlorophyll content (so that $n = 190$ samples) and symbiont cell density ($n = 172$ samples), we focused on the standard visual score ($n = 213$ samples).

We determined a genome-wide significance threshold of $P = 4.3 \times 10^{-8}$ for the quantile-normalized visual score by considering the distribution of *P* values obtained from 10^5 permutation tests (fig. S34). With this approach, no single SNP in our GWAS is genome-wide significant, indicating the absence of common, major effect loci associated with variation in bleaching. In particular, the minimum *P* value in the *sacsin* gene is high (0.01); for related phenotypes of total chlorophyll content and symbiont cell density, however, SNPs in *sacsin* are among the top signals (minimal *P* values are 5.05×10^{-5} and 2.38×10^{-5}, respectively),

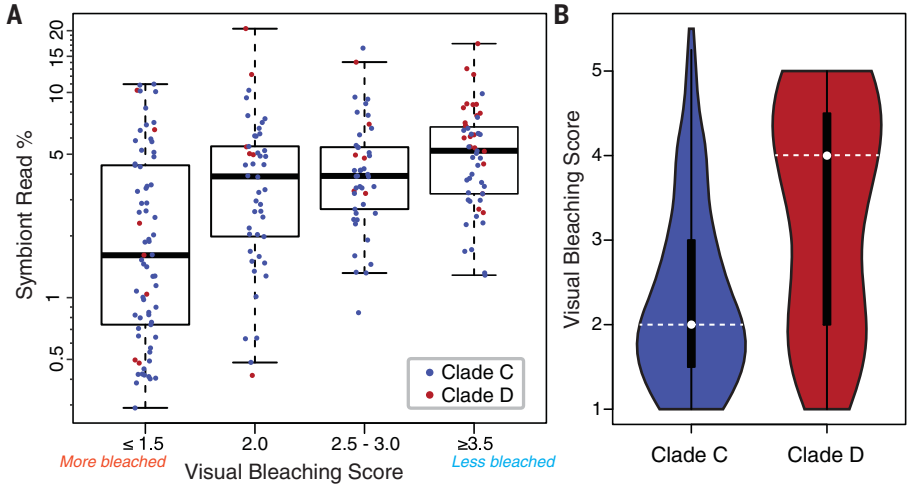


Fig. 5. Relationship between symbiont species and bleaching response revealed with high-throughput sequencing. (**A**) The total proportion of reads mapping to draft symbiont genomes was measured for each sample. Bins for the visual scores were chosen so that approximately equal numbers of samples were included in each. Points are color-coded according to the dominant symbiont genus found in each sample. The *y* axis is shown on the $\log_{10}$ scale. (**B**) Samples dominated by *Durusdinium* (formerly *Symbiodinium* clade D) have a significantly higher ($P = 2.46 \times 10^{-7}$ by means of a two-tailed Mann-Whitney *U* test) visual score as compared with samples dominated by *Cladocopium* (formerly *Symbiodinium* clade C). White dotted lines in each violin plot represent the median.

suggesting that the lack of signal in the visual score GWAS may reflect lack of power. For visual score, the top peaks were a region encompassing two genes on chromosome 13 and one with three genes on chromosome 14 [also detected with the software BIMBAM (fig. S33) (70)] (Fig. 6A). These associations remain to be confirmed, but among the three genes on chromosome 14 is malate synthase, a key component of the glyoxylate cycle, a pathway not found in metazoans outside of cnidarians (71). Malate synthase is up-regulated in response to thermal stress in *A. palmata* and implicated in the coral stress response (72, 73).

Predicting individual responses in natural populations

The GWAS results in 213 diploid individuals indicate that variation in bleaching is not due to common loci of large effect but instead is polygenic, as has also been suggested for thermal tolerance in *A. hyacinthus* (22), so that our study is underpowered to identify individual causal loci. Nonetheless, when combined, the estimated effects of variants across the genome may be predictive of phenotype (25). We therefore constructed a polygenic score [PGS; also called breeding value (23)] using LD-based clumping (for a *P* value threshold of 10^{-5}) (fig. S35) and the effect sizes estimated in the GWAS (74). Because we lacked a true out-of-sample validation set, we assessed the prediction accuracy of the PGS using a jackknife cross-validation procedure with an 85/15% training-test split in each partition; similar results were also obtained for 80/20% and 90/10% training-test splits (fig. S36).

Given that both environmental and genetic factors contribute to variation in bleaching, we combined our PGS with other predictors, including environmental effects and the dominant symbiont species type. Using a cross-validation procedure, we assessed the improvement in prediction accuracy gained from including these predictors in a linear model (Fig. 6B). Compared with a model including only batch effects and genetic principal components, we observed a large increase in prediction accuracy using four principal components for the environmental variables as well as a substantial improvement when we included data obtained from our genomic sequencing—namely, the proportion of *Durusdinium* reads out of all symbiont reads and the PGS (Fig. 6B). Including the PGS alone provided a small but significant increase in prediction accuracy [$P = 0.023$; mean incremental coefficient of determination (R^2) = 0.026]; the same was true in models that used the top six or eight environmental PCs instead (which in total explain 99% of the variance in environmental variables; $P \leq 0.029$) (fig. S37). In total, the model explains an estimated ~62% of the variance in quantile-normalized visual scores in our sample. However, moving forward it will be important to test its performance on a distinct sample rather than using cross-validation as we have of necessity done here.

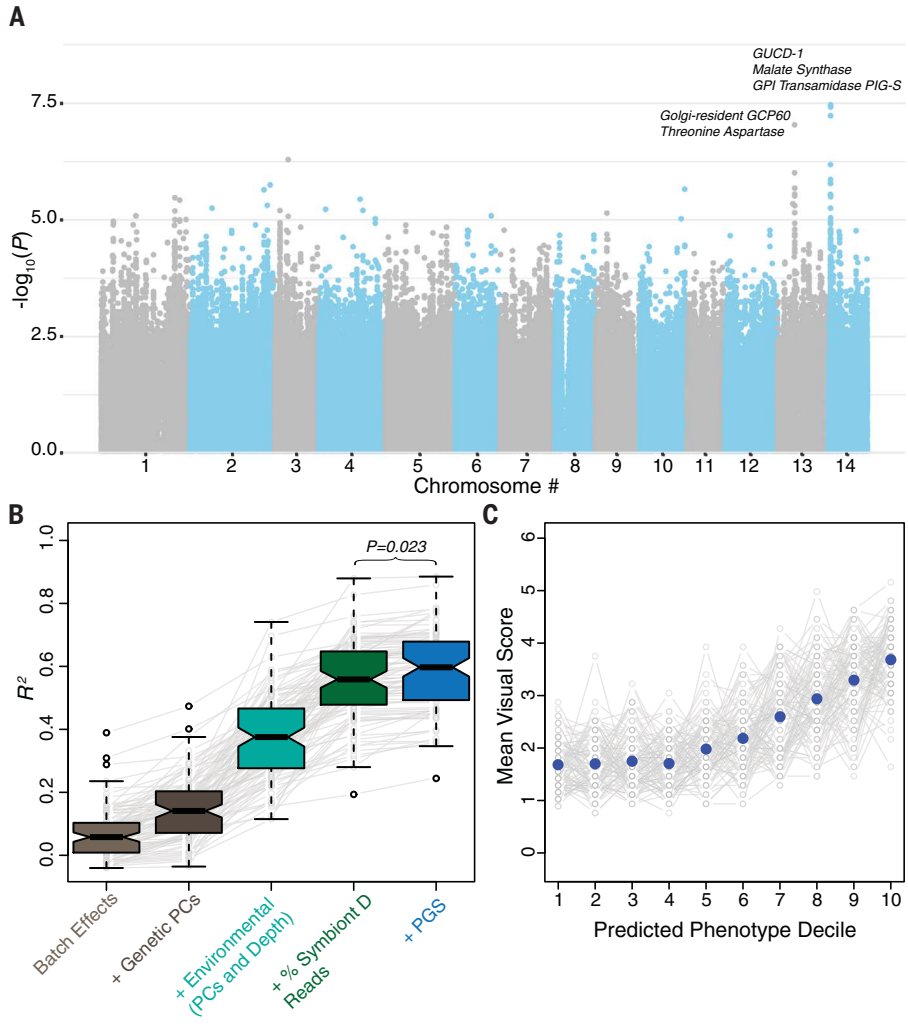


Fig. 6. A GWAS and prediction accuracy for bleaching response. (**A**) A Manhattan plot of *P* values obtained from testing for an association of each SNP with variation in visual score. Shown are the names of all genes in the two top peaks, on chromosome 13 and 14. (**B**) The increase in prediction accuracy (measured as R^2) for linear models with different sets of predictors. Each boxplot shows the distribution of R^2 obtained for 100 jackknife cross-validation partitions, with a random 15% of the total sample withheld as a test set in each. Batch effects include the collection date and the sequencing batch. Each boxplot shows the R^2 for sets of predictors in a linear model. Mann-Whitney *U* tests were performed to test for differences in the distribution of R^2 between models. The average number of SNPs used in the PGS in each jackknife partition is 15.32. (**C**) Using the full predictive model, individuals with the highest predicted trait values have the highest visual scores on average. Within each jackknife partition, the test set was divided into deciles according to the predicted phenotype. The blue dot for each decile indicates the mean visual score, and gray lines connect deciles in a given jackknife partition.

Implications

Understanding how species will respond to increasing temperatures is key to ensuring adequate protection through conventional management approaches and to supporting interventions aimed at restoration or facilitating adaptation to prevent future losses. In this work, we focused on bleaching in the coral *A. millepora*, a trait of central ecological importance, demonstrating the feasibility of predicting interindividual response from genomic and environmental data. The lack of genome-wide significant associations suggests that the differences in bleaching response are due to loci that individually explain only a small proportion of the variance. As has been the case in other species, we therefore expect that as sample sizes increase, so will the prediction accuracy of the PGS and hence of the model. With larger samples, it will also become possible to test for interactions between the coral genome and environmental factors or symbiont types

as well as to estimate the heritability of bleaching in the wild and hence learn more about the potential adaptive response.

In the development of effective conservation strategies, a predictive model of bleaching can help distinguish those individuals more likely to be tolerant on reefs across different shelf positions, latitudes, and environmental conditions, informing current spatial protection strategies and predictions about reef futures (*75*, *76*). As an illustration, using our cross-validation procedure and full predictive model, individuals in the top quartile of predicted phenotype values have significantly higher visual scores on average than those in the lowest quartile ($P = 8.8 \times 10^{-16}$ by means of a one-tailed Wilcoxon signed rank test) (Fig. 6C). The approach may also help to identify the most tolerant individuals to use for managed translocations and selective breeding for assisted evolution–based conservation strategies (*77*). In that regard, the lack of population structure in *A. millepora* over large geographic distances makes it an ideal system because one could conceivably breed corals with high predicted trait values from multiple locations. Alternatively, such predictions could be used to increase the abundance of such corals in local populations rather than moving them through assisted gene flow, which carries risks and ethical concerns.

Beyond the application to corals, our study demonstrates the feasibility of high-quality genome assembly, imputation, and GWASs in a nonmodel organism. Moreover, our work highlights the potential of combining genomic and environmental data to predict phenotypes of ecological importance—notably, traits that are likely to play a role in the response to climate change.

Materials and methods

Genome assembly and annotation

The de novo genome assembly of *A. millepora* was constructed by using a combination of PacBio reads and Illumina paired-end reads with 10X Genomics Chromium barcodes. Sequencing was performed by the New York Genome Center and Cold Spring Harbor Laboratory. We constructed PacBio-only assemblies using the software Canu (*78*) and Falcon (*79*) and used the 10X barcodes and paired-end reads for scaffolding and error correction. The high–molecular weight DNA used for the assembly was extracted from an adult specimen, which also contained symbiont DNA. To remove potential symbiont contigs from the coral genome assembly, we additionally generated paired-end reads from two pools of aposymbiotic larvae and mapped these reads to our draft assembly. Using an unsupervised clustering approach, we identified and removed assembled contigs that appeared to be derived from the symbiont and masked regions of the assembly with no reads from the larvae. To assemble chromosome-scale scaffolds, we used previously published linkage maps (*19*, *80*) inferred from SNP and restriction site–associated DNA sequencing (RAD-seq) markers to linearly arrange contigs. Previously published transcriptome data (*27*, *28*) were used to predict gene models and annotate the resulting assembly.

Sample collection and phenotyping

Coral samples were collected from 12 reefs in the central region of the GBR at the height of mass bleaching between 14 March and 1 April 2017. For each colony, the bleaching score was estimated visually by using the Coral Color Reference Card (*33*) to the nearest color, including half increments. During collection dives, samples were sent to the surface every 5 min and immediately fixed in liquid nitrogen for further analysis. The total chlorophyll content was estimated in three technical replicates by measuring absorbance at 665, 649, and 632 nm on a microplate reader (synergy H4, BioTek Instruments, USA). The total coral host protein was similarly measured in triplicate by using the DC Protein Assay (Bio-Rad Laboratories, USA) following the manufacturers microplate protocol. Ecological conditions at each reef were measured with environmental and spatial variables recorded at a resolution of 1 km^2 (*31*).

Whole-genome sequencing, variant calling, and genotype imputation

The libraries for samples sequenced at high coverage were prepared by using the Illumina TruSeq polymerase chain reaction (PCR)–free protocol and sequenced on an Illumina HiSeqX platform. Libraries for samples sequenced at low coverage were prepared for multiplexed shotgun genotyping with Tn5 transposase and tagmentation (*56*). The Tn5 libraries were sequenced on an Illumina HiSeq4000 platform across 10 lanes. Illumina adapters and polyadenylate [poly(A)] sequences were trimmed (*81*) and low-quality reads were removed before mapping to our *A. millepora* genome assembly with *bwa-mem* (*82*). Variants were called by using Genome Analysis Toolkit (GATK) and filtered according to GATK best practices (*83*). The algorithm we used for genotype imputation in the low-coverage sequencing data are based on the copying model of Li and Stephens (*84*) and implemented in the software *loimpute* developed by Gencove and available at www.gencove.com.

Population genetic analyses

We used the software PLINK (*74*) to estimate the decay of pairwise LD as a function of physical distance, to perform PCA, and to estimate F_{ST}. To estimate nucleotide diversity, we used a custom wrapper for vcftools (*85*) to take into account missing data and uncalled sites. We inferred historical changes in N_e using the software PSMC (*35*), for which we increased the stringency of the quality filter to 30 and used a bin size of 10 base pairs (bp). We used EEMS (*39*) to visualize effective migration rates across the 12 sampled reefs, setting the number of MCMC iterations to 5×10^6. The software lostruct (*48*) was used to identify localized genomic regions in which patterns of relatedness differ from what is typical for the genome, under default parameters. We used the read-aware phasing module implemented in SHAPEIT2 (*55*) to infer haplotypes in the high-coverage genomes. To identify localized genomic regions with high frequencies of the two most common haplotypes, we calculated the h_{12} summary statistic in window sizes of 100 SNPs using a custom Python script (*47*).

