Molecular Phylogenetics and Evolution 111 (2017) 132–148

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Coral hybridization or phenotypic variation? Genomic data reveal gene flow between *Porites lobata* and *P. Compressa*

Z.H. Forsman [a,*], I.S.S. Knapp [a], K. Tisthammer [b], D.A.R. Eaton [c], M. Belcaid [a], R.J. Toonen [a]

[a] Hawaii Institute of Marine Biology, Kāneʻohe, Hawaiʻi, United States
[b] Kewalo Marine Laboratory, Honolulu, Hawaiʻi, United States
[c] Yale University, Department of Ecology and Evolutionary Biology, New Haven, Connecticut, United States

ARTICLE INFO

Article history:
Received 12 November 2016
Revised 26 March 2017
Accepted 26 March 2017
Available online 31 March 2017

Keywords:
Cnidarians
Bioinformatics/phyloinformatics
Polymorphism
Hybridization
Invertebrates
Metagenomics
Systematics

ABSTRACT

Major gaps remain in our understanding of the ecology, evolution, biodiversity, biogeography, extinction risk, and adaptive potential of reef building corals. One of the central challenges remains that there are few informative genetic markers for studying boundaries between species, and variation within species. Reduced representation sequencing approaches, such as RADseq (Restriction site Associated DNA sequencing) have great potential for resolving such relationships. However, it is necessary to identify loci in order to make inferences for endosymbiotic organisms such as corals. Here, we examined twenty-one coral holobiont ezRAD libraries from Hawaiʻi, focusing on *P. lobata* and *P. compressa*, two species with contrasting morphology and habitat preference that previous studies have not resolved.

We used a combination of *de novo* assembly and reference mapping approaches to identify and compare loci: we used reference mapping to extract and compare nearly complete mitochondrial genomes, ribosomal arrays, and histone genes. We used *de novo* clustering and phylogenomic methods to compare the complete holobiont data set with coral and symbiont subsets that map to transcriptomic data. In addition, we used reference assemblies to examine genetic structure from SNPs (Single Nucleotide Polymorphisms). All approaches resolved outgroup taxa but failed to resolve *P. lobata* and *P. compressa* as distinct, with mito-nuclear discordance and shared mitochondrial haplotypes within the species complex. The holobiont and 'coral transcriptomic' datasets were highly concordant, revealing stronger genetic structure between sites than between coral morphospecies. These results suggest that either branching morphology is a polymorphic trait, or that these species frequently hybridize. This study provides examples of several approaches to acquire, identify, and compare loci across metagenomic samples such as the coral holobiont while providing insights into the nature of coral variability.

1. Introduction

Many lineages of the Tree of Life are cryptic and challenging to differentiate, requiring informative molecular markers for determining species boundaries, species ranges, evaluation of extinction risk, and for determining appropriate conservation actions (Purvis et al., 2000; Purvis, 2008). Basal metazoan branches in particular are deeply divergent lineages with few available genomic resources or orthologous genetic markers that are sufficiently conserved to be reliably amplified, sequenced, and aligned across taxa (Voolstra et al., 2017). Mitochondrial or ribosomal markers are the best studied and most widely used, however mitochondrial loci in several basal metazoans including cnidarians evolve 10–20 times slower than vertebrate mitochondria (Hellberg, 2006; Huang et al., 2008). With the exception of a few notable examples, coral mitochondrial markers are generally uninformative below the genus level (Flot et al., 2008; Eytan et al., 2009; Schmidt-Roach et al., 2012; Forsman et al., 2013; Keshavmurthy et al., 2013; Luck et al., 2013; Pinzón et al., 2013). The most widely used alternatives to mitochondrial markers are nuclear ribosomal genes and transcribed spacers; however, the evolution of these multicopy regions is complex, resulting in divergent paralogous variants co-occurring within individual genomes (Odorico and Miller, 1997; van Oppen et al., 2000; Vollmer and Palumbi, 2004; Forsman et al., 2006; Stat et al., 2012). Additional markers are challenging to acquire from non-model organisms, and therefore have rarely been used for systematic, phylogenetic, or phylogeographic studies.

Next-generation sequencing has resulted in the rapid proliferation of genomic data for non-model organisms (Andrews and

* Corresponding author.
E-mail address: zac@hawaii.edu (Z.H. Forsman).

http://dx.doi.org/10.1016/j.ympev.2017.03.023
1055-7903/

Luikart, 2014; Puritz et al., 2014a,b). In particular, Restriction site-Associated DNA sequencing (RADseq) is a cost effective, reduced genomic representation approach enabling the comparison of thousands of loci, thereby providing new insights into challenging problems, such as: recent adaptive radiations (Rundell and Price, 2009; Wagner et al., 2012), phylogenetic relationships over deep evolutionary timescales (Rubin et al., 2012; Cariou et al., 2013; Hipp et al., 2014), and hybridization and species boundaries (Wagner et al., 2012; Hipp et al., 2014; Takahashi et al., 2014; Herrera and Shank, 2015). RADseq data typically results in many thousands of short (~30–500 bp) sequences adjacent to restriction enzyme cut sites and there are a wide variety of RAD protocols suitable to various research questions (Davey et al., 2012; Cariou et al., 2013; Reitzel et al., 2013; Puritz et al., 2014b; Andrews et al., 2016). Alignment of these short reads results in 'stacks' of RAD loci providing 'vertical' depth of coverage for a particular locus necessary to distinguish biological heterozygosity from PCR and sequencing errors. These short sequences can make the identification and clustering of loci problematic in the absence of a reference genome, therefore some RAD protocols generate longer alignments of sequence reads that are slightly staggered (either due to random DNA shearing or incomplete digestion), effectively trading the depth of vertical coverage at some loci for greater horizontal coverage across fragments (Etter et al., 2011; Toonen et al., 2013; Andrews and Luikart, 2014; Puritz et al., 2014b). Increasing 'horizontal' coverage affords more confident identification of the locus, which is particularly advantageous when metagenomic samples are examined, such as for the coral holobiont -- a close association of a diverse variety of organisms including the coral animal, obligate symbiotic dinoflagellate algae, bacteria, fungi, microbes, and gut contents.

Corals are among the most recalcitrant groups of organisms to study in part because of the complexity of the holobiont, but in addition they are phenotypically plastic, long-lived colonial animals with overlapping generation times that occupy a wide range of depths and habitats over an enormous geographic range. These factors likely contribute to high levels of genetic and morphological variation, which in combination with a lack of informative molecular markers, has constrained our understanding of the foundational species of an increasingly threatened ecosystem. Genetic markers with higher resolving power are urgently needed for improved study of population genetics, biogeography, species boundaries, taxonomy, and evolutionary history (Stat et al., 2012; Herrera and Shank, 2015), as well as for understanding geographic distributions, which are fundamental for determining if species are endemic, rare, or threatened with extinction (Brainard et al., 2011).

For this study we focus on the coral genus *Porites* (Link, 1807), which has long been a prime example of 'the species problem' due to notoriously difficult species identification and confusing patterns of morphological variation (Vaughan, 1907; Brakel, 1977). Species boundaries within the genus remain poorly understood and are the subject of ongoing debate with several unresolved species complexes (Brakel, 1977; Jameson, 1997; Forