Article

Current Biology

Morphological stasis masks ecologically divergent coral species on tropical reefs

Highlights

- Widespread Indo-Pacific coral represents a cryptic species complex
- Substantial ecological differentiation despite ancient morphological stasis
- Current taxonomic framework for corals underestimates ecologically relevant diversity

Authors

Pim Bongaerts, Ira R. Cooke, Hua Ying, ..., Mark A. Ragan, Madeleine J.H. van Oppen, Ove Hoegh-Guldberg

Correspondence

pbongaerts@calacademy.org

In brief

Extensive cryptic diversity is being uncovered in reef-building corals. Bongaerts et al. demonstrate through a comprehensive genomic assessment of a widespread coral that such cryptic diversity can mask ecologically and physiologically divergent species and highlight the critical need to acknowledge and start capturing this hidden diversity.

Bongaerts et al., 2021, Current Biology *31*, 2286–2298
June 7, 2021
https://doi.org/10.1016/j.cub.2021.03.028

CellPress

CellPress
OPEN ACCESS

Current Biology

Article

Morphological stasis masks ecologically divergent coral species on tropical reefs

Pim Bongaerts,[1,2,3,20,21,*] Ira R. Cooke,[4,5,18] Hua Ying,[6,18] Dagmar Wels,[2,18] Stijn den Haan,[2,7,18] Alejandra Hernandez-Agreda,[1] Christopher A. Brunner,[2,8,9] Sophie Dove,[2,3] Norbert Englebert,[2,3] Gal Eyal,[3,10] Sylvain Forêt,[6] Mila Grinblat,[4,5,8] Kyra B. Hay,[2] Saki Harii,[11] David C. Hayward,[6] Yu Lin,[12] Morana Mihaljević,[2,13] Aurelie Moya,[8,14] Paul Muir,[15] Frederic Sinniger,[11] Patrick Smallhorn-West,[8] Gergely Torda,[8] Mark A. Ragan,[16,19] Madeleine J.H. van Oppen,[9,17,19] and Ove Hoegh-Guldberg[2,3,19]

[1]California Academy of Sciences, San Francisco, CA 94118, USA
[2]Global Change Institute, The University of Queensland, Brisbane, QLD 4072, Australia
[3]ARC Centre of Excellence for Coral Reef Studies, School of Biological Sciences, The University of Queensland, Brisbane, QLD 4072, Australia
[4]Department of Molecular and Cell Biology, James Cook University, Townsville, QLD 4811, Australia
[5]Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, Townsville, QLD 4811, Australia
[6]Division of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, ACT 2601, Australia
[7]Department of Environmental Sciences and Policy, Central European University, Budapest 1051, Hungary
[8]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia
[9]Australian Institute of Marine Science, Townsville, QLD 4810, Australia
[10]The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel
[11]Tropical Biosphere Research Center, University of the Ryukyus, Motobu, Okinawa 905-0227, Japan
[12]Research School of Computer Science, Australian National University, Canberra, ACT 2601, Australia
[13]Science Lab UZH, University of Zurich, Zurich 8057, Switzerland
[14]Department of Biology, University of Konstanz, Konstanz 78457, Germany
[15]Queensland Museum, Townsville, QLD 4810, Australia
[16]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia
[17]School of BioSciences, The University of Melbourne, Parkville, VIC 3010, Australia
[18]These authors contributed equally
[19]Senior author
[20]Twitter: @pimbongaerts
[21]Lead contact
*Correspondence: pbongaerts@calacademy.org
https://doi.org/10.1016/j.cub.2021.03.028

SUMMARY

Coral reefs are the epitome of species diversity, yet the number of described scleractinian coral species, the framework-builders of coral reefs, remains moderate by comparison. DNA sequencing studies are rapidly challenging this notion by exposing a wealth of undescribed diversity, but the evolutionary and ecological significance of this diversity remains largely unclear. Here, we present an annotated genome for one of the most ubiquitous corals in the Indo-Pacific (*Pachyseris speciosa*) and uncover, through a comprehensive genomic and phenotypic assessment, that it comprises morphologically indistinguishable but ecologically divergent lineages. Demographic modeling based on whole-genome resequencing indicated that morphological crypsis (across micro- and macromorphological traits) was due to ancient morphological stasis rather than recent divergence. Although the lineages occur sympatrically across shallow and mesophotic habitats, extensive genotyping using a rapid molecular assay revealed differentiation of their ecological distributions. Leveraging "common garden" conditions facilitated by the overlapping distributions, we assessed physiological and quantitative skeletal traits and demonstrated concurrent phenotypic differentiation. Lastly, spawning observations of genotyped colonies highlighted the potential role of temporal reproductive isolation in the limited admixture, with consistent genomic signatures in genes related to morphogenesis and reproduction. Overall, our findings demonstrate the presence of ecologically and phenotypically divergent coral species without substantial morphological differentiation and provide new leads into the potential mechanisms facilitating such divergence. More broadly, they indicate that our current taxonomic framework for reef-building corals may be scratching the surface of the ecologically relevant diversity on coral reefs, consequently limiting our ability to protect or restore this diversity effectively.

INTRODUCTION

Tropical coral reefs are known for their high levels of biodiversity, harboring hundreds of thousands of macroscopic and an unknown number of microscopic species.[1] Interestingly, the contribution of reef-building corals remains surprisingly moderate, with only ~750–850 valid species accepted worldwide.[2–5] Under the current taxonomic framework, species are distinguished primarily based on diagnostic skeletal characteristics, an approach that originated from the early days of coral reef science, when underwater observations were extremely challenging.[6,7] Skeletal traits at both the corallite and colony level are often highly variable within species and across environments, posing a major challenge to coral taxonomy.[8] Conversely, skeletal traits are known to converge and can even obscure deep phylogenetic divergences at the family level,[9,10] highlighting further challenges to the use of the skeletal morphospace as a taxonomic framework. Molecular approaches have helped resolve some of these difficulties by differentiating morphological plasticity from actual species traits,[11,12] confirming species separation in the context of subtle morphological differences,[13–15] and clarifying deeper evolutionary relationships within the order.[9,10] However, molecular studies have also uncovered a wealth of undescribed diversity within taxonomic species that cannot be readily explained.[11,16–24]

The notion that scleractinian coral diversity may be far greater than acknowledged through conventional taxonomy is not novel. A seminal review published over 2 decades ago highlighted the ubiquitous presence of sibling species in the marine realm and stressed the importance of exploring its ecological relevance.[25] Since then, molecular studies have indeed exposed "taxonomically cryptic diversity" (i.e., genetically distinct taxa that have been erroneously classified under a single species name) within many coral genera.[21,26] However, much of this cryptic diversity is being exposed through population genetic studies designed to relate genetic patterns to geography,[26] and consequently, it usually remains unclear to what extent they represent phenotypically and ecologically distinct entities (but see exceptions[27–31]), including whether lineages are truly morphologically cryptic. In fact, well-studied examples of coral species complexes are characterized by substantial morphological differentiation,[14,16,17,20,27] and we still know little about the potential for ecological or phenotypic differentiation when gross morphology is largely constrained. The traditional use of only a few genetic loci has further impeded an assessment of the evolutionary context of cryptic diversification and the respective roles of neutral versus selective processes on ecological or reproductive traits.[28,32] Overall, the ecological relevance of cryptic diversity in corals remains poorly understood and is rarely considered in ecological studies or conservation planning.

To address this knowledge gap, we conducted a comprehensive assessment to evaluate the nature of cryptic diversity in *Pachyseris speciosa* ("serpent coral"; Figure 1A), one of the most ubiquitous and abundant species across the Indo-Pacific.[3] This zooxanthellate coral has one of the widest bathymetric distributions on tropical coral reefs—from close to the surface to lower mesophotic depths (~5–95 m)[33]—and has therefore been the focus of studies assessing its ecological opportunism.[34–37] Our initial assessment of its genetic structure indicated the presence of undescribed sympatric lineages, and we therefore used this as an opportunity to explore the extent to which cryptic diversity can obscure genomic, ecological, and phenotypic divergence and how reproductive isolation may be maintained in such closely related lineages. Specifically, our study (summarized in Figure 1) involved the generation of an annotated genome for *P. speciosa* using long-read sequencing, which we used as a reference for reduced-representation sequencing (nextRAD) of populations ranging from across the geographic and bathymetric distribution of this species. Subsequent whole-genome resequencing (WGS) of representative coral colonies was implemented for demographic modeling and for the development of a cleaved amplified polymorphic sequence (CAPS) assay to assess ecological, phenotypic, and reproductive differentiation in the uncovered lineages from Australia. Overall, our findings demonstrate the potential for ecological and phenotypic divergence without substantial morphological differentiation and highlight that conventional skeleton-based taxonomy may be substantially underestimating the ecologically relevant diversity of scleractinian corals on tropical coral reefs.

RESULTS AND DISCUSSION

Genome assembly and annotation

We generated a highly contiguous reference genome for *Pachyseris speciosa*, assembled through PacBio single-molecule long-read sequencing (Figure 1D; Table S2). The assembled genome size is 984 Mb, representing one of the largest coral genomes reported to date and comprising 2,368 contigs with a N50 size of 766.6 kb (largest contig 4.6 Mb; Table S2). In total, 39,160 protein-coding genes were predicted—a number comparable to that of two other recently sequenced corals from the "robust" clade (i.e., one of the two major clades of extant corals)[39,40]—with the gene models being 90.5% complete and 5.1% partial (based on conserved single-copy metazoan orthologous genes; Table S2). *De novo* repeat annotation revealed that 52.2% of the assembly is occupied by repetitive elements, which are generally better resolved in long-read assemblies.[40] The dominance of transposable elements compared with other robust corals indicates that the larger genome size may be substantially driven by expansion of those transposons (Table S2).[41]

Pachyseris speciosa represents a sympatric species complex

The annotated genome assembly was used as a reference for nextRAD sequencing, where we targeted *P. speciosa* colonies (n = 501) from 32 sites in Australasia (the Great Barrier Reef [GBR], the Western Coral Sea [WCS], and Papua New Guinea [PNG]), Okinawa (OKI), and Israel (ISR) (Figures 1E and S1; Table S1). Genetic structuring based on principal-component and neighbor-joining analyses revealed the presence of four highly divergent lineages (Figures 2 and S1C). One lineage represents the geographically separated ISR population (Gulf of Aqaba) and indicates the existence of a distinct species in the Red Sea region, consistent with patterns observed in other coral genera.[24,42,43] The other three lineages occurred sympatrically at each of the sampled reefs in Australasia, hereafter arbitrarily referred to as "green," "blue," and "red" lineage, with the

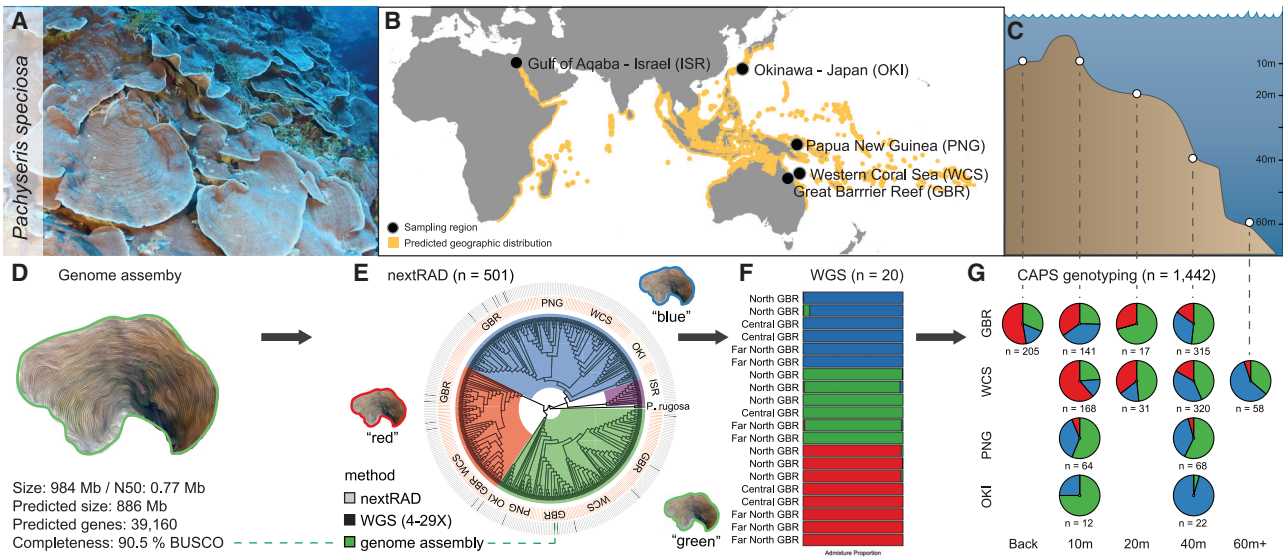

Figure 1. Overview of the study system and sequencing/genotyping approach

(A) *Pachyseris speciosa* colonies at 40 m depth (Western Coral Sea).

(B) Predicted geographic distribution[38] of *Pachyseris speciosa* sensu lato and the five sampling regions.

(C) Sampled habitats and their corresponding water depths (±3 m).

(D) Genome assembly summary statistics.

(E) Circular phylogenetic tree summarizing the reduced-representation sequencing data (nextRAD; 8,536 SNPs) and showing the partitioning into four major lineages. Sample region is indicated in the inner surrounding circle and sample selection for whole-genome sequencing is indicated in the outer surrounding circle.

(F) Admixture proportion (PCAngsd) based on whole-genome resequencing (WGS; ~37M SNPs) of representative samples from the three lineages occurring on the Great Barrier Reef.

(G) Geographic and ecological distribution of the lineages across habitats and regions (excluding Israel) based on the CAPS genotyping assay (1–3 markers per sample) merged with the nextRAD data (note that assignments for Okinawa are based exclusively on nextRAD data).

assembled reference genome representing a genotype belonging to the green lineage. Subsequent structure analyses using a maximum-likelihood framework (snapclust) indicated an optimum between four and six clusters, and the additional two clusters (for k = 5 and 6) correspond to two populations from OKI that grouped with, respectively, the blue and green lineages but were substantially differentiated from those in eastern Australia (Figures S1D and S1E). Beyond this expected geographic differentiation of sub-tropical OKI, the overall structuring demonstrates the presence of genetically distinct lineages within the currently acknowledged species, *P. speciosa*.

Phylogenomic analyses demonstrate that the *P. speciosa* lineages represent a closely related species complex, compared to congeners *Pachyseris rugosa* and *Pachyserisinattesa* (Figures 3A, S2A, and S2B). Differentiation between the red, green, and blue lineages is observed across the genome, and global mean F_{ST} values are much higher (average = 0.1256; range: 0.1032–0.1613) than between geographic regions within these lineages (average = 0.0207; range: 0.0130–0.0339; Figures 2 and S1E). The two lineages in OKI are more related to the green and blue lineages, respectively, from Australia than to each other (global mean F_{ST} = 0.1324; Figure S1E) and consistently grouped together with those lineages in both concatenated and species tree phylogenies (Figures 3A, S2A, and S2B). The "purple" lineage exclusive to ISR (Red Sea) is, on average, the most divergent of all the lineages and was consistently placed as ancestral to the blue and "dark blue" lineages phylogenies. Deeper phylogenetic relationships between the lineages are less clear, with low bootstrap support for the short backbone internodes of the three major clades (red, "green/dark green," and "purple/blue/dark blue"; Figures 3, S2A, and S2B). The apparent restricted distribution of the red lineage in Australasia indicates that it may have originated in that region, and similarly, the consistent ancestral placement of the purple lineage may indicate a potential origin of that clade ("blue/purple") in the Indian Ocean.

Indications of admixture (i.e., samples with a maximum assignment <0.95) between the three lineages were minimal with the exception of one F1 hybrid on the GBR (Figures 2 and S1F). This coral had a roughly equal assignment to the green and blue cluster and a heterozygous genotype for 35 out of 39 genotyped SNPs that were alternatively fixed between corals belonging to these two clusters. Admixture signatures of most other putatively admixed samples were indicative of analytical artifacts (e.g., due to admixture with an unsampled population/lineage; Figure S1F). Overall, the limited admixture and presence of only a single F1 hybrid (with its reproductive viability being unknown), indicates that the three lineages have evolved reproductive isolation. This, combined with their widespread sympatric distribution, provides further support that these lineages represent distinct species (with their formal description underway; P.M. and P.B., unpublished data), adding to the ever-growing number of (taxonomically) cryptic species within the Scleractinia.[21,26]

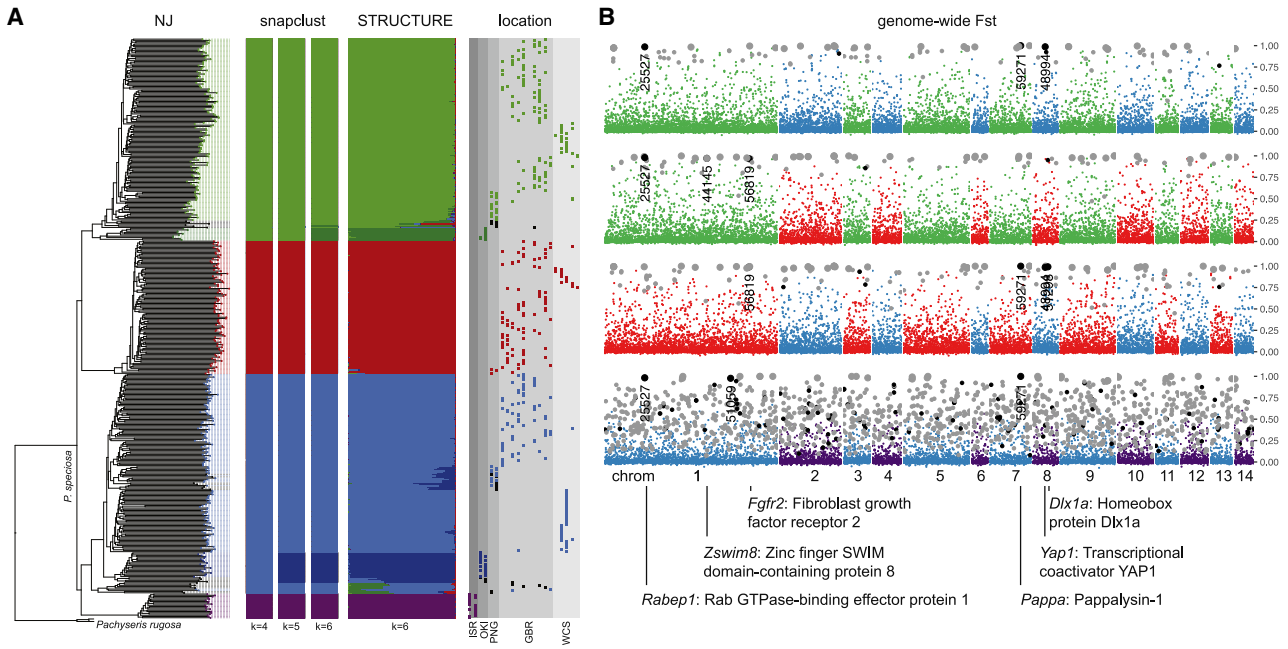

Figure 2. Genetic structuring and genome-wide differentiation of *P. speciosa* lineages based on nextRAD sequencing

(A) Neighbor-joining tree with matching snapclust (k = 4–6) and STRUCTURE (k = 6) assignments for each coral colony (nextRAD; 1,861 distance-filtered SNPs). Dots in the "location" column indicate sampling population (geographically arranged within region) with the color indicating STRUCTURE assignment (potentially admixed samples in black). The *Pachyseris rugosa* outgroup (n = 3) is not included in the clustering analyses.

(B) Manhattan plots showing genome-wide distribution of F_{ST} values for different *P. speciosa* lineage comparisons (nextRAD; showing 13,760 SNPs). SNPs are plotted along pseudochromosomes (mapped to *Acropora millepora* for visualization purposes), showing only those that mapped and with the alternating colors indicating the two lineages compared in each plot. Gray and black dots indicate SNPs that are alternatively fixed (large dot) or identified as pcadapt outliers (medium-sized dot), with a black color indicating missense variants. Alternatively, fixed missense variants relating to the inter-comparison of the three Australasian lineages are additionally labeled with their gene ID and the corresponding UniProt gene and protein names (note that two variants, identified as located in *Hecw2* and an unidentified gene, are not mentioned here, as they did not map to *A. millepora* chromosomes).

See also Figure S1.

Morphologically cryptic lineages despite ancient divergence

Gross morphological variation in *P. speciosa* has been characterized as highly constrained, compared to the much more variable congener *P. rugosa*.[2,45] This lack of macromorphological variation was confirmed in an assessment of *in situ* photographs of sampled colonies (n = 157) from the GBR and WCS. We did observe variability in the height, continuity, and symmetry of the carinae (i.e., the characteristic "ridges" of this species); however, this variation did not partition across the three different lineages (Figure S3A). Similarly, an examination of qualitative traits using skeletal specimens exposed variation in some of the traits (mostly related to the carinae), but again, this was not partitioned across the three different lineages (n = 36; Table S3), nor did it align with one of the other five currently accepted species described in this genus.[5] An initial assessment of micro-skeletal features using scanning electron microscopy also did not reveal differentiating characteristics between the three lineages (n = 15; Figure S3B).

In contrast to the lineages being morphologically indistinguishable, they are divergent genomically with 146 alternatively fixed SNPs among the three lineages in Australia (0.5% of 29,287 SNPs based on nextRAD sequencing; Figure 2B). This included 8 missense variants (i.e., a nonsynonymous substitution producing a different amino acid) located in genes of which 7 had homology to UniProtDB genes (Figure 2B). This included the fibroblast growth factor receptor 2 gene (*Fgfr2*) thought to be involved in environmental sensing and larval metamorphosis,[46–49] the *Dlx1a* gene from the homeobox-containing superfamily generally associated with morphogenesis,[50] and the *Pappa* gene encoding the pregnancy-associated plasma protein-A (which is upregulated during spawning in *Orbicella franksi* and *Orbicella annularis*).[51] Outlier analyses using pcadapt identified a total of 459 SNPs, which included 19 additional missense variants. This included genes associated with the circadian clock (*Hcrtr2*), skeletal matrix (CADN_ACRMI, *Fat1*), and the thermal stress response (*Sacs*).[52–54] When comparing the three lineages with the purple lineage from ISR, we identified alternatively fixed missense variants in 16 additional genes. Several of these (*Sacs*, *Pxdn*, *Ipo9*, and *Park2*) belonged to gene families known to be involved in the coral heat stress response,[54–57] potentially related to the thermal resilience generally associated with corals from this region.[58]

Whole-genome resequencing of representative green, blue, and red colonies from the GBR (n = 20 at 4–27× coverage) further confirmed the strong genetic differentiation with minimal admixture (Figure 1F). To assess whether the three lineages have diverged recently and/or experienced differences in demographic history, we used the pairwise sequentially Markovian coalescent (PSMC)[59] approach for a subset of colonies that

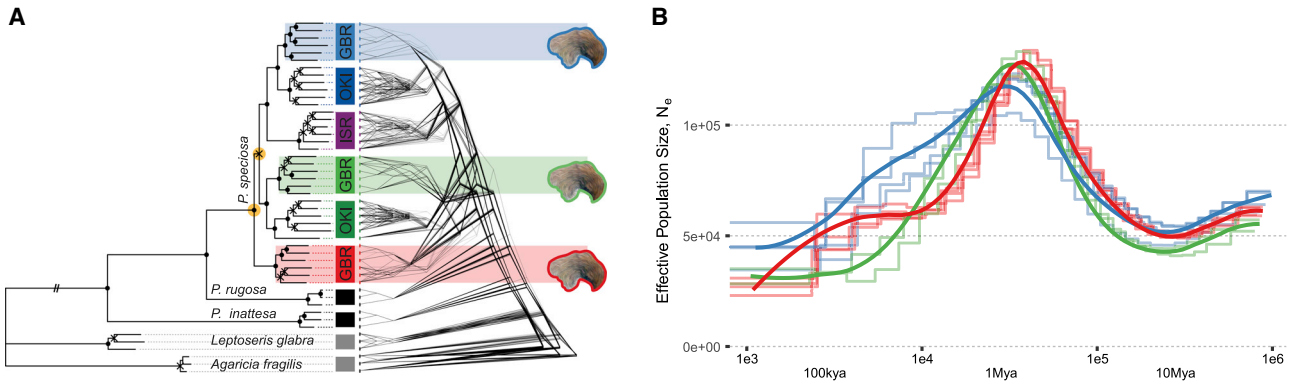

Figure 3. Phylogenetic relationships and demographic history of *P. speciosa* lineages

(A) Maximum-likelihood consensus tree (raxml) based on concatenated loci (nextRAD; ~1,7M distinct alignment patterns), with individual bootstraps (n = 400) shown mirrored on the right. Dots on nodes indicate high bootstrap support (>95%), with crosses indicating a lower bootstrap support. Yellow circles highlight uncertain backbone internodes (based on raxml and tetrad phylogenies).

(B) Demographic history as inferred through whole-genome resequencing. Changes in inferred, effective population size as estimated using MSMC for 10 deeply sequenced coral colonies (WGS; ~0.8–1M SNPs) representing the three *P. speciosa* lineages occurring sympatrically in Australia. Smoothed lines in bold represent averages for each of the three lineages. The timescale is shown in units of numbers of generations in the past, with the additional x axis showing time in the past, assuming a generation time of 35 years. A mutation rate of $4.83e^{-8}$ was assumed for all calculations.[44]

See also Figure S2.

were sequenced at greater depth (14–27× coverage; n = 10). Modeled demographic histories showed much greater variation between than within lineages and fail to converge, suggesting an ancient divergence as far back as 10 mya (Figure 3B) and that the cryptic nature is due to morphological stasis rather than recent divergence. All lineages peaked in effective population size around 1–3 mya, which is consistent with PSMC analyses in recent studies on other robust[44] and complex[60] corals. Given an ancient divergence, the short backbone internode observed in phylogenetic analyses (Figures 3 and S2) may be indicative of an "ancient rapid radiation,"[61] with the observed geographic patterns highlighting the potential role of vicariance events associated with oceanographic barriers across the Indian Ocean and the Indo-Australian Archipelago during the Late Miocene/Pliocene.[62] We also constructed circularized mitochondrial genomes from the whole-genome resequencing data (~19 kbp in length), which in contrast only showed a limited assortment of mitochondrial haplotypes into lineage-specific groups (Figure S2D), potentially reflecting incomplete lineage sorting and/or ancient introgression, combined with very low mutation rates typical of anthozoans.[63,64] It also reiterates how mitochondrial regions in corals for species-level phylogenetics should be used cautiously.[40]

Ecological and phenotypic differentiation of the lineages

Given the lack of diagnostic morphological features discriminating the three *P. speciosa* lineages, we designed a CAPS assay based on the nextRAD data to allow for high-throughput or field-based genotyping. We used the rapid assay to expand the number of identified genotypes (n = 1,442) to include a broad range of reef habitats and locations and assessed whether there are other ecological and/or phenotypic traits that distinguish the three lineages. The genotyping confirmed the ubiquitous sympatric distribution, with 39 out of 45 sampled Australasian sites containing representatives of all three lineages (Figure 4A). However, there was a significant overall effect of habitat and region on the relative abundances of the three lineages, with pairwise tests indicating significant differences between shallow (back-reef and 10 m) and mesophotic (40 m and 60 m) habitats (confined to GBR and WCS; Table S4). When assessing the proportions of individual lineages over depth, the red lineage was significantly more abundant on shallow (back reef and 10 m) as compared to intermediate (20 m) and mesophotic (40 and 60 m) habitats, with the opposite pattern observed for the green and blue lineages (Figure 4B; Table S4). Although the red lineage was not found in OKI, the divergent green and blue lineages appeared to be associated with, respectively, shallow and mesophotic depths (Figures 1G and 4A), indicating there may also be distinct ecological distributions in that geographic region.

To examine potential phenotypic differentiation, we quantitatively assessed skeletal and physiological differences between lineages in Australia. Principal component analyses based on coral host skeletal traits (four traits; n = 54), physiological traits (protein and lipid content; n = 69), and Symbiodiniaceae physiological traits (cell density and five pigment traits; n = 70) revealed separation between lineages despite high levels of variance (Figures 4D–4F). The green and red lineages differed significantly across all three trait groups, whereas the blue and red lineages differed in skeletal traits and blue and green in symbiont physiological traits (Table S4). In terms of skeletal traits, the mean density of septa (i.e., vertical blades inside the corallite) was significantly lower in the red lineage, but not morphologically diagnostic, given the overlapping ranges (Figure S4). Observed ranges of physiological traits were similar to those reported in other studies that include *P. speciosa*,[34,35] but separating the three lineages revealed differences in protein content, Symbiodiniaceae density, and photosynthetic pigment concentrations (Figure S4). Although *Pachyseris speciosa* sensu lato is usually restricted to shaded or deeper environments, it has been considered an efficient autotroph across its entire depth range.[36]

CellPress
OPEN ACCESS

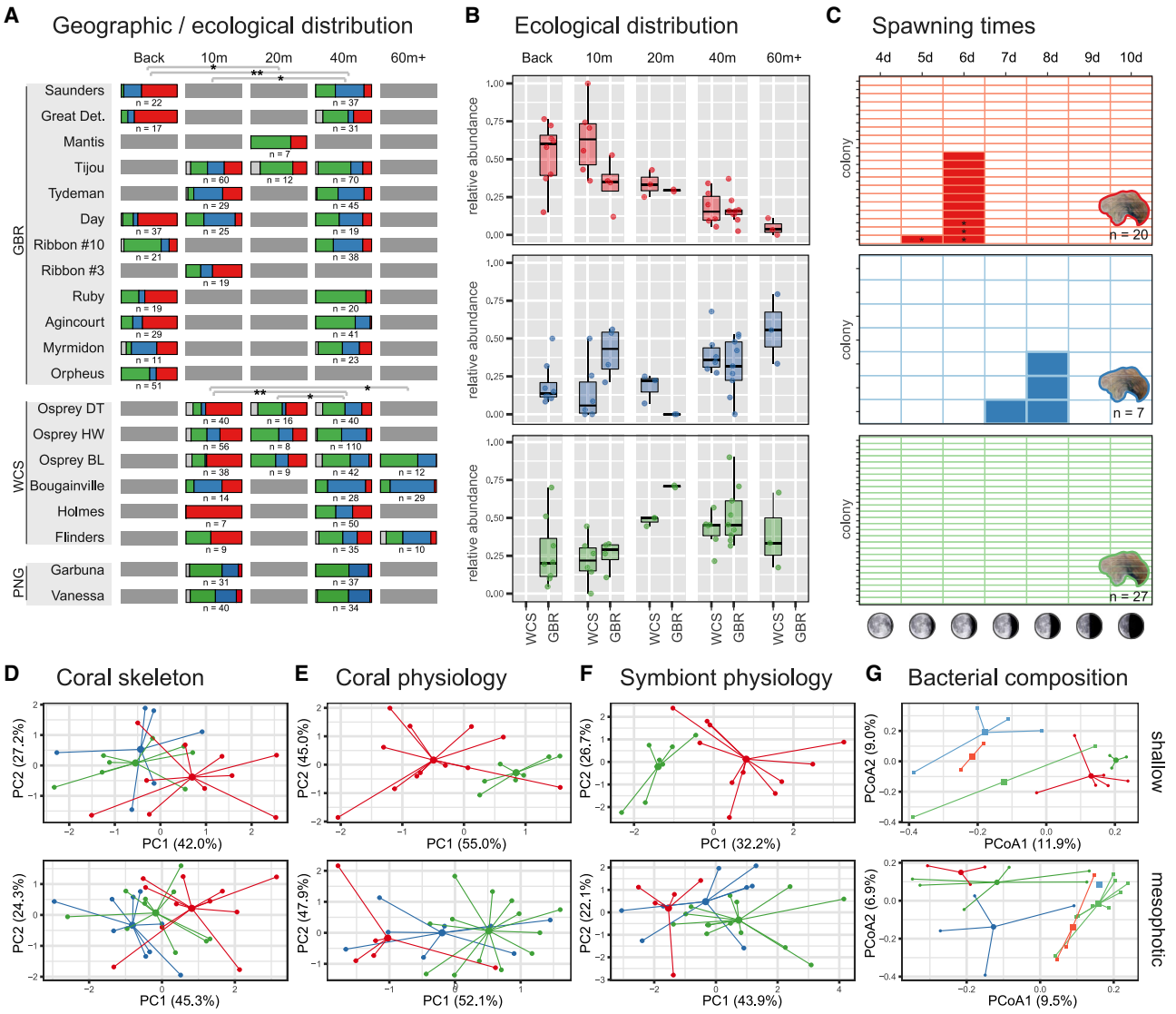

Figure 4. Ecological, phenotypic, and reproductive characterization of *P. speciosa* lineages

(A) Proportional composition of *P. speciosa* lineages across geographic locations and depth/habitat (based on CAPS and nextRAD genotyping). Gray background indicates that no sampling was conducted. Gray lines indicate significant pairwise differences (after Bonferroni correction), where *$p < 0.05$ and **$p < 0.01$.

(B) Boxplots summarizing relative abundances of *P. speciosa* lineages for each habitat and region (Great Barrier Reef and Western Coral Sea). Dots represent the proportions for each individual population (n = 41; corresponding to those in A).

(C) Binary heatmap indicating the spawning status of 54 *P. speciosa* colonies (rows) over time (columns) and grouped by lineage. Colonies were collected at Orpheus Island (Great Barrier Reef) and monitored *ex situ* between 3 and 9 days after the full moon in November 2017. Opaque cells indicate the release of eggs (asterisk) or sperm (no symbol). Empty cells indicate that the colony was isolated and monitored but that no gametes were observed (e.g., no gamete release was observed for the green lineage).

(D) Principal component analysis (PCA) for four host skeletal traits.

(E) PCA for two coral host physiological traits: protein and lipid content.

(F) PCA for six Symbiodiniaceae photophysiological traits.

(G) Principal coordinate analysis for bacterial community structure (based on 16S amplicon sequencing). Coral colonies are compared within their own depth groups: shallow (top plots; 10–20 m) and mesophotic (bottom plots; 40–60 m), with most colonies originating from either 10 or 40 m. Skeletal and physiological traits were measured for samples from the Western Coral Sea, whereas the bacterial composition included samples from both the Great Barrier Reef (square shapes) and Western Coral Sea (circle shapes) regions.

See also Figure S4 and Tables S4 and S5.

However, we observed a tendency for different lineages to vary in how they trade symbiont densities for chlorophyll concentration per symbiont cell, where lineages with thicker tissues (greater protein concentrations) hosted greater symbiont densities (Figure S4). Although sample sizes were small relative to the observed variance, the results demonstrate the existence

of phenotypic differences between the morphologically cryptic lineages, pointing toward distinct physiological strategies.

To explore whether the phenotypic differences may be attributed to differences in coral-associated bacteria or algal symbionts (Symbiodiniaceae), these microbial communities were compared among the three lineages. Associated bacterial communities were assessed by genotyping the host lineages from samples of a previously published 16S rRNA gene metabarcoding dataset (n = 43).[37] No significant differences were found in the bacterial community structure, richness, and diversity of the three lineages (Figure S2E), but regional patterns were present (WCS and GBR; Figure 4G; Table S4), confirming earlier findings.[37] Three bacterial operational taxonomic units (OTUs) from the genera *Corynebacterium* and *Gluconacetobacter* were consistently found in the three lineages, across regions. Despite the lack of differences in the overall bacterial communities, there were two unique OTUs that were only found in the blue and red lineage, respectively. To investigate potential differences in lineage-associated Symbiodiniaceae communities, we screened the nextRAD data for contaminating chloroplast or mitochondrial Symbiodiniaceae loci. We found three organellar loci that were genotyped for several hundred samples, but these were largely invariant within Australian samples. This is in line with a previous study (from West Australia) that observed a single Symbiodiniaceae type in *P. speciosa* over depth,[34] although some depth partitioning was observed in the Red Sea.[35] More-detailed mitochondrial characterization was conducted by aligning whole-genome resequencing data from the three lineages to the *Cladocopium goreaui* genome (ITS2 type C1);[65] this revealed a highly reticulated haplotype network based on an 8-kb mitochondrial region with 99.7% similarity across samples (n = 16). Out of the eleven haplotypes, two were shared between the green and blue lineages, whereas the remaining haplotypes were found only once or multiple times but in a single lineage (Figure S2C). Overall, microbial associations appeared to be either primarily driven by environment (in the case of the bacterial communities) or were highly consistent (Symbiodiniaceae), indicating that observed phenotypic differences between lineages were unlikely driven by distinct microbial associations.

Reproductive observations

Given the limited admixture between the sympatrically occurring *P. speciosa* lineages, we monitored the reproductive behavior of shallow colonies from Orpheus Island on the GBR. Previous reports from that location indicated spawning of *P. speciosa* between 5 and 6 days after the full moon.[66] Colonies were collected from the field just after the full moon in November 2017, genotyped using the CAPS assay (which identified 20 red, 7 blue, and 27 green colonies), and monitored *ex situ* for spawning from 3 to 9 days after the full moon. On day 5, half of the colonies of the red lineage released gametes (with one colony releasing also on day 4), with no colonies from the other two lineages releasing gametes (Figure 4C). On day 7, nearly half of the colonies from the blue lineage released gametes (with one also releasing on day 6; Figure 4C), with again no colonies from the other two lineages releasing gametes. No gamete release was observed for the green lineage within the monitoring period, although the green colony from which the draft genome was constructed spawned 8 days after the full moon in December 2014 (STAR Methods). Overall, the average spawning time of male colonies was 19 min after sunset (n = 13) versus an average of 43 min for female colonies (n = 4; Table S5), in line with general observations of males spawning earlier in broadcast spawning corals.[67] Attempts to preserve unfertilized eggs in filtered seawater (at ambient temperature) for experimental inter-lineages crosses were not successful (complete degradation within 24 h).

The observed temporal segregation in the timing of gamete release observed during the November 2017 spawning provides a potential mechanism to explain the minimal admixture between these sympatrically occurring lineages. Temporal reproductive isolation is a common strategy for minimizing interspecific gamete encounters in scleractinian corals, with the timing difference ranging from hours[68,69] to months.[70,71] Rather than diurnal or seasonal cycles,[72–74] the 2-day difference in gamete release observed here (between colonies from the red and blue lineages) could be due to entrainment by the lunar cycle. Although the role of temporal reproductive isolation will have to be confirmed through further work, the identification of fixed high-impact gene variants related to environmental sensing, development, and gametogenesis provides initial leads to the genomic basis of this potential prezygotic reproductive barrier.

Given the predicted split-spawning (i.e., mass spawning occurring across at least 2 consecutive months) in 2017, we maintained a subset of the colonies (n = 36) for additional spawning monitoring after the full moon of December in that year. Observations undertaken for those colonies saw gamete release for only one colony of the blue lineage and one of the green lineages (both on day 4 after the full moon). In total, two-thirds of the colonies from the red lineage released gametes, with release observed in the period between day 3 and day 8, with most of the release occurring on day 4 (8 colonies versus 1–3 colonies on the other days; Table S5). Although these observations confirmed the occurrence of a split spawning, they may not reflect natural release patterns, given the extended time these colonies were removed from natural cues (e.g., exposure to lunar cycle and lack of exposure to tides). In addition, their prolonged close proximity to one another in the single tank ("raceway") in which they were kept may have affected the timing of release (e.g., due to chemical signaling).[73] The observations confirm the potential for experimental inter-lineage crossings under artificial conditions to assess the additional or alternative role of gametic incompatibility in reproductive isolation.

Geographic and habitat structuring within lineages

Within each of the three Australasian lineages, principal component analyses identified considerable substructuring, which was driven to a large extent by sampling region (GBR, WCS, and PNG; Figures 5A and 5C). However, additional substructure was identified (2 to 3 apparent clusters) in the GBR samples that could not be linked to subregions, locations, or habitat (Figure 5C). Discriminant analysis of principal components (DAPC) further confirmed the genetic differentiation among the three sampling regions (Figure 5B), indicating restricted gene flow between these regions, with some exceptions indicating recent admixture. Although genetic differentiation between the GBR and WCS has been observed previously in other coral species,[75,76] here, we identify further differentiation between

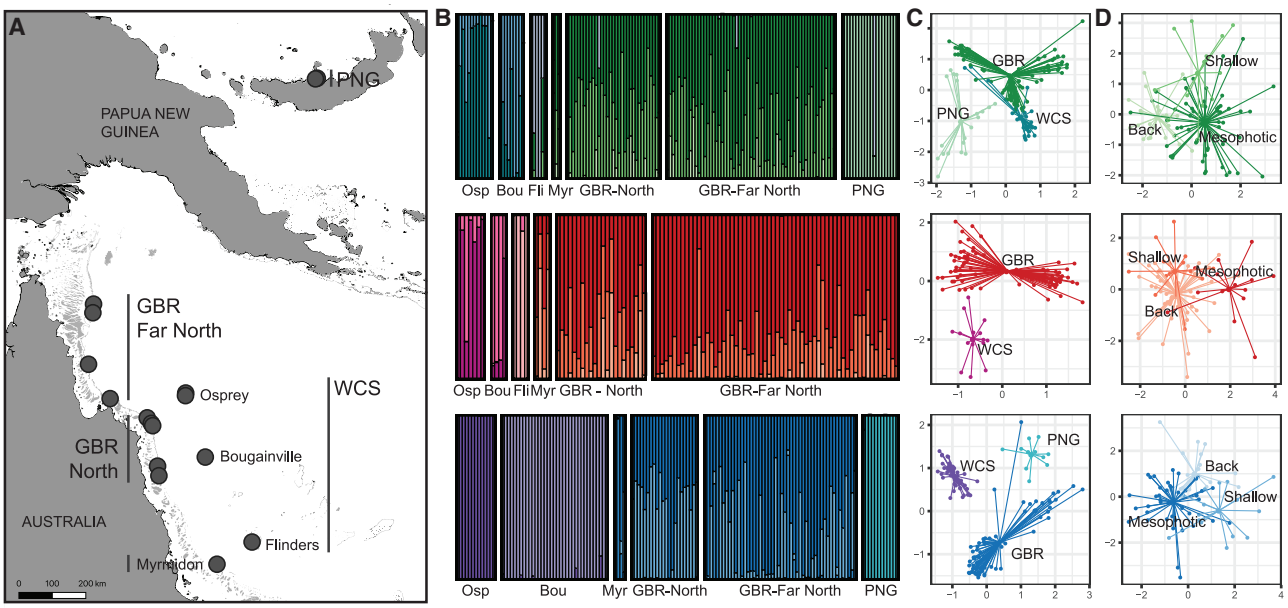

Figure 5. Genetic structuring within lineages by region and habitat

(A) Map showing the Australasian sampling locations/regions included in the DAPC plots.

(B) DAPC assignment plots for each of the three lineages (from top to bottom: green, red, and blue) using location as prior (but ignoring habitat and grouping "North" and "Far North" locations on the Great Barrier Reef).

(C) PCA for each of the three lineages, showing structuring by region as well as unexplained substructuring within the Great Barrier Reef populations.

(D) DAPC scatterplot for the Great Barrier Reef using habitat as prior. PCA and DAPC scatterplots show individual colonies connected to, respectively, region and habitat centroids, and analyses are based on nextRAD data (6,129–6,576 SNPs) after the removal of outliers.

See also Figure S5.

individual WCS atolls, confirming their rather isolated nature as they are surrounded by deep oceanic water. In contrast, only very subtle substructuring was identified between the "Northern" and "Far Northern" regions of the GBR (Figure 5C). The inability to detect distinct location- or region-associated clusters on the GBR could indicate the potential for gene flow over large distances, a pattern observed for other coral species and facilitated by the high density of reefs along the outer shelf.[75–79] Nonetheless, given the presence of substructuring within each of the locations (a pattern frequently observed in scleractinian corals),[22] such "panmixia" should be interpreted with caution.

Where other studies found a clear partitioning of cryptic lineages across habitats,[29,30] we observed substantial overlap between the depth distributions of the three *P. speciosa* lineages (e.g., the red lineage is present at shallow and mesophotic depths but at lower relative abundances in the latter). Surveys during the 2016 mass bleaching found that most *Pachyseris* colonies down to 25 m were bleached, although fewer than half of the colonies at 40 m depth were bleached.[80] Opportunistic genotyping of 14 healthy and 3 bleached colonies collected from 40 m during this event found healthy representatives of all three lineages (8 green, 5 red, and 1 blue healthy colony versus 2 green and 1 blue bleached colony). The broad depth range of these lineages combined with the strong depth attenuation of bleaching in *Pachyseris*[80] raises the question whether deep populations may benefit those at shallow depth by acting as a refuge and source of reproduction.[81] Unfortunately, genetic differentiation between habitats (within each lineage) could not be adequately assessed at individual locations, given the small sample sizes after splitting into cryptic lineages. When merging individual locations on the GBR, DAPC showed some discrimination between habitat clusters (Figures 5D, S5A, and S5B). This separation was driven by many low-contributing and several high-contributing alleles, indicating there may be potential limitations to such vertical connectivity (Figure S5C). However, as confounding factors may be at play (due to merging locations), this should be further explored through a replicated sample design across multiple locations with sufficient colonies from each lineage at each depth.

Conclusions

Tens of thousands of multicellular species have been described in association with tropical coral reefs, yet this is estimated to represent <10% of their actual species richness.[1] Reef-building corals (Scleractinia) have been considered to contribute relatively little to this estimated discrepancy, with the taxonomy considered to be relatively complete compared to other less-studied coral reef taxa.[1] Molecular approaches have indeed corroborated many of the current taxonomic species[82] but have also unveiled extensive undescribed diversity within the traditional morphological boundaries.[21,26] This pattern is likely to be accelerated given the transition to genome-wide sequencing approaches, solving the resolution issues associated with traditional sequence markers.[31,83,84] The inability to discriminate this diversity as readily identifiable taxonomic units in the field (or even through close skeletal examination) has raised concerns about the practicality and necessity of incorporating it into our systematic framework, particularly as the lack of

morphological differentiation can lead to the assumption of negligible physiological differentiation. Although we found no diagnostic morphological characters distinguishing the three *Pachyseris* lineages in this study, the genome-wide differentiation was accompanied by substantial differences in ecological and physiological traits, demonstrating how morphological stasis can mask substantial ecological divergence. Failure to recognize such cryptic diversity will result in erroneous interpretations of species distributions, extinction risk, and spatial genetic differentiation,[21,26] all with critical ramifications for conservation management. In light of the rapid degradation of the world's coral reefs, it is critical to acknowledge and start capturing this hidden diversity to improve our understanding and ability to protect these fragile ecosystems.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Specimen used for genome sequencing
  - Coral specimen collections
- METHOD DETAILS
  - Overall summary of methods
  - Genome assembly and annotation
  - Reduced-representation sequencing
  - CAPS marker development and genotyping
  - Whole-genome re-sequencing
  - Characterization of microbial communities
  - Phenotypic characterization
  - Reproductive characterization
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Ecological differentiation
  - Phenotypic differentiation
  - Microbial community structure

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.03.028.

ACKNOWLEDGMENTS

We thank David Whillas, Jaap Barendrecht, David Harris, Sara Naylor, and Annamieke van den Heuvel for support in the field or in the lab, as well as Underwater Earth, The Ocean Agency, and crews from Reef Connections, Mike Ball Dive Expeditions, SY Ethereal, and the Waitt Foundation. The authors acknowledge the Reef Future Genomics (ReFuGe) 2020 Consortium, of which this study was part, as organized by the Great Barrier Reef Foundation. This project was supported (in chronological order) by the XL Catlin Seaview Survey (2012–2014; funded by the XL Catlin Group in partnership with Underwater Earth and The University of Queensland), an Accelerate Partnerships grant from the Department of Science Information Technology Innovation and the Arts of the Queensland Government (2014–2016), an Australian Research Council Discovery Early Career Researcher Award (2016–2018; DE160101433 awarded to P.B.), and the Hope for Reefs Initiative at the California Academy of Sciences (2018–2020). Additional support was received from the Australian Research Council Centre for Excellence in Coral Reef Studies at The University of Queensland (awarded to O.H.-G.) and through a Natural Sciences grant no. 24133 from the Mitsubishi Foundation (awarded to S.H.). Substantial sea time was generously provided by the Waitt Foundation and the Joy Foundation. The genome assembly and whole-genome sequencing was funded by the Great Barrier Reef Foundation "Resilient Coral Reefs Successfully Adapting to Climate Change" program in collaboration with the Australian Government, Bioplatforms Australia through the National Collaborative Research Infrastructure Strategy, Rio Tinto, and a family foundation.

AUTHOR CONTRIBUTIONS

P.B., I.R.C., H.Y., D.W., S.d.H., M.A.R., M.J.H.v.O., and O.H.-G. conceived and designed the research. P.B., D.W., S.d.H., C.A.B., S.D., N.E., G.E., M.G., K.B.H., D.C.H., M.M., P.M., and G.T. contributed to lab work. P.B., N.E., G.E., K.B.H., S.H., A.M., F.S., P.S.-W., and G.T. contributed to specimen collections. P.B., I.R.C., H.Y., D.W., S.d.H., A.H.-A., S.F., and Y.L. contributed to analyses. P.B. wrote the paper, and all authors contributed manuscript edits.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 5, 2020
Revised: February 13, 2021
Accepted: March 9, 2021
Published: April 2, 2021

REFERENCES

1. Fisher, R., O'Leary, R.A., Low-Choy, S., Mengersen, K., Knowlton, N., Brainard, R.E., and Caley, M.J. (2015). Species richness on coral reefs and the pursuit of convergent global estimates. Curr. Biol. *25*, 500–505.
2. Veron, J.E.N., Stafford-Smith, M.G., Turak, E., and DeVantier, L.M. (2020). Corals of the world. http://www.coralsoftheworld.org/page/home/.
3. DeVantier, L., and Turak, E. (2017). Species richness and relative abundance of reef-building corals in the Indo-West Pacific. Diversity *9*, 25.
4. Madin, J.S., Anderson, K.D., Andreasen, M.H., Bridge, T.C., Cairns, S.D., Connolly, S.R., Darling, E.S., Diaz, M., Falster, D.S., Franklin, E.C., et al. (2016). The Coral Trait Database, a curated database of trait information for coral species from the global oceans. Sci. Data *3*, 160017.
5. WoRMS Editorial Board (2020). World Register of Marine Species. https://www.marinespecies.org.
6. Best, M.B., Boekschoten, G.J., and Oosterbaan, A. (1984). Species concept and ecomorph variation in living and fossil Scleractinia. Palaeontogr. Am. *54*, 70–79.
7. Knowlton, N., and Jackson, J.B. (1994). New taxonomy and niche partitioning on coral reefs: jack of all trades or master of some? Trends Ecol. Evol. *9*, 7–9.
8. Todd, P.A. (2008). Morphological plasticity in scleractinian corals. Biol. Rev. Camb. Philos. Soc. *83*, 315–337.
9. Fukami, H., Budd, A.F., Paulay, G., Solé-Cava, A., Allen Chen, C., Iwao, K., and Knowlton, N. (2004). Conventional taxonomy obscures deep divergence between Pacific and Atlantic corals. Nature *427*, 832–835.
10. Huang, D., Licuanan, W.Y., Baird, A.H., and Fukami, H. (2011). Cleaning up the 'Bigmessidae': molecular phylogeny of scleractinian corals from Faviidae, Merulinidae, Pectiniidae and Trachyphylliidae. BMC Evol. Biol. *11*, 37.
11. Schmidt-Roach, S., Miller, K.J., Lundgren, P., and Andreakis, N. (2014). With eyes wide open: a revision of species within and closely related to the *Pocillopora damicornis* species complex (Scleractinia; Pocilloporidae) using morphology and genetics. Zool. J. Linn. Soc. *170*, 1–33.

12. Forsman, Z.H., Knapp, I.S.S., Tisthammer, K., Eaton, D.A.R., Belcaid, M., and Toonen, R.J. (2017). Coral hybridization or phenotypic variation? Genomic data reveal gene flow between *Porites lobata* and *P. Compressa*. Mol. Phylogenet. Evol. *111*, 132–148.

13. Knowlton, N., Weil, E., Weigt, L.A., and Guzmán, H.M. (1992). Sibling species in *Montastraea annularis*, coral bleaching, and the coral climate record. Science *255*, 330–333.

14. Weil, E., and Knowton, N. (1994). A multi-character analysis of the Caribbean coral *Montastraea annularis* (Ellis and Solander, 1786) and its two sibling species, *M. faveolata* (Ellis and Solander, 1786) and *M. franksi* (Gregory, 1895). Bull. Mar. Sci. *55*, 151–175.

15. Forsman, Z.H., Barshis, D.J., Hunter, C.L., and Toonen, R.J. (2009). Shape-shifting corals: molecular markers show morphology is evolutionarily plastic in *Porites*. BMC Evol. Biol. *9*, 45.

16. Bongaerts, P., Riginos, C., Ridgway, T., Sampayo, E.M., van Oppen, M.J., Englebert, N., Vermeulen, F., and Hoegh-Guldberg, O. (2010). Genetic divergence across habitats in the widespread coral *Seriatopora hystrix* and its associated *Symbiodinium*. PLoS ONE *5*, e10871.

17. Frade, P.R., Reyes-Nivia, M.C., Faria, J., Kaandorp, J.A., Luttikhuizen, P.C., and Bak, R.P.M. (2010). Semi-permeable species boundaries in the coral genus *Madracis*: introgression in a brooding coral system. Mol. Phylogenet. Evol. *57*, 1072–1090.

18. Stefani, F., Benzoni, F., Yang, S.-Y., Pichon, M., Galli, P., and Chen, C.A. (2011). Comparison of morphological and genetic analyses reveals cryptic divergence and morphological plasticity in *Stylophora* (Cnidaria, Scleractinia). Coral Reefs *30*, 1033.

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

20. Warner, P.A., van Oppen, M.J., and Willis, B.L. (2015). Unexpected cryptic species diversity in the widespread coral *Seriatopora hystrix* masks spatial-genetic patterns of connectivity. Mol. Ecol. *24*, 2993–3008.

21. Richards, Z.T., Berry, O., and van Oppen, M.J.H. (2016). Cryptic genetic divergence within threatened species of *Acropora* coral from the Indian and Pacific Oceans. Conserv. Genet. *17*, 577–591.

22. Gélin, P., Fauvelot, C., Bigot, L., Baly, J., and Magalon, H. (2017). From population connectivity to the art of striping Russian dolls: the lessons from *Pocillopora* corals. Ecol. Evol. *8*, 1411–1426.

23. Sinniger, F., Prasetia, R., Yorifuji, M., Bongaerts, P., and Harii, S. (2017). *Seriatopora* diversity preserved in upper mesophotic coral ecosystems in Southern Japan. Front. Mar. Sci. *4*, 155.

24. Arrigoni, R., Berumen, M.L., Stolarski, J., Terraneo, T.I., and Benzoni, F. (2019). Uncovering hidden coral diversity: a new cryptic lobophylliid scleractinian from the Indian Ocean. Cladistics *35*, 301–328.

25. Knowlton, N. (1993). Sibling species in the sea. Annu. Rev. Ecol. Syst. *24*, 189–216.

26. Sheets, E.A., Warner, P.A., and Palumbi, S.R. (2018). Accurate population genetic measurements require cryptic species identification in corals. Coral Reefs *37*, 549–563.

27. González, A.M., Prada, C.A., Ávila, V., and Medina, M. (2018). Ecological speciation in corals. In Population Genomics: Marine Organisms, M. Oleksiak, and O. Rajora, eds. (Springer), pp. 303–324.

28. Rose, N.H., Bay, R.A., Morikawa, M.K., and Palumbi, S.R. (2018). Polygenic evolution drives species divergence and climate adaptation in corals. Evolution *72*, 82–94.

29. Bongaerts, P., Riginos, C., Hay, K.B., van Oppen, M.J., Hoegh-Guldberg, O., and Dove, S. (2011). Adaptive divergence in a scleractinian coral: physiological adaptation of *Seriatopora hystrix* to shallow and deep reef habitats. BMC Evol. Biol. *11*, 303.

30. Prada, C., and Hellberg, M.E. (2013). Long prereproductive selection and divergence by depth in a Caribbean candelabrum coral. Proc. Natl. Acad. Sci. USA *110*, 3961–3966.

31. Rosser, N.L., Thomas, L., Stankowski, S., Richards, Z.T., Kennington, W.J., and Johnson, M.S. (2017). Phylogenomics provides new insight into evolutionary relationships and genealogical discordance in the reef-building coral genus *Acropora*. Proc. Biol. Sci. *284*, 20162182.

32. Bongaerts, P., Riginos, C., Brunner, R., Englebert, N., Smith, S.R., and Hoegh-Guldberg, O. (2017). Deep reefs are not universal refuges: reseeding potential varies among coral species. Sci. Adv. *3*, e1602373.

33. Bongaerts, P., Sampayo, E.M., Bridge, T.C.L., Ridgway, T., Vermeulen, F., Englebert, N., Webster, J.M., and Hoegh-Guldberg, O. (2011). *Symbiodinium* diversity in mesophotic coral communities on the Great Barrier Reef: a first assessment. Mar. Ecol. Prog. Ser. *439*, 117–126.

34. Cooper, T.F., Ulstrup, K.E., Dandan, S.S., Heyward, A.J., Kühl, M., Muirhead, A., O'Leary, R.A., Ziersen, B.E., and Van Oppen, M.J. (2011). Niche specialization of reef-building corals in the mesophotic zone: metabolic trade-offs between divergent *Symbiodinium* types. Proc. Biol. Sci. *278*, 1840–1850.

35. Ziegler, M., Roder, C.M., Büchel, C., and Voolstra, C.R. (2015). Mesophotic coral depth acclimatization is a function of host-specific symbiont physiology. Front. Mar. Sci. *2*, 4.

36. Radice, V.Z., Hoegh-Guldberg, O., Fry, B., Fox, M.D., and Dove, S.G. (2019). Upwelling as the major source of nitrogen for shallow and deep reef-building corals across an oceanic atoll system. Funct. Ecol. *33*, 1120–1134.

37. Hernandez-Agreda, A., Leggat, W., Bongaerts, P., and Ainsworth, T.D. (2016). The microbial signature provides insight into the mechanistic basis of coral success across reef habitats. MBio *7*, e00560-16.

38. IUCN (International Union for Conservation of Nature) (2019). Pachyseris speciosa. The IUCN Red List of Threatened Species. https://www.iucnredlist.org.

39. Romano, S.L., and Palumbi, S.R. (1996). Evolution of scleractinian corals inferred from molecular systematics. Science *271*, 640–642.

40. Ying, H., Cooke, I., Sprungala, S., Wang, W., Hayward, D.C., Tang, Y., Huttley, G., Ball, E.E., Forêt, S., and Miller, D.J. (2018). Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. Genome Biol. *19*, 175.

41. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. *27*, 722–736.

42. Flot, J.-F., Blanchot, J., Charpy, L., Cruaud, C., Licuanan, W.Y., Nakano, Y., Payri, C., and Tillier, S. (2011). Incongruence between morphotypes and genetically delimited species in the coral genus *Stylophora*: phenotypic plasticity, morphological convergence, morphological stasis or interspecific hybridization? BMC Ecol. *11*, 22.

43. Arrigoni, R., Berumen, M.L., Mariappan, K.G., Beck, P.S.A., Hulver, A.M., Montano, S., Pichon, M., Strona, G., Terraneo, T.I., and Benzoni, F. (2020). Towards a rigorous species delimitation framework for scleractinian corals based on RAD sequencing: the case study of *Leptastrea* from the Indo-Pacific. Coral Reefs *39*, 1001–1025.

44. Prada, C., Hanna, B., Budd, A.F., Woodley, C.M., Schmutz, J., Grimwood, J., Iglesias-Prieto, R., Pandolfi, J.M., Levitan, D., Johnson, K.G., et al. (2016). Empty niches after extinctions increase population sizes of modern corals. Curr. Biol. *26*, 3190–3194.

45. Veron, J.E.N., and Pichon, M. (1980). Scleractinia of Eastern Australia. III. Families Agariciidae, Siderastreidae, Fungiidae, Oculiniidae, Merulinidae, Pectiniidae, Caryophylliidae, Dendrophylliidae (Australian National University Press for the Australian Institute of Marine Science).

46. Rentzsch, F., Fritzenwanker, J.H., Scholz, C.B., and Technau, U. (2008). FGF signalling controls formation of the apical sensory organ in the cnidarian *Nematostella vectensis*. Development *135*, 1761–1769.

47. Reyes-Bermudez, A., Villar-Briones, A., Ramirez-Portilla, C., Hidaka, M., and Mikheyev, A.S. (2016). Developmental progression in the coral *Acropora digitifera* is controlled by differential expression of distinct regulatory gene networks. Genome Biol. Evol. *8*, 851–870.

48. Strader, M.E., Aglyamova, G.V., and Matz, M.V. (2018). Molecular characterization of larval development from fertilization to metamorphosis in a reef-building coral. BMC Genomics *19*, 17.

49. Cleves, P.A., Strader, M.E., Bay, L.K., Pringle, J.R., and Matz, M.V. (2018). CRISPR/Cas9-mediated genome editing in a reef-building coral. Proc. Natl. Acad. Sci. USA *115*, 5235–5240.

50. Takechi, M., Adachi, N., Hirai, T., Kuratani, S., and Kuraku, S. (2013). The Dlx genes as clues to vertebrate genomics and craniofacial evolution. Semin. Cell Dev. Biol. *24*, 110–118.

51. Gonzalez Angel, A.M. (2019). Genomic and transcriptomic insights on the Orbicella species complex (The Pennsylvania State University), Ph.D. dissertation.

52. Ramos-Silva, P., Kaandorp, J., Huisman, L., Marie, B., Zanella-Cléon, I., Guichard, N., Miller, D.J., and Marin, F. (2013). The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. Mol. Biol. Evol. *30*, 2099–2112.

53. Rosenberg, Y., Doniger, T., Harii, S., Sinniger, F., and Levy, O. (2017). Canonical and cellular pathways timing gamete release in *Acropora digitifera*, Okinawa, Japan. Mol. Ecol. *26*, 2698–2710.

54. Mayfield, A.B., Chen, Y.-J., Lu, C.-Y., and Chen, C.-S. (2018). The proteomic response of the reef coral *Pocillopora acuta* to experimentally elevated temperatures. PLoS ONE *13*, e0192001.

55. Voolstra, C.R., Schnetzer, J., Peshkin, L., Randall, C.J., Szmant, A.M., and Medina, M. (2009). Effects of temperature on gene expression in embryos of the coral *Montastraea faveolata*. BMC Genomics *10*, 627.

56. van Oppen, M.J.H., Bongaerts, P., Frade, P., Peplow, L.M., Boyd, S.E., Nim, H.T., and Bay, L.K. (2018). Adaptation to reef habitats through selection on the coral animal and its associated microbiome. Mol. Ecol. *27*, 2956–2971.

57. Liew, Y.J., Howells, E.J., Wang, X., Michell, C.T., Burt, J.A., Idaghdour, Y., and Aranda, M. (2020). Intergenerational epigenetic inheritance in reef-building corals. Nat. Clim. Chang. *10*, 254–259.

58. Fine, M., Gildor, H., and Genin, A. (2013). A coral reef refuge in the Red Sea. Glob. Change Biol. *19*, 3640–3647.

59. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925.

60. Mao, Y., Economo, E.P., and Satoh, N. (2018). The roles of introgression and climate change in the rise to dominance of Acropora corals. Curr. Biol. *28*, 3373–3382.e5.

61. Rothfels, C.J., Larsson, A., Kuo, L.-Y., Korall, P., Chiou, W.-L., and Pryer, K.M. (2012). Overcoming deep roots, fast rates, and short internodes to resolve the ancient rapid radiation of eupolypod II ferns. Syst. Biol. *61*, 490–509.

62. Cowman, P.F., and Bellwood, D.R. (2013). Vicariance across major marine biogeographic barriers: temporal concordance and the relative intensity of hard versus soft barriers. Proc. Biol. Sci. *280*, 20131541.

63. van Oppen, M.J., Willis, B.L., and Miller, D.J. (1999). Atypically low rate of cytochrome b evolution in the scleractinian coral genus *Acropora*. Proc. Biol. Sci. *266*, 179–183.

64. Shearer, T.L., Van Oppen, M.J.H., Romano, S.L., and Wörheide, G. (2002). Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol. Ecol. *11*, 2475–2487.

65. Liu, H., Stephens, T.G., González-Pech, R.A., Beltran, V.H., Lapeyre, B., Bongaerts, P., Cooke, I., Aranda, M., Bourne, D.G., Forêt, S., et al. (2018). *Symbiodinium* genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. Commun. Biol. *1*, 95.

66. Babcock, R.C., Bull, G.D., Harrison, P.L., Heyward, A.J., Oliver, J.K., Wallace, C.C., and Willis, B.L. (1986). Synchronous spawnings of 105 scleractinian coral species on the Great Barrier Reef. Mar. Biol. *90*, 379–394.

67. Levitan, D.R. (2005). Sex-specific spawning behavior and its consequences in an external fertilizer. Am. Nat. *165*, 682–694.

68. Fukami, H., Omori, M., Shimoike, K., Hayashibara, T., and Hatta, M. (2003). Ecological and genetic aspects of reproductive isolation by different spawning times in Acropora corals. Mar. Biol. *142*, 679–684.

69. Levitan, D.R., Fukami, H., Jara, J., Kline, D., McGovern, T.M., McGhee, K.E., Swanson, C.A., and Knowlton, N. (2004). Mechanisms of reproductive isolation among sympatric broadcast-spawning corals of the *Montastraea annularis* species complex. Evolution *58*, 308–323.

70. Rosser, N.L. (2015). Asynchronous spawning in sympatric populations of a hard coral reveals cryptic species and ancient genetic lineages. Mol. Ecol. *24*, 5006–5019.

71. Gilmour, J.P., Underwood, J.N., Howells, E.J., Gates, E., and Heyward, A.J. (2016). Biannual spawning and temporal reproductive isolation in *Acropora* corals. PLoS ONE *11*, e0150916.

72. Baird, A.H., Guest, J.R., and Willis, B.L. (2009). Systematic and biogeographical patterns in the reproductive biology of scleractinian corals. Annu. Rev. Ecol. Evol. Syst. *40*, 551–571.

73. Levitan, D.R., Fogarty, N.D., Jara, J., Lotterhos, K.E., and Knowlton, N. (2011). Genetic, spatial, and temporal components of precise spawning synchrony in reef building corals of the *Montastraea annularis* species complex. Evolution *65*, 1254–1270.

74. Kaniewska, P., Alon, S., Karako-Lampert, S., Hoegh-Guldberg, O., and Levy, O. (2015). Signaling cascades and the importance of moonlight in coral broadcast mass spawning. eLife *4*, e09991.

75. van Oppen, M.J., Lutz, A., De'ath, G., Peplow, L., and Kininmonth, S. (2008). Genetic traces of recent long-distance dispersal in a predominantly self-recruiting coral. PLoS ONE *3*, e3401.

76. Lukoschek, V., Riginos, C., and van Oppen, M.J. (2016). Congruent patterns of connectivity can inform management for broadcast spawning corals on the Great Barrier Reef. Mol. Ecol. *25*, 3065–3080.

77. Van Oppen, M.J., Peplow, L.M., Kininmonth, S., and Berkelmans, R. (2011). 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.

78. Torda, G., Lundgren, P., Willis, B.L., and van Oppen, M.J. (2013). Revisiting the connectivity puzzle of the common coral *Pocillopora damicornis*. Mol. Ecol. *22*, 5805–5820.

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

80. Frade, P.R., Bongaerts, P., Englebert, N., Rogers, A., Gonzalez-Rivero, M., and Hoegh-Guldberg, O. (2018). Deep reefs of the Great Barrier Reef offer limited thermal refuge during mass coral bleaching. Nat. Commun. *9*, 3447.

81. Bongaerts, P., and Smith, T.B. (2019). Beyond the "Deep Reef Refuge" hypothesis: a conceptual framework to characterize persistence at depth. In Mesophotic Coral Ecosystems Coral Reefs of the World, Y. Loya, K.A. Puglise, and T.C.L. Bridge, eds. (Springer International Publishing), pp. 881–895.

82. Veron, J. (2013). Overview of the taxonomy of zooxanthellate Scleractinia. Zool. J. Linn. Soc. *169*, 485–508.

83. Quattrini, A.M., Faircloth, B.C., Dueñas, L.F., Bridge, T.C.L., Brugler, M.R., Calixto-Botía, I.F., DeLeo, D.M., Forêt, S., Herrera, S., Lee, S.M.Y., et al. (2018). Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: new approaches to long-standing problems. Mol. Ecol. Resour. *18*, 281–295.

84. Quek, R.Z.B., Jain, S.S., Neo, M.L., Rouse, G.W., and Huang, D. (2020). Transcriptome-based target-enrichment baits for stony corals (Cnidaria: Anthozoa: Scleractinia). Mol. Ecol. Resour. *20*, 807–818.

85. Simpson, J.T. (2014). Exploring genome characteristics and sequence quality without a reference. Bioinformatics *30*, 1228–1235.

86. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast

reference-free genome profiling from short reads. Bioinformatics *33*, 2202–2204.

87. Huang, S., Kang, M., and Xu, A. (2017). HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics *33*, 2577–2579.
88. Chaisson, M.J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics *13*, 238.
89. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.
90. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666.
91. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. *34*, W435–W439.
92. Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics *5*, 59.
93. Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics *12*, 491.
94. Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods *9*, 173–175.
95. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. *35*, 543–548.
96. Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. *7*, e1002195.
97. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
98. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., and Thibault, J. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 11.10.1–11.10.33.
99. Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics *27*, 3070–3071.
100. Talevich, E., Invergo, B.M., Cock, P.J., and Chapman, B.A. (2012). Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. BMC Bioinformatics *13*, 209.
101. Beugin, M.-P., Gayet, T., Pontier, D., Devillard, S., and Jombart, T. (2018). A fast likelihood solution to the genetic clustering problem. Methods Ecol. Evol. *9*, 1006–1016.
102. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.
103. Jakobsson, M., and Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics *23*, 1801–1806.
104. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
105. Privé, F., Luu, K., Vilhjálmsson, B.J., and Blum, M.G.B. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. Mol. Biol. Evol. *37*, 2153–2154.
106. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and 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 (Austin) *6*, 80–92.
107. Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F.J., Lippman, Z.B., and Schatz, M.C. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. *20*, 224.
108. Kofler, R., and Schlötterer, C. (2012). Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. Bioinformatics *28*, 2084–2085.
109. Eaton, D.A.R., and Overcast, I. (2020). ipyrad: interactive assembly and analysis of RADseq datasets. Bioinformatics *36*, 2592–2594.
110. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
111. Yu, G. (2020). Using ggtree to visualize data on tree-like structures. Curr. Protoc. Bioinformatics *69*, e96.
112. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.
113. Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. *11*, 94.
114. Eaton, D.A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics *30*, 1844–1849.
115. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3–new capabilities and interfaces. Nucleic Acids Res. *40*, e115.
116. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.
117. Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. Genetics *210*, 719–731.
118. Schiffels, S., and Wang, K. (2020). MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. Methods Mol. Biol. *2090*, 147–166.
119. Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–a baiting and iterative mapping approach. Nucleic Acids Res. *41*, e129.
120. Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., and Stadler, P.F. (2013). MITOS: improved *de novo* metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. *69*, 313–319.
121. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.
122. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.
123. Leigh, J.W., and Bryant, D. (2015). POPART: full-feature software for haplotype network construction. Methods Ecol. Evol. *6*, 1110–1116.
124. Clement, M., Posada, D., and Crandall, K.A. (2000). TCS: a computer program to estimate gene genealogies. Mol. Ecol. *9*, 1657–1659.
125. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.
126. Anderson, M., Gorley, R., and Clarke, K.P. (2008). PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods (PRIMER-E).
127. Blin, N., and Stafford, D.W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res. *3*, 2303–2308.

128. Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods *13*, 1050–1054.

129. Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA *6*, 11.

130. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358–1370.

131. Cooke, I., Ying, H., Forêt, S., Bongaerts, P., Strugnell, J.M., Simakov, O., Zhang, J., Field, M.A., Rodriguez-Lanetty, M., Bell, S.C., et al. (2020). Genomic signatures in the coral holobiont reveal host adaptations driven by Holocene climate change and reef specific symbionts. Sci. Adv. *6*, eabc6318.

132. Chifman, J., and Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. Bioinformatics *30*, 3317–3324.

133. Gower, G., Tuke, J., Rohrlach, A.B., Soubrier, J., Llamas, B., Bean, N., and Cooper, A. (2018). Population size history from short genomic scaffolds: how short is too short? bioRxiv. https://doi.org/10.1101/382036.

134. Sunagawa, S., Woodley, C.M., and Medina, M. (2010). Threatened corals provide underexplored microbial habitats. PLoS ONE *5*, e9554.

135. Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. *73*, 5261–5267.

136. DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. *72*, 5069–5072.

137. Budd, A.F., Fukami, H., Smith, N.D., and Knowlton, N. (2012). Taxonomic classification of the reef coral family Mussidae (Cnidaria: Anthozoa: Scleractinia). Zool. J. Linn. Soc. *166*, 465–529.

138. Terraneo, T.I., Berumen, M.L., Arrigoni, R., Waheed, Z., Bouwmeester, J., Caragnano, A., Stefani, F., and Benzoni, F. (2014). *Pachyseris inattesa* sp. n. (Cnidaria, Anthozoa, Scleractinia): a new reef coral species from the Red Sea and its phylogenetic relationships. ZooKeys *433*, 1–30.

139. Strychar, K.B., Coates, M., Sammarco, P.W., Piva, T.J., and Scott, P.T. (2005). Loss of *Symbiodinium* from bleached soft corals *Sarcophyton ehrenbergi*, *Sinularia* sp. and *Xenia* sp. J. Exp. Mar. Biol. Ecol. *320*, 159–177.

140. Dove, S.G., Lovell, C., Fine, M., Deckenback, J., Hoegh-Guldberg, O., Iglesias-Prieto, R., and Anthony, K.R. (2008). Host pigments: potential facilitators of photosynthesis in coral symbioses. Plant Cell Environ. *31*, 1523–1533.

141. Whitaker, J.R., and Granum, P.E. (1980). An absolute method for protein determination based on difference in absorbance at 235 and 280 nm. Anal. Biochem. *109*, 156–159.

142. Folch, J., Lees, M., and Sloane Stanley, G.H. (1957). A simple method for the isolation and purification of total lipides from animal tissues. J. Biol. Chem. *226*, 497–509.

143. Dunn, S.R., Thomas, M.C., Nette, G.W., and Dove, S.G. (2012). A lipidomic approach to understanding free fatty acid lipogenesis derived from dissolved inorganic carbon within cnidarian-dinoflagellate symbiosis. PLoS ONE *7*, e46801.

144. Zapata, M., Rodríguez, F., and Garrido, J.L. (2000). Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases. Mar. Ecol. Prog. Ser. *195*, 29–45.

145. Dove, S., Ortiz, J.C., Enríquez, S., Fine, M., Fisher, P., Iglesias-Prieto, R., Thornhill, D., and Hoegh-Guldberg, O. (2006). Response of holosymbiont pigments from the scleractinian coral *Montipora monasteriata* to short-term heat stress. Limnol. Oceanogr. *51*, 1149–1158.

146. Harwell, M.R., Rubinstein, E.N., Hayes, W.S., and Olds, C.C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. J. Educ. Stat. *17*, 315–339.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Genome assembly	This paper	NCBI: PRJNA686157
Raw sequence data for genome and transcriptome assembly	This paper	ENA: PRJEB23386
Raw sequence data for reduced-representation (nextRAD) sequencing	This paper	NCBI: PRJNA701715
Whole-genome resequencing	This paper	NCBI: PRJNA686482
16S rRNA amplicon sequencing	Hernandez-Agreda et al.[37]	NCBI: PRJNA328211
Oligonucleotides		
Pspe-Green-CfoI/HhaI: F: ACCTGGT GACCTTTGCCATA; R: TCTGTCAGT AGAGGGAGGGG	This paper	N/A
Pspe-Blue-HaeIII: F: CCGTTTCTTC GTCAGCCATT; R: CACATCGCTCT TCTTCCGTT	This paper	N/A
Pspe-Red-Taqα1: F: TAATCGCACT GCTAGGGACG, R: CTTGGTCTGTT TGTAGCCGT	This paper	N/A
Software and algorithms		
FastQC v0.11.6	N/A	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
sga.preqc	Simpson[85]	https://github.com/jts/sga/wiki/preqc
GenomeScope	Vurture et al.[86]	https://github.com/schatzlab/genomescope
CANU v1.5	Koren et al.[41]	https://github.com/marbl/canu
HaploMerger2	Huang et al.[87]	https://github.com/mapleforest/HaploMerger2
BLASR	Chaisson and Tesler[88]	https://github.com/PacificBiosciences/blasr
RepeatModeler v1.0.11	N/A	http://www.repeatmasker.org/RepeatModeler
RepeatMasker v4.0.8	N/A	http://www.repeatmasker.org/
Trinity	Grabherr et al.[89]	https://github.com/trinityrnaseq/trinityrnaseq
PSyTrans	N/A	https://github.com/sylvainforet/psytrans
PASA	Haas et al.[90]	https://github.com/PASApipeline/PASApipeline/wiki
AUGUSTUS	Stanke et al.[91]	http://bioinf.uni-greifswald.de/augustus/
SNAP	Korf[92]	https://github.com/KorfLab/SNAP
MAKER2	Holt and Yandell[93]	http://www.yandell-lab.org/software/maker.html
transposonPSI	N/A	http://transposonpsi.sourceforge.net
hhblits	Remmert et al.[94]	https://toolkit.tuebingen.mpg.de/tools/hhblits
BUSCO v3	Waterhouse et al.[95]	https://busco-archive.ezlab.org/
HMMER v3	Eddy[96]	http://hmmer.org/
Rad-seq script collection	N/A	https://github.com/pimbongaerts/radseq
TrimGalore	N/A	https://github.com/FelixKrueger/TrimGalore
BWA-MEM	Li and Durbin[97]	https://github.com/lh3/bwa
GATK (v3/4) / Picard Tools	Van der Auwera et al.[98]	https://gatk.broadinstitute.org/hc/en-us
vcf_clone_detect.py	This study	https://github.com/pimbongaerts/radseq
adegenet	Jombart and Ahmed[99]	https://github.com/thibautjombart/adegenet
Bio.Phylo	Talevich et al.[100]	https://biopython.org/
snapclust (adegenet)	Beugin et al.[101]	https://github.com/thibautjombart/adegenet
structure_mp.py	Bongaerts et al.[32]	https://github.com/pimbongaerts/radseq

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
STRUCTURE v2.3.4	Pritchard et al.[102]	https://web.stanford.edu/group/pritchardlab/structure.html
CLUMPP v1.1.2	Jakobsson and Rosenberg[103]	https://rosenberglab.stanford.edu/clumpp.html
vcftools	Danecek et al.[104]	https://github.com/vcftools/vcftools
pcadapt v4	Privé et al.[105]	https://github.com/bcm-uga/pcadapt
SnpEff	Cingolani et al.[106]	https://pcingola.github.io/SnpEff/
RaGOO	Alonge et al.[107]	https://github.com/malonge/RaGOO
Gowinda	Kofler and Schlötterer[108]	https://sourceforge.net/p/gowinda/wiki/Main/
ipyrad v0.9.62 / tetrad	Eaton and Overcast[109]	https://github.com/dereneaton/ipyrad
RAxML	Stamatakis[110]	https://cme.h-its.org/exelixis/web/software/raxml/
ggtree	Yu[111]	https://github.com/YuLab-SMU/ggtree
ape	Paradis and Schliep[112]	https://cran.r-project.org/web/packages/ape/index.html
DAPC (adegenet)	Jombart et al.[113]	https://github.com/thibautjombart/adegenet
PyRAD v3.0.66	Eaton[114]	https://github.com/dereneaton/pyrad
pyrad_find_caps_markers.py	This study	https://github.com/pimbongaerts/radseq
Primer3	Untergasser et al.[115]	https://bioinfo.ut.ee/primer3-0.4.0/
ANGSD v0.913	Korneliussen et al.[116]	https://github.com/ANGSD/angsd
PCAngsd v0.973	Meisner and Albrechtsen[117]	https://github.com/Rosemeis/pcangsd
Msmc2 v1.1.0	Schiffels and Wang[118]	https://github.com/stschiff/msmc2
MITObim v1.9	Hahn et al.[119]	https://github.com/chrishah/MITObim
MITOS	Bernt et al.[120]	http://mitos.bioinf.uni-leipzig.de/index.py
Samtools v1.6 / bcftools v1.9	Li et al.[121]	https://samtools.github.io/
Geneious v11.0.2	Kearse et al.[122]	https://www.geneious.com/
PopArt 1.7	Leigh and Bryant[123]	http://popart.otago.ac.nz/index.shtml
TCS	Clement et al.[124]	http://bioresearch.byu.edu/tcs/
QIIME v1.9	Caporaso et al.[125]	http://qiime.org/
vegan	N/A	https://cran.r-project.org/web/packages/vegan/index.html
PRIMER v7 / PERMANOVA+	Anderson et al.[126]	https://www.primer-e.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be sent to Pim Bongaerts (pbongaerts@calacademy.org).

Materials availability

This study did not generate any new reagents.

Data and code availability

The accession number for the *Pachyseris speciosa* genome assembly reported in this paper is NCBI: PRJNA686157, with the raw sequence data for the genome and transcriptome assemblies available through ENA: PRJEB23386. Raw sequencing data for the reduced-representation (nextRAD) sequencing is available through NCBI: PRJNA701715, and for the whole-genome resequencing through NCBI: PRJNA686482. Gene models, variant call datasets, electronic notebooks, and scripts are accessible through https://github.com/pimbongaerts/pachyseris.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Specimen used for genome sequencing

For the reference genome, sperm from a single colony of *Pachyseris speciosa* was collected at Orpheus Island Research Station (under GBRMPA permit G14/36802.1). The colony was collected on 8 December 2014 from the Island's fringing reef (S18.608°, E146.489°) from 6 m depth, and kept in a flow-through aquarium with 0.5 micron filtered seawater. Every afternoon, one hour before sunset, the colony was placed in a container with as little filtered seawater as possible to cover the whole colony. Broadcast spawning of sperm was first observed on 14 December 2014 at 18:35 (exactly at sunset). The colony's holding water was then centrifuged in 50 mL falcon tubes to concentrate the sperm into a pellet, using an Eppendorf

5702 R centrifuge at 3,000 g for 15 min. The resulting sperm pellet was snap-frozen in liquid nitrogen and stored at −80°C. Additional tissue from the colony was sampled on 15 December 2014, and stored at −80°C for transcriptome sequencing.

Coral specimen collections

For population-level assessments, small fragments from a total of ~2,500 *P. speciosa* colonies were collected (details in Table S1) from the Great Barrier Reef and Coral Sea atolls in Australia, Kimbe Bay in Papua New Guinea, Okinawa in Japan, and Eilat in Israel. Primary collections were performed across three distinct reef habitats: the back-reef (10 m ± 3), shallow slope (10 m ± 3) and deep slope (40 m ± 3), with additional populations collected at intermediate (20 m ± 3) and lower mesophotic depths (60-85 m). Samples were collected using SCUBA or a remotely operated vehicle (ROV; Seabotix vLBV300) between 2012-2017. Additional outgroup samples for phylogenomic analyses were collected from *Pachyseris rugosa* (Great Barrier Reef), *Pachyseris inattesa* (Eilat), and *Leptoseris* (cf.) *glabra* (Great Barrier Reef). Small fragments of the collected specimens were stored in salt-saturated buffer solution (containing 20% DMSO and 0.5 M EDTA) and/or in molecular-grade ethanol, and for a proportion of specimens a skeletal voucher was bleached, rinsed in freshwater and dried. The overall predicted geographic distribution of *P. speciosa* was downloaded from the IUCN Red List of Threatened Species.[38]

METHOD DETAILS

Overall summary of methods

We assembled and annotated a *de novo* reference genome using PacBio sequencing (~95X coverage) of a sperm sample from a *Pachyseris speciosa* colony from Orpheus Island on the Great Barrier Reef. This was used as a reference for reduced-representation sequencing (nextRAD) of *P. speciosa* colonies (n = 501) from shallow and mesophotic habitats in five different regions (Great Barrier Reef, Western Coral Sea, Papua New Guinea, Okinawa and Israel). Whole-genome re-sequencing of representative *P. speciosa* samples (n = 20) was then undertaken for historic demographic modeling at ~5X or ~20X coverage. Given the lack of diagnostic morphological characteristics distinguishing the lineages, we designed a cleaved amplified polymorphic sequence (CAPS) assay to increase the number of genotyped colonies (n = 1,442), and assess ecological, phenotypic and reproductive differences. Morphological differences between lineages were assessed through *in situ* photographs (n = 157 colonies), examination of qualitative traits in bleach-dried skeletons (n = 36 colonies), and micro-skeletal features using scanning electron microscopy (n = 15 colonies). Quantitative measurements were undertaken for five skeletal characters (n = 54 colonies). Physiological characterization was undertaken through protein and lipid quantification, Symbiodiniaceae cell counts, and photopigment quantification using high-performance liquid chromatography (HPLC) (n = 73 colonies). Spawning behavior was assessed through reproductive monitoring of colony fragments (n = 54) during November/December 2017 at the Orpheus Island Research Station. Microbiome characterization was undertaken by genotyping host colonies (n = 43 colonies) from a previous study based on 16S rRNA amplicon sequencing.[37]

Genome assembly and annotation

High-molecular-weight DNA was extracted from sperm through homogenization in liquid nitrogen (based on the method of Blin and Stafford[127]). Initial sequencing was undertaken on the Illumina HiSeq 2500 platform following the PCR-free library construction protocol developed by the Broad Institute (https://software.broadinstitute.org/software/discovar/blog/). For the genome assembly, long-read sequencing was then undertaken on the PacBio Sequel platform at the Ramaciotti Centre for Genomics, across 92 SMRT cells to ~100x coverage. RNA was extracted from tissue collected from the same colony, with directional RNA-seq libraries sequenced on an Illumina HiSeq 2500 at the Australian Genome Research Facility (AGRF). Prior to the assembly, genome size and heterozygosity rate were estimated based on the paired-end short reads. FastQC v0.11.6 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was first applied to screen the base quality, GC-content, overrepresented k-mers, and adaptors. Genome size estimates were calculated using sga.preqc[85] and GenomeScope[86] based on a k-mer size of 31, and were respectively 886.1 and 749.6 Mb (Table S2).

The final genome assembly was achieved in three stages. First, the genome assembly was conducted using CANU v1.5[41] with default settings. This resulted in 10,783 contigs, 1,788 Mb assembled sequences and N50 size of 328 Kb. This is in line with the observation that the assembled size is nearly twice as large as the true genome size when the heterozygosity rate is high in a diploid genome.[128] To reduce allelic redundancy, HaploMerger2[87] was then employed, which successfully merged 44% of the sequences. Finally, PacBio long reads were mapped to the assembly using BLASR[88] and the mean coverage was calculated for each contig. Contigs whose GC contents were greater than 45% (genome GC% = 39%) and read coverages were lower than 50X (~half of the expected coverage) were considered as putative contaminants. These contigs were used as queries to perform blastn search (E-value $\leq 1e^{-05}$) against the NCBI non-redundant nucleotide database (https://www.ncbi.nlm.nih.gov). Contigs that contained more sequences significantly similar to non-metazoan sequences than to metazoan sequences were removed.

De novo identification of repeat classes was accomplished with RepeatModeler v1.0.11 (http://www.repeatmasker.org/RepeatModeler) with parameter "-engine ncbi." The classifier utilized two *de novo* repeat finding programs, RECON and RepeatScout, and was built upon RepBase v20181026.[129] The resulting repeat library was used as input by RepeatMasker v4.0.8 (http://www.repeatmasker.org) to generate the repeat annotation. Protein-coding gene annotation was performed as described Ying et al.:[40] *de novo* and genome-guided transcriptome assembly was performed using Trinity,[89] followed by PSyTrans to remove

Symbiodiniaceae transcripts (https://github.com/sylvainforet/psytrans). Transcripts were then assembled to the genome assembly using PASA,[90] from which a set of likely ORFs were generated. Based on their protein coding ability and completeness, these ORFs were carefully assessed to produce a high confidence and non-redundant training gene set. This was used to train AUGUSTUS[91] and SNAP,[92] and the resulting parameters were employed by the corresponding program from MAKER. The *ab initio* gene models were predicted using the MAKER2[93] pipeline. In addition, putative transposable elements in the gene models were excluded based on transposonPSI (http://transposonpsi.sourceforge.net) and hhblits[94] searches to transposon databases.

Both the genome assembly and gene model datasets were tested for the completeness of conserved core genes using Benchmarking Universal Single-Copy Orthologs (BUSCO v3[95]) under default parameters. The metazoan gene set (odb9), which contains 978 orthologs, was employed as the reference dataset. To further validate the gene models, the predicted protein sequences were matched against the Swiss-Prot database and PFAM-A protein domain database. Swiss-Prot database (2018-08) was downloaded from UniProt FTP site (ftp://ftp.uniprot.org) and blastp was performed (E-value $\leq 1e^{-05}$). The annotated coral proteins were used as queries and the curated database proteins were used as targets. The target coverage was defined as the percentage of the target length in the alignment. To identify well-defined protein domains, HMMER v3[96] was used to perform alignments to PFAM-A v31.0 hmm profile, and those with a combined E-value and c-E-value lower than $1e^{-05}$ were selected.

Reduced-representation sequencing

Coral gDNA extraction was performed using a modified salt-extraction method,[32] reducing Symbiodiniaceae contamination through several centrifugation steps (“separation” method), unless this resulted in insufficient gDNA yield (< 150 ng gDNA) in which case the extraction was performed directly on the sampled tissue (“standard” method; ~25% of sequenced samples). Quality and yield of gDNA were assessed using gel electrophoresis and a Qubit fluorometer to select a subset of higher-quality samples within each sampled population for downstream sequencing (giving preference to those where endosymbiont contamination was reduced; n = 678). This included 3 replicates of the same sperm sample used for the reference genome, 7 additional technical replicates, 9 outgroup samples (*Pachyseris rugosa*, *Pachyseris inattesa*, and *Leptoseris* (cf.) *glabra*) and a Symbiodiniaceae sample isolated from a *P. speciosa* colony using fluorescence-activated cell sorting (as described in Bongaerts et al.[32]). Library preparation was carried out using the nextRAD method (Nextera-fragmented, reductively-amplified DNA; SNPsaurus, LLC), which uses a selective primer sequence (rather than restriction enzymes) to genotype loci consistently between samples. Genomic DNA was purified using AMPure XP beads, and then fragmented and ligated with adaptor sequences using Nextera reagent (Illumina, Inc). Fragmented DNA was then PCR amplified (73°C for 26 cycles) with one of the primers matching the adaptor and extending into the genomic DNA using a 9 bp selective sequence (“GTGTAGAGG”). Libraries were sequenced across 4 NextSeq 500 (Illumina, Inc) lanes using 150 bp single-end chemistry and following the manufacturer’s recommended protocol. Parsing and analyses are detailed in an electronic notebook (https://github.com/pimbongaerts/pachyseris), using generic Python scripts located in a separate repository (https://github.com/pimbongaerts/radseq).

TrimGalore (https://github.com/FelixKrueger/TrimGalore) was used to trim Nextera adapters and low-quality ends (PHRED < 20), while discarding reads shorter than 30 bp. Reads of each sample were then mapped to our *P. speciosa* genome using BWA-MEM.[97] To evaluate levels of contamination across the two extraction methods and the varying gDNA yields, we also independently mapped the reads against the *Cladocopium goreaui* genome[65] (ITS2 type C1) (Figure S1A). Variant calling (for reads mapping to the *P. speciosa* genome) was undertaken using the “UnifiedGenotyper” from the GATK pipeline,[98] and hard-filtered using “VariantFiltration” for bi-allelic single-nucleotide polymorphisms (SNPs) with a minimum coverage of 10X, a minimum genotype quality of 30, and a minimum allele frequency of 0.01. Initially, the overall dataset was reduced to those SNPs that were genotyped for at least 80% of samples and those samples that were genotyped for at least 50% of SNPs (n = 501). Genotyping accuracy was verified from three replicate sperm samples (99.8% similarity), as well as seven regular replicate pairs (separate gDNA extraction and library preparation; 98.9%–99.6% similarity), with only the highest-performing sample of each replicate set retained in the eventual dataset. Duplicate genotypes (clones) were identified through the “vcf_clone_detect” script (https://github.com/pimbongaerts/radseq), using a conservative manual threshold (< 97%) based on the determined genotyping accuracy (from technical replicates) and the distribution of pairwise genetic similarities across all samples (Hamming-based), retaining the sample with the least missing data from each set of duplicates (total of 24 potential clones removed). Given that the opportunistic sampling could have led to accidental resampling of colonies, no interpretations were made regarding clonality rates.

To assess overall genetic structuring, we visualized the structure of the overall SNP dataset (with 468 remaining samples; 8,536 SNPs) using a Principal Component Analysis (PCA) in adegenet,[99] and a neighbor-joining (NJ) tree based on genetic distance (Hamming-based) using the Phylo module in the Biopython[100] library. Both indicated the presence of 4 major clusters in our data (Figures 2 and S1C), which was then further explored using snapclust,[101] a maximum likelihood approach based on the Expectation-Maximization (EM) algorithm that offers goodness-of-fit statistics. We evaluated three different statistics (Akaike, Bayesian and Kullback Information Criteria) using the “choose.k” function under increasing numbers of clusters (up to 20) across five replicate, subsampled datasets (thinned to ensure a minimum distance of 2,5 Kbp between SNPs). We then ran snapclust for the indicated optimum window (k = 4 to k = 6; Figure S1D), again using the same five replicate datasets, using the “Ward” algorithm to define initial group assignments, and with 50 iterations of the EM. At k = 5 and k = 6, two additional geographic populations were separated out from the four most divergent clusters, which was also supported in the NJ tree (Figure 2), and k = 6 was therefore chosen for subsequent analyses. Using the “structure_mp”[32] script (https://github.com/pimbongaerts/radseq) we then also ran STRUCTURE v2.3.4[102] for 20 replicate, subsampled datasets (again thinned to ensure a minimum distance of 2,5 Kbp between SNPs) to further assess potential

signatures of admixture between the 6 clusters. Runs were conducted using the admixture model with correlated allele frequencies and not considering priors (50,000 repeats after a burn-in of 100,000). Individual runs were then aligned using CLUMPP v1.1.2,[103] and assessed for the presence of "ghost clusters" (i.e., clusters with no fully assigned samples), only retaining those runs that have a maximum ancestry assignment of at least 0.99 across all clusters. Samples were then separated out to clusters using a "lenient" ($\geq$0.8) and "stringent" ($\geq$0.95) mean ancestry assignment cut-off (marking samples below those thresholds as "unassigned"). The assignment was compared to clustering in the NJ tree by coloring branch tips according to their "lenient" STRUCTURE assignments (Figure 2). Potential admixed samples were identified by extracting those with a maximum ancestry assignment of < 0.95 across clusters. The potential of one specific sample to represent an F1 hybrid was assessed by quantifying heterozygosity for SNPs that were established to be alternatively fixed for the two clusters the sample was assigned to.

To assess the extent and nature of divergence between the three lineages, a reduced SNP dataset was generated with only "stringently" assigned samples, grouped by cluster and geographic region. Pairwise genome-wide, mean F_{ST} values[130] were calculated between all groups (45 pairwise comparisons) to assess overall divergence between lineages versus geographic regions using vcftools.[104] Highly divergent SNPs were extracted by calculating pairwise allele frequency differentials (AFDs) between the three Australian lineages (using the Great Barrier Reef and Coral Sea regions as "replicates"), and identifying those SNPs that across both regions were alternatively fixed (AFD of $\geq$ 0.95 to allow for some genotyping error). In addition, divergent SNPs were identified through pairwise comparisons of lineages (grouping the Great Barrier Reef and Coral Sea together) using the first principal component (after verifying it only separates samples based on lineage) in pcadapt v4,[105] and based on q-values with an expected false discovery rate lower than 5%. Divergent SNPs were also identified using both methods for the Red Sea lineage (Israel). Genetic variants were annotated and functional effects predicted using SnpEff,[106] with divergent SNPs manually assessed based on their UniProt IDs. For visualization purposes only, we used RaGOO[107] to assemble our 2,368 genomic scaffolds into pseudomolecules by mapping them to a chromosome-level assembly of *Acropora millepora*,[79] under the expectation that broad-scale synteny is expected to be conserved (e.g., as demonstrated for *Acropora digitifera* and *Nematostella*).[131] We obtained Gene Ontology (GO) terms for all SNPs using the UniProt portal (https://www.uniprot.org), and then used Gowinda[108] to conduct GO enrichment analyses. Gowinda was run using the original genomic scaffolds in the "gene" mode (assuming complete linkage disequilibrium between SNPs within a gene) for the alternatively fixed and pcadapt outlier SNP sets using 1,000,000 simulations and a 1,000 bp window upstream and downstream. This window size represents a commonly used cut-off, to conservatively consider SNPs within limited distance outside of gene introns/exons. Assessments of these SNPs for gene ontology (GO) enrichment did not identify significantly enriched GO terms (although admittedly only a small proportion of the overall genome was assessed through the reduced representation sequencing).

Phylogenetic relationships between the lineages were explored through a separate dataset where we selected 6 samples for each of the 6 uncovered *P. speciosa* lineages (selecting those with the highest number of genotyped SNPs), as well as 3 *P. rugosa* samples, 3 *P. inattesa* samples, 3 *Leptoseris* (cf.) *glabra* samples, and 3 *Agaricia fragilis* samples (from Bongaerts et al.[32]). We again used TrimGalore to trim Nextera adapters and low-quality ends (PHRED < 20), while discarding reads shorter than 30 bp. These samples (n = 48) were then mapped to our *P. speciosa* reference genome using ipyrad v0.9.62[109] under default parameters. The resulting loci were concatenated to generate maximum likelihood trees with RAxML,[110] using the "GTRGAMMA" substation model and automatic bootstrapping ("autoMRE"), which resulted in 400 bootstraps and 1,731,480 distinct alignment patterns. We also conducted species tree inference based on SNPs using the SVDQuartets[132] algorithm as implemented in tetrad,[109] using full quartet sampling and SNP subsampling to ensure a minimum distance of 2,5 Kbp. Consensus trees and variation over the bootstrap replicates were visualized using the ggtree[111] and ape[112] packages.

Genetic structuring within each of the three Australasian lineages was assessed by splitting the dataset based on the "lenient" assignment cut-off (0.8), and removing samples below that threshold. These three datasets were filtered for SNPs genotyped for at least 80% of samples (within each lineage), and with a maximum observed heterozygosity threshold of 0.5 (to filter out potential paralogs). For each lineage, we also created a dataset that was more representative of neutral genomic diversity by removing SNPs that were identified as outliers using pcadapt v4[105] based on q-values with an expected false discovery rate lower than 10%. We used pcadapt as it is generally robust under hierarchical population genetic structure and does not require *a priori* population information, which is important given our nested sampling (region, location and habitat) and low population sizes at the deepest level (habitat) associated with the splitting of the dataset (into three cryptic lineages). For both "neutral" and "overall" datasets, we then assessed overall genetic structuring using principal component analysis (PCA in adegenet[99]), and evaluated structuring across habitats, locations, and regions using discriminant analysis of principal components (DAPC[113]).

CAPS marker development and genotyping

In order to develop a rapid and cost-effective diagnostic assay for three cryptic Australasian lineages, we screened the nextRAD sequence loci for lineage-diagnostic mutations in the sequence motifs of commonly available restriction enzymes. As this was undertaken prior to the establishment of the reference genome, clustering and variant calling of the nextRAD data was first analyzed *de novo* using PyRAD v3.0.66[114] using a clustering threshold of 88%, a minimum sequence coverage of 6, and a maximum of 4 sites with a PHRED quality below 20. Sites that could be targeted using cleaved amplified polymorphic sequence (CAPS) markers were then identified using the pyrad_find_caps_markers.py script (https://github.com/pimbongaerts/radseq). An initial host genome assembly and realigned whole-genome resequencing data was then used to manually align potential target loci, so that primers could be designed to flank the loci (using Primer3[115]), and the presence of additional adjacent restriction sites could be evaluated. After

initial amplification, digestion and reproducibility screening for 20 different primer pairs, three markers were selected (named "Pspe-Green-CfoI/HhaI," "Pspe-Blue-HaeIII," and "Pspe-Red-Taqα1") and further tested on samples with a known lineage assignment (based on nextRAD) data to confirm genotyping reliability (3 out of 120 samples showed a different assignment compared to that based on the RAD-seq data).