Genome-wide association tests and construction of a PGS

We performed GWASs for bleaching measured as either a visual score, a total chlorophyll content, or a symbiont cell density. For each GWAS, we removed samples with missing phenotype data. The set of covariates included in all GWAS models were the top four PCs of environmental and spatial variables, the first two genetic PCs, the collection date, the sequencing batch, collection depth, and the relative proportion of *Durusdinium* reads (out of all symbiont reads). All GWASs were performed by using standard linear regression implemented in PLINK (*74*), and phenotypes were quantile-normalized. Only common variants [minor allele frequency (MAF) ≥ 0.05] and sites with less than 10% of genotypes missing were considered. We additionally performed a GWAS for visual score using a LMM implemented in GEMMA (*67*, *68*) and in a Bayesian framework using the software BIMBAM (*70*). We set cutoffs for genome-wide significance for each phenotype by performing permutation tests. Specifically, we shuffled the values of each phenotype across samples 10,000 times; for each permutation, we performed a standard linear GWAS using the same covariates and assessed the minimal P value. We then set the genome-wide significance threshold as the 95th percentile of the distribution of these minimal permuted P values.

Because we lacked a true out-of-sample validation set to assess the accuracy of PGSs constructed from GWAS, we used a jackknife cross-validation (CV) procedure to subset the samples into 100 partitions of training and test sets. In each partition, we randomly withheld 15% of samples as the test set and selected the other 85% of individuals as the training set. We recalculated the genetic PCs for the training set and used the loadings from this PCA for the genetic PCs in the test set. For each jackknife partition, a standard GWAS for quantile-normalized visual score was performed on the training set. To build the PGS, we used a

P value threshold of 1×10^{-5} followed by LD-clumping to choose sets of approximately independent SNPs. For each set of selected SNPs, the PGS was calculated by summing the allelic dosages weighted by their estimated effect size. To assess prediction accuracy, a PGS was then estimated in the test set. We assessed the prediction accuracy of the PGS along with other factors that may influence bleaching (such as environmental principal components, depth, and proportion of *Durusdinium*) in linear models. Mann-Whitney *U* tests were performed to test for differences in predicted trait values across jackknife partitions across different sets of predictors.

REFERENCES AND NOTES

1. C. Bellard, C. Bertelsmeier, P. Leadley, W. Thuiller, F. Courchamp, Impacts of climate change on the future of biodiversity. *Ecol. Lett.* **15**, 365–377 (2012). doi: 10.1111/j.1461-0248.2011.01736.x; pmid: 22257223
2. C. Parmesan, G. Yohe, A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003). doi: 10.1038/nature01286; pmid: 12511946
3. S. A. Henson *et al*., Rapid emergence of climate change in environmental drivers of marine ecosystems. *Nat. Commun.* **8**, 14682 (2017). doi: 10.1038/ncomms14682; pmid: 28267144
4. M. Pacifici *et al*., Assessing species vulnerability to climate change. *Nat. Clim. Chang.* **5**, 215–224 (2015). doi: 10.1038/nclimate2448
5. A. A. Hoffmann, C. M. Sgrò, Climate change and evolutionary adaptation. *Nature* **470**, 479–485 (2011). doi: 10.1038/nature09670; pmid: 21350480
6. I.-C. Chen, J. K. Hill, R. Ohlemüller, D. B. Roy, C. D. Thomas, Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024–1026 (2011). doi: 10.1126/science.1206432; pmid: 21852500
7. G.-R. Walther *et al*., Ecological responses to recent climate change. *Nature* **416**, 389–395 (2002). doi: 10.1038/416389a; pmid: 11919621
8. J. Merilä, A. P. Hendry, Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evol. Appl.* **7**, 1–14 (2014). doi: 10.1111/eva.12137; pmid: 24454544
9. M. Kelly, Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philos. Trans. R. Soc. London B Biol. Sci.* **374**, 20180176 (2019). doi: 10.1098/rstb.2018.0176; pmid: 30966963
10. D. J. Ayre, T. P. Hughes, Climate change, genotypic diversity and gene flow in reef-building corals. *Ecol. Lett.* **7**, 273–278 (2004). doi: 10.1111/j.1461-0248.2004.00585.x
11. T. C. LaJeunesse *et al*., Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Curr. Biol.* **28**, 2570–2580.e6 (2018). doi: 10.1016/j.cub.2018.07.008; pmid: 30100341
12. N. Traylor-Knowles, Heat stress compromises epithelial integrity in the coral, *Acropora hyacinthus*. *PeerJ* **7**, e6510 (2019). doi: 10.7717/peerj.6510; pmid: 30828497
13. T. P. Hughes *et al*., Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017). doi: 10.1038/nature21707; pmid: 28300113
14. T. P. Hughes *et al*., Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science* **359**, 80–83 (2018). doi: 10.1126/science.aan8048; pmid: 29302011
15. T. P. Hughes *et al*., Global warming transforms coral reef assemblages. *Nature* **556**, 492–496 (2018). doi: 10.1038/s41586-018-0041-2; pmid: 29670282
16. D. J. Ayre, T. P. Hughes, Genotypic diversity and gene flow in brooding and spawning corals along the Great Barrier Reef, Australia. *Evolution* **54**, 1590–1605 (2000). doi: 10.1111/j.0014-3820.2000.tb00704.x; pmid: 11108587
17. M. J. H. Van Oppen, L. M. Peplow, S. Kininmonth, R. Berkelmans, Historical and contemporary factors shape the population genetic structure of the broadcast spawning coral, *Acropora millepora*, on the Great Barrier Reef. *Mol. Ecol.* **20**, 4899–4914 (2011). doi: 10.1111/j.1365-294X.2011.05328.x; pmid: 22026459
18. J. R. Guest *et al*., Coral community response to bleaching on a highly disturbed reef. *Sci. Rep.* **6**, 20717 (2016). doi: 10.1038/srep20717; pmid: 26876092
19. G. B. Dixon *et al*., CORAL REEFS. Genomic determinants of coral heat tolerance across latitudes. *Science* **348**, 1460–1462 (2015). doi: 10.1126/science.1261224; pmid: 26113720
20. M. V. Matz, E. A. Treml, G. V. Aglyamova, L. K. Bay, Potential and limits for rapid genetic adaptation to warming in a Great Barrier Reef coral. *PLOS Genet.* **14**, e1007220 (2018). doi: 10.1371/journal.pgen.1007220; pmid: 29672529
21. K. E. Dziedzic, H. Elder, H. Tavalire, E. Meyer, Heritable variation in bleaching responses and its functional genomic basis in reef-building corals (*Orbicella faveolata*). *Mol. Ecol.* **28**, 2238–2253 (2019). doi: 10.1111/mec.15081; pmid: 30913323
22. N. H. Rose, R. A. Bay, M. K. Morikawa, S. R. Palumbi, Polygenic evolution drives species divergence and climate adaptation in corals. *Evolution* **72**, 82–94 (2018). doi: 10.1111/evo.13385; pmid: 29098686
23. D. S. Falconer, T. F. C. Mackay, *Introduction to Quantitative Genetics* (Pearson, ed. 4, 1996).
24. N. R. Wray, K. E. Kemper, B. J. Hayes, M. E. Goddard, P. M. Visscher, Complex trait prediction from genome data: Contrasting EBV in livestock to PRS in humans. *Genetics* **211**, 1131–1141 (2019). doi: 10.1534/genetics.119.301859; pmid: 30967442
25. E. S. Buckler *et al*., The genetic architecture of maize flowering time. *Science* **325**, 714–718 (2009). doi: 10.1126/science.1174276; pmid: 19661422
26. A. V. Khera *et al*., Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018). doi: 10.1038/s41588-018-0183-z; pmid: 30104762
27. A. Moya *et al*., Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO_2-driven acidification during the initiation of calcification. *Mol. Ecol.* **21**, 2440–2454 (2012). doi: 10.1111/j.1365-294X.2012.05554.x; pmid: 22490231
28. D. Bhattacharya *et al*., Comparative genomics explains the evolutionary success of reef-forming corals. *eLife* **5**, e13288 (2016). doi: 10.7554/eLife.13288; pmid: 27218454
29. J. Benevenuto, L. F. V. Ferrão, R. R. Amadeu, P. Munoz, How can a high-quality genome assembly help plant breeders? *Gigascience* **8**, giz068 (2019). doi: 10.1093/gigascience/giz068; pmid: 31184361
30. K. Nadachowska-Brzyska, R. Burri, L. Smeds, H. Ellegren, PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol. Ecol.* **25**, 1058–1072 (2016). doi: 10.1111/mec.13540; pmid: 26797914
31. S. A. Matthews *et al*., High-resolution characterization of the abiotic environment and disturbance regimes on the Great Barrier Reef, 1985-2017. *Ecology* **100**, e02574 (2019). doi: 10.1002/ecy.2574; pmid: 30645776