For the amplications, we used 0.5–1.0 μl of DNA, 1 μl 10x PCR buffer (Invitrogen), 0.3 μl 50 mM MgCl2, 0.2 μl 10 mM dNTPs, 0.5 μl for both the forward and reverse primer (10 μM), 0.07 μl of Platinum *Taq* DNA Polymerase (Invitrogen) and dH_20 water to a total volume of 10 μl per reaction. The cycling protocol was: 1 × 94°C (4 min); 31 × [1 min at 94°C, 1 min at 61°C, 1 min at 72°C]; 1 × 72°C (6 min). The restriction digest mix contained 1 μL enzyme buffer (10X), 0.1 or 0.05 μL restriction enzyme depending on the concentration (10.000 u/mL or 20.000 U/mL), 0.5 μL PCR product, and dH_20 water to a total volume of 10 μl per digest. The restriction digest was run for 1 hour at 37°C (for CfoI, HhaI, or HaeIII) or 65°C (for TaqαI), and visualized by running 4 μl of digest product on a 3% agarose gel with GelRed stain (120V for 40 min) immediately afterward. Field assays were performed using the miniPCR thermal cycler and blueGel visualization system (Amplyus, Cambridge, MA, USA), using agarose gels with TBE buffer. Overall, 1,119 samples were genotyped using the assay (1-3 CAPS markers per sample), including those for the physiological, microbial, and reproductive assessments in this study, and together with the RAD-seq data this led to a total of 1,442 genotyped samples across regions.

Whole-genome re-sequencing

Representative samples (n = 20) of the three distinct lineages (as identified by the nextRAD sequencing) were selected for whole-genome re-sequencing. These samples originated from back-reef habitats (10 m) in three regions on the Great Barrier Reef (Central GBR: Myrmidon Reef, Northern GBR: Ribbon Reef 10, and Far Northern GBR: Great Detached Reef). Genomic DNA was purified using AMPure XP beads (1.8:1 beads/DNA ratio), with separated barcoded libraries prepared for each individual using Illumina's Nextera kit following the manufacturer's recommended protocol. Pooled libraries were sequenced on four Illumina Hiseq 2500 (Illumina, Inc) lanes using 100 bp paired-end chemistry resulting in an average coverage of ~5X. Four additional lanes of sequencing were conducted for ten representative samples (representing the three different lineages from both Myrmidon Reef and Great Detached Reef) to obtain higher coverage (~20X) for demographic analyses. All commands used to analyze whole genome resequencing data and perform demographic analyses are detailed in an electronic notebook (https://github.com/iracooke/pachyseris_wgs).

Raw reads were first pre-processed and mapped against the *P. speciosa* genome based on GATK best practices.[115] Adapters were marked using Picard,[115] mapping was performed using BWA-MEM;[97] marking shorter split hits as secondary but with all other parameters at their defaults, and PCR duplicates were marked using Picard. Since low coverage samples had insufficient depth to reliably call genotypes, we used a genotype likelihood approach to investigate genetic structuring and assess admixture at the whole genome level. To support this approach ANGSD v0.913[116] was used to call SNPs and calculate genotype likelihoods. Variant sites (SNPs) were retained if they had a p value less than $1e^{-06}$ (GATK probability model), and a minor allele frequency greater than 5%. Genotype likelihoods called using ANGSD were used to explore genetic structuring and admixture using PCAngsd v0.973.[117]

Demographic histories for each of the ten deeply sequenced colonies were inferred based on the distribution of heterozygous sites using the PSMC' method[59] implemented in msmc2 v1.1.0.[118] This method was used in favor of approaches based on allele frequency spectra due to the substantial substructure encountered within our focal lineages. In order to avoid known inaccuracies due to fragmentation[133] or miscalled genotypes the analysis was restricted to scaffolds larger than 1Mb (a total of 390Mb) and genomic regions with repeats, excessively low or high read coverage were excluded. Demographic histories for each colony were inferred by performing 100 bootstrap replicates and taking the average. Bootstrap data was generated by randomly sampling the genome in 0.5Mb chunks and arranging these into 30 scaffolds of length 20Mb per replicate. Since this approach allows inference of historical effective population sizes for each sequenced coral colony,[59] it was possible to assess the consistency of these estimates within lineages. In order to translate msmc results into real timescales and effective population sizes the spontaneous mutation rate, μ and generation time, g are required. In the absence of data required to independently calculate these parameters we used estimates (μ = 4.83e-8; g = 35y) recently published for *Orbicella*[44] as these are the closest available in a phylogenetic sense. It should be noted that there is considerable uncertainty in these estimates which affects the timescale and effective population sizes, however the shape of curves shown in this figure are unaffected by changes in these parameters.

A reference mitochondrial genome for *P. speciosa* was assembled using deeply sequenced data from a single colony from the "red" lineage, sampled at Myrmidon back reef (Central GBR). Mitochondrial reads were extracted from whole genome data, assembled and scaffolded into a single contig of length 19,507 bp using MITObim v1.9[119] with the complete mitochondrial genome of *Acropora digitifera* (GenBank: NC_022830) as a bait. Manual inspection revealed overlapping sequence at both ends indicative of a circular sequence. After trimming redundant bases a circular genome of length 19,007 bp with a single gap of length 30 bp was produced. Annotation of this genome was performed with MITOS.[120]

Consensus mitochondrial sequences for each of the whole-genome sequenced colonies were generated by mapping raw reads to the reference mitochondrial genome. Mapping was performed with BWA-MEM[97] and resulted in a high coverage (> 3000x) bam file for each colony. Consensus sequences (n = 20) were then called using samtools v1.6 and bcftools v1.9[121] These were imported into Geneious v11.0.2[122] as an alignment and trimmed to remove a 468 bp region where manual inspection of read coverage suggested a potential misassembly or unresolved repeat. The trimmed alignment was exported to nexus format and visualized with PopArt v1.7[123] as a TCS[124] network.

Characterization of microbial communities

Basic characterization of associated Symbiodiniaceae was done by extracting "contaminant" nextRAD loci matching the plastid genome of *Cladocopium goreaui*[65] (ITS2 type C1). For this, all loci from the *de novo* analyzed dataset (constructed with PyRAD for the CAPS development) were mapped against the *C. goreaui* plastid and mitochondrial genomes using BWA-MEM,[97] retaining only loci that were genotyped for at least 100 samples. The loci were visually assessed, and the two most informative nextRAD loci were extracted (both matched a single plastid genome scaffold) and evaluated. We also mapped the Nextera whole-genome re-sequencing to the *Cladocopium goreaui* mitochondrial genome using BWA-MEM,[97] then called the consensus sequence for a 7 kb high-coverage region of 16 *P. speciosa* samples using samtools v1.6,[121] and assessed haplotypes using PopArt v1.7.[123]

Microbiome characterization was carried out in Hernandez-Agreda et al.,[37] and we now determined the host genotypes for 43 of those samples and reanalyzed the data for lineage-associated patterns. DNA was extracted using the modified protocol[134] of MoBio PowerPlant Pro DNA isolation kit, with the bacterial 16S rRNA gene PCR amplified using the 27F/519R (v1-v3 region) primers.[37] Library preparation was carried out using the Illumina TruSeq DNA library protocol and sequenced on the Illumina MiSeq (300 bp paired-end). Sequence data were analyzed using Quantitative Insights Into Microbial Ecology (QIIME v1.9[125]). After discarding low-quality sequences (ambiguous base calls, sequences < 200 bp and homopolymers > 6 bp), barcodes, primers, and chimeras were removed . Operational Taxonomic Units were defined and identified with clustering at 97% similarity and RDP classifier[135] and GreenGenes[136] database. After removing chloroplasts, mitochondria, unidentified, and unassigned OTUs, OTU tables were normalized using fourth root transformation and standardized by total by sample. A second OTU table was generated by transforming data into presence/absence. Core microbiome (100%) was identified for each genotype using QIIME script 'compute_core_microbiome.py'.

Phenotypic characterization

Morphological and physiological characterization was undertaken for a subset of samples from the Western Coral Sea. First, gross morphological appearance was visually assessed for colonies with *in situ* photographs of genotyped colonies (n = 157; Figure S3A). Bleach-dried skeletons were examined under a stereomicroscope to assess the presence of discriminating, qualitative skeletal characteristics (n = 36; Table S3). Quantitative measurements undertaken for five skeletal characters were measured for a subset of skeletons (n = 54-89): septa per 5 mm, diameter secondary corallite, ridge height, valley width between ridge bottoms, valley width between ridge tops (characters adapted from Budd et al.[137] and Terraneo et al.[138]). Measurements were conducted in triplicate, except for the diameter of secondary corallite which was dependent on the number of secondary corallites present. Additionally, a small number of representative fragments was then visually assessed using scanning electron microscopy (SEM) to screen for obvious discriminating characters (n = 15; Figure S3B).

Physiological characterization was undertaken for a subset of *P. speciosa* samples (n = 73) that were collected from 10, 20, 40 and 60 m depth at three different Osprey sites, and were snap-frozen in liquid nitrogen and subsequently stored at −20°C. Coral tissue was removed using an air-brush and 10 mL filtered phosphate buffer. Following centrifugation of the homogenate, the supernatant was frozen (−80°C) for host protein analysis. The pellet was resuspended in 3 mL filtered phosphate buffer and evenly separated into three aliquots stored at −80°C until further processing for symbiont cell count and pigment quantification with high-performance liquid chromatography (HPLC), keeping the third sample as a back-up. Symbiodiniaceae density (per cm^2) was determined with a Neubauer Improved Bright-Line haemocytometer and a Olympus BX43 light microscope with 6 replicate counts of diluted samples.[139,140] The Symbiodiniaceae counts were normalized against the surface area of the sample skeleton, which was determined by weighing the coral skeleton three times before and after wax dipping. A calibration curve of objects with known surface area was generated to calculate the surface area using the below equation. The Whitaker and Granum[141] methodology was used to obtain the water-soluble protein content (mg cm^{-2}) from the host tissue suspension from the air-brushing procedure and standardized to the surface area (cm^2). The total lipid concentration (mg cm^{-2}) was determined using a modified method of Folch et al.[142] as described by Dunn et al.[143] Frozen coral fragments of 2-3 cm^2 were used and resulting lipid content was standardized to the surface area (cm^2) of each fragment. Pigment concentration in Symbiodiniaceae cells were determined by high-performance liquid chromatography (HPLC). Pigment concentrations (chlorophyll α, chlorophyll c2, peridin, fucoxanthin, diatoxanthin (Dtx), and diadinoxanthin (Ddx)) were measured with a HPLC using the method as described by Zapata et al.[144] and Dove et al.[145] and normalized to surface area (cm^2), Symbiodiniaceae cell count (pg $cell^{-1}$), and chlorophyll α (Chl α^{-1}). Lastly, the total xanthophyll pool per chlorophyll α ((Dtx + Ddx) Chl α^{-1}) was calculated.

Reproductive characterization

Reproductive behavior was assessed at the Orpheus Island Research Station, where 54 large colony fragments were collected from two nearby reef sites between 4-6 November, 2017. Coral colonies were kept in a raceway with flow-through, filtered seawater, constant air bubbling, and a shading canvas to mimic a shaded environment. Colonies were genotyped as described above in the CAPS genotyping section (with a subset of colonies genotyped in the field, and all colonies eventually genotyped in the lab). Between 7-13 November, 2017 colonies were isolated into individual containers and monitored for spawning from 45 minutes before to 2 hours after sunset. Given the predicted split-spawning, 35 of the originally collected coral colonies were kept in a single raceway until the December spawning. Between 7-12 December, 2017, colonies were again isolated into individual containers and monitored for spawning from 30-45 minutes before to 1.5 hours after sunset. Spawning was

defined as the vigorous release of gametes (no ''dribbling'' of either sperm or eggs was observed). All colonies were returned to the reef after spawning observations as per permitting requirements.

QUANTIFICATION AND STATISTICAL ANALYSIS

Ecological differentiation

To assess ecological distributions in the Australasian region (n = 1,312; only considering populations with at least 7 samples), we used PERMANOVA as implemented in the adonis function of the R vegan package (https://cran.r-project.org/web/packages/vegan/index.html) based on Bray-Curtis dissimilarities. We tested for differences in relative proportions of the three lineages between habitats and regions (considering only the 10 m and 40 m habitats as those were sampled across all three Australasian regions) using location as strata, followed by pairwise testing of all habitats within each region (with *p-value*s adjusted using Bonferroni correction for multiple comparisons). To assess differences in the relative abundances across habitats (for each lineage separately), we used one-way ANOVA (Type II) testing for the GBR and WCS regions, followed by pairwise testing (Tukey's test and *p-value*s corrected with single-step method) between habitats. Assumptions of normality and homoscedasticity were fulfilled (a square root transformation was applied to the relative abundances of the ''green'' lineage).

Phenotypic differentiation

Statistical differences in quantitative morphological and physiological traits were assessed collectively using a permutational multivariate analysis of variance (PERMANOVA+[126]), and individually with one-way ANOVA. Three multivariate matrices were generated to evaluate host and symbiont characteristics collectively: morphological traits (septa (per 5mm), width between ridge tops, valley width, ridge height), host (protein and lipids (mg cm^{-2})) and symbiont physiological traits (Chlorophyll α and c_2 (Chl α^{-1}), Chlorophyll α (pg $cell^{-1}$), Peridin and Fucoxanthin (Chl α^{-1}), Xanthophyll pool (Chl α^{-1}) and Symbiodiniaceae (cm^{-2})). Collinearity was evaluated in normalized matrices before the analyses. Differences between lineages and habitats were evaluated on Euclidean distance matrices using Type III sum of squares and 9,999 permutations. The factor ''site'' was not considered because adequate replication lacked to test it comprehensively (samples for physiology were collected prior to discovering the cryptic diversity), however the three sites were located within 10km of each other. A principal component analysis was performed on the multivariate matrices, and only considering samples with no missing data. Univariate analyses were carried out to compare lineages within each depth/habitat (considering those with n > 3) with one-way ANOVA (type II). Non-normal and heteroscedastic data was transformed, unless assumptions were still not met (protein data). Here the ANOVA analysis was still used as it is robust against these violations.[146] Significant results were further examined with a Tukey's test (*p-value*s were adjusted using single-step method correction for multiple comparisons).

Microbial community structure

Differences in bacterial communities between genotypes, habitats, and regions were evaluated with a permutational multivariate analysis of variance (PERMANOVA+,[126] Type III sum of squares and 9,999 permutations) on Bray-Curtis (structure), Sorensen (composition) and Euclidian (richness, diversity, Delta+ and Lambda+) distances. Chloroplasts, mitochondria, unidentified, and unassigned were removed, and OTU tables were normalized using fourth root transformation and standardized by total by sample. A second OTU table was generated by transforming data into presence/absence.