32. S. DiPerna, M. Hoogenboom, S. Noonan, K. Fabricius, Effects of variability in daily light integrals on the photophysiology of the corals *Pachyseris speciosa* and *Acropora millepora*. *PLOS ONE* **13**, e0203882 (2018). doi: 10.1371/journal.pone.0203882; pmid: 30240397
33. U. E. Siebeck, N. J. Marshall, A. Klüter, O. Hoegh-Guldberg, Monitoring coral bleaching using a colour reference card. *Coral Reefs* **25**, 453–460 (2006). doi: 10.1007/s00338-006-0123-8
34. Z. T. Richards, D. J. Miller, C. C. Wallace, Molecular phylogenetics of geographically restricted *Acropora* species: Implications for threatened species conservation. *Mol. Phylogenet. Evol.* **69**, 837–851 (2013). doi: 10.1016/j.ympev.2013.06.020; pmid: 23850500
35. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011). doi: 10.1038/nature10231; pmid: 21753753
36. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014). doi: 10.1038/ng.3015; pmid: 24952747
37. Y. Mao, E. P. Economo, N. Satoh, The roles of introgression and climate change in the rise to dominance of *Acropora* corals. *Curr. Biol.* **28**, 3373–3382.e5 (2018). doi: 10.1016/j.cub.2018.08.061; pmid: 30344117
38. C. Prada *et al*., Empty niches after extinctions increase population sizes of modern corals. *Curr. Biol.* **26**, 3190–3194 (2016). doi: 10.1016/j.cub.2016.09.039; pmid: 27866895
39. D. Petkova, J. Novembre, M. Stephens, Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* **48**, 94–100 (2016). doi: 10.1038/ng.3464; pmid: 26642242
40. P. Souter *et al*., Location and disturbance affect population genetic structure in four coral species of the genus *Acropora* on the Great Barrier Reef. *Mar. Ecol. Prog. Ser.* **416**, 35–45 (2010). doi: 10.3354/meps08740
41. S. W. Davies, E. A. Treml, C. D. Kenkel, M. V. Matz, Exploring the role of Micronesian islands in the maintenance of coral genetic diversity in the Pacific Ocean. *Mol. Ecol.* **24**, 70–82 (2015). doi: 10.1111/mec.13005; pmid: 25407355
42. W. Huang *et al*., Genetic diversity and large-scale connectivity of the scleractinian coral *Porites lutea* in the South China Sea. *Coral Reefs* **37**, 1259–1271 (2018). doi: 10.1007/s00338-018-1724-8
43. V. Lukoschek, C. Riginos, M. J. H. van Oppen, Congruent patterns of connectivity can inform management for broadcast spawning corals on the Great Barrier Reef. *Mol. Ecol.* **25**, 3065–3080 (2016). doi: 10.1111/mec.13649; pmid: 27085309
44. J. T. Ladner, S. R. Palumbi, Extensive sympatry, cryptic diversity and introgression throughout the geographic distribution of two coral species complexes. *Mol. Ecol.* **21**, 2224–2238 (2012). doi: 10.1111/j.1365-294X.2012.05528.x; pmid: 22439812
45. H. Levene, Genetic Equilibrium when more than one ecological niche is available. *Am. Nat.* **87**, 331–333 (1953). doi: 10.1086/281792
46. S. Yeaman, S. P. Otto, Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**, 2123–2129 (2011). doi: 10.1111/j.1558-5646.2011.01277.x; pmid: 21729066
47. N. R. Garud, P. W. Messer, E. O. Buzbas, D. A. Petrov, Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLOS Genet.* **11**, e1005004 (2015). doi: 10.1371/journal.pgen.1005004; pmid: 25706129
48. H. Li, P. Ralph, Local PCA shows how the effect of population structure differs along the genome. *Genetics* **211**, 289–304 (2019). doi: 10.1534/genetics.118.301747; pmid: 30459280
49. D. A. Parfitt *et al*., The ataxia protein sacsin is a functional co-chaperone that protects against polyglutamine-expanded ataxin-1. *Hum. Mol. Genet.* **18**, 1556–1565 (2009). doi: 10.1093/hmg/ddp067; pmid: 19208651
50. M. Ménade *et al*., Structures of ubiquitin-like (Ubl) and Hsp90-like domains of sacsin provide insight into pathological mutations. *J. Biol. Chem.* **293**, 12832–12842 (2018). doi: 10.1074/jbc.RA118.003939; pmid: 29945973
51. R. Cunning, R. A. Bay, P. Gillette, A. C. Baker, N. Traylor-Knowles, Comparative analysis of the *Pocillopora damicornis* genome highlights role of immune system in coral evolution. *Sci. Rep.* **8**, 16134 (2018). doi: 10.1038/s41598-018-34459-8; pmid: 30382153
52. E. M. Hemond, S. T. Kaluziak, S. V. Vollmer, The genetics of colony form and function in Caribbean *Acropora* corals. *BMC Genomics* **15**, 1133 (2014). doi: 10.1186/1471-2164-15-1133; pmid: 25519925
53. R. Cunning, A. C. Baker, Not just who, but how many: The importance of partner abundance in reef coral symbioses. *Front. Microbiol.* **5**, 400 (2014). doi: 10.3389/fmicb.2014.00400; pmid: 25136339
54. A. B. Mayfield, Y.-J. Chen, C.-Y. Lu, C.-S. Chen, The proteomic response of the reef coral *Pocillopora acuta* to experimentally elevated temperatures. *PLOS ONE* **13**, e0192001 (2018). doi: 10.1371/journal.pone.0192001; pmid: 29385204
55. O. Delaneau, B. Howie, A. J. Cox, J.-F. Zagury, J. Marchini, Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013). doi: 10.1016/j.ajhg.2013.09.002; pmid: 24094745
56. S. Picelli *et al*., Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014). doi: 10.1101/gr.177881.114; pmid: 25079858
57. L. Huang *et al*., Genotype-imputation accuracy across worldwide human populations. *Am. J. Hum. Genet.* **84**, 235–250 (2009). doi: 10.1016/j.ajhg.2009.01.013; pmid: 19215730
58. B. L. Browning, Y. Zhou, S. R. Browning, A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018). doi: 10.1016/j.ajhg.2018.07.015; pmid: 30100085
59. D. P. Manzello *et al*., Role of host genetics and heat-tolerant algal symbionts in sustaining populations of the endangered coral *Orbicella faveolata* in the Florida Keys with ocean warming. *Glob. Change Biol.* **25**, 1016–1031 (2019). doi: 10.1111/gcb.14545; pmid: 30552831
60. P. M. Visscher *et al*., Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLOS Genet.* **10**, e1004269 (2014). doi: 10.1371/journal.pgen.1004269; pmid: 24721987

61. P. M. Visscher *et al*., Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLOS Genet.* **2**, e41 (2006). doi: 10.1371/journal.pgen.0020041; pmid: 16565746
62. J. E. Parkinson, I. B. Baums, The extended phenotypes of marine symbioses: Ecological and evolutionary consequences of intraspecific genetic diversity in coral-algal associations. *Front. Microbiol.* **5**, 445 (2014). doi: 10.3389/fmicb.2014.00445; pmid: 25202306
63. T. F. Cooper *et al*., Environmental factors controlling the distribution of symbiodinium harboured by the coral *Acropora millepora* on the Great Barrier Reef. *PLOS ONE* **6**, e25536 (2011). doi: 10.1371/journal.pone.0025536; pmid: 22065989
64. R. Berkelmans, M. J. H. van Oppen, The role of zooxanthellae in the thermal tolerance of corals: A 'nugget of hope' for coral reefs in an era of climate change. *Proc. Biol. Sci.* **273**, 2305–2312 (2006). doi: 10.1098/rspb.2006.3567; pmid: 16928632
65. M. K. Morikawa, S. R. Palumbi, Using naturally occurring climate resilient corals to construct bleaching-resistant nurseries. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10586–10591 (2019). doi: 10.1073/pnas.1721415116; pmid: 31061118
66. A. Klueter, J. Trapani, F. I. Archer, S. E. McIlroy, M. A. Coffroth, Comparative growth rates of cultured marine dinoflagellates in the genus *Symbiodinium* and the effects of temperature and light. *PLOS ONE* **12**, e0187707 (2017). doi: 10.1371/journal.pone.0187707; pmid: 29186143
67. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012). doi: 10.1038/ng.2310; pmid: 22706312
68. X. Zhou, A unified framework for variance component estimation with summary statistics in genome-wide association studies. *Ann. Appl. Stat.* **11**, 2027–2051 (2017). doi: 10.1214/17-AOAS1052; pmid: 29515717
69. P. R. Muir, P. A. Marshall, A. Abdulla, J. D. Aguirre, Species identity and depth predict bleaching severity in reef-building corals: Shall the deep inherit the reef? *Proc. Biol. Sci.* **284**, 20171551 (2017). doi: 10.1098/rspb.2017.1551; pmid: 29021175
70. B. Servin, M. Stephens, Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLOS Genet.* **3**, e114 (2007). doi: 10.1371/journal.pgen.0030114; pmid: 17676998
71. F. A. Kondrashov, E. V. Koonin, I. G. Morgunov, T. V. Finogenova, M. N. Kondrashova, Evolution of glyoxylate cycle enzymes in Metazoa: Evidence of multiple horizontal transfer events and pseudogene formation. *Biol. Direct* **1**, 31 (2006). doi: 10.1186/1745-6150-1-31; pmid: 17059607
72. N. R. Polato, N. S. Altman, I. B. Baums, Variation in the transcriptional response of threatened coral larvae to elevated temperatures. *Mol. Ecol.* **22**, 1366–1382 (2013). doi: 10.1111/mec.12163; pmid: 23331636
73. R. M. Wright, G. V. Aglyamova, E. Meyer, M. V. Matz, Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *BMC Genomics* **16**, 371 (2015). doi: 10.1186/s12864-015-1540-2; pmid: 25956907
74. S. Purcell *et al*., PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
75. I. B. Baums, A restoration genetics guide for coral reef conservation. *Mol. Ecol.* **17**, 2796–2811 (2008). doi: 10.1111/j.1365-294X.2008.03787.x; pmid: 18482262
76. J. E. Parkinson *et al*., Molecular tools for coral reef restoration: Beyond biomarker discovery. *Conserv. Lett.* 10.31219/osf.io/9k7ah (2019).
77. M. J. H. van Oppen, J. K. Oliver, H. M. Putnam, R. D. Gates, Building coral reef resilience through assisted evolution. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2307–2313 (2015). doi: 10.1073/pnas.1422301112; pmid: 25646461
78. S. Koren *et al*., Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017). doi: 10.1101/gr.215087.116; pmid: 28298431
79. C.-S. Chin *et al*., Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016). doi: 10.1038/nmeth.4035; pmid: 27749838
80. S. Wang, L. Zhang, E. Meyer, M. V. Matz, Construction of a high-resolution genetic linkage map and comparative genome analysis for the reef-building coral *Acropora millepora*. *Genome Biol.* **10**, R126 (2009). doi: 10.1186/gb-2009-10-11-r126; pmid: 19900279
81. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). doi: 10.1093/bioinformatics/btu170; pmid: 24695404
82. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010). doi: 10.1093/bioinformatics/btp698; pmid: 20080505
83. A. McKenna *et al*., The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi: 10.1101/gr.107524.110; pmid: 20644199
84. N. Li, M. Stephens, Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003). pmid: 14704198
85. P. Danecek *et al*., The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011). doi: 10.1093/bioinformatics/btr330; pmid: 21653522
86. M. Nei, W. H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5269–5273 (1979). doi: 10.1073/pnas.76.10.5269; pmid: 291943
87. C. Shinzato *et al*., Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320–323 (2011). doi: 10.1038/nature10249; pmid: 21785439
88. J. Felsenstein, G. A. Churchill, A Hidden Markov Model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* **13**, 93–104 (1996). doi: 10.1093/oxfordjournals.molbev.a025575; pmid: 8583911
89. Z. Fuller, zfuller5280/CoralGenomes: Public release. Zenodo (2020). doi: 10.5281/zenodo.3727666

ACKNOWLEDGMENTS

We thank members of the Australian Institute of Marine Science's coral bleaching group for assistance with sample collections and R. Durbin, A. Harpak, S. Myers, J. Pritchard, M. Schumer, and the Andolfatto, Przeworski, and Sella laboratories for helpful comments or discussions. We are particularly grateful to J. Sorkness from the Austin Reef Club for providing a live *A. millepora* fragment as a source of high-quality DNA for genome sequencing. We are also thankful to K. Dougan, A. Bellantuono, C. Granados-Cifuentes, and M. Rodriguez-Lanetty for making the *Durusdinium* draft genome sequence available. **Funding:** This project was funded by the Australian Institute of Marine Science, the Australian government's National Environmental Science Program, the Agouron Institute, and Columbia University. L.A.M. was supported by an AIMS@JCU Ph.D. scholarship and a National Environmental Science Program Tropical Water Quality Hub grant. M.M. was supported by National Science Foundation grant IOS1755277. We acknowledge computing resources from Columbia University's Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) contract C090171. The coral collections were conducted under Great Barrier Reef Marine Park Authority collection permit G16/38488.1. **Author contributions:** Z.L.F., L.K.B., M.M., and M.P. designed the project; Z.L.F. and M.P. drafted the manuscript, alongside L.K.B. and M.M.; V.J.L.M., L.A.M., N.C., M.M., and L.K.B. collected samples and performed phenotyping; Z.L.F., Y.L., J.S., and L.S. performed analyses; P.A. and J.P. provided expertise and technical support; J.P. and L.S. conducted the sequencing validation; Z.L.F., L.K.B., and M.P. funded the research; and L.K.B., M.M., and M.P. jointly supervised the work. **Competing interests:** The authors declare no competing interests. **Data availability:** The reference genome assembly is available under BioProject PRJNA633778. All paired-end reads are deposited on the Sequence Read Archive under accession nos. SAMN13447100-13447355 and under BioProject PRJNA593014. The phased reference haplotype panel and imputation pipeline are freely available on www.gencove.com. All scripts and associated data are available through Zenodo (*89*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6501/eaba4674/suppl/DC1
Materials and Methods
Figs. S1 to S37
Tables S1 to S8
References (*90*–*133*)

View/request a protocol for this paper from *Bio-protocol*.

5 December 2019; accepted 1 June 2020
10.1126/science.aba4674

Science

Population genetics of the coral *Acropora millepora*: Toward genomic prediction of bleaching

Zachary L. Fuller, Veronique J. L. Mocellin, Luke A. Morris, Neal Cantin, Jihanne Shepherd, Luke Sarre, Julie Peng, Yi Liao, Joseph Pickrell, Peter Andolfatto, Mikhail Matz, Line K. Bay and Molly Przeworski

Science **369** (6501), eaba4674.
DOI: 10.1126/science.aba4674

Conservation help from genomics

Corals worldwide are under threat from rising sea temperatures and pollution. One response to heat stress is coral bleaching—the loss of photosynthetic endosymbionts that provide energy for the coral. Fuller *et al.* present a high-resolution genome of the coral *Acropora millepora* (see the Perspective by Bay and Guerrero). They were able to perform population genetic analyses with samples sequenced at lower coverage and conduct genome-wide association studies. These data were combined to generate a polygenic risk score for bleaching that can be used in coral conservation.

Science this issue p. eaba4674; see also p. 249

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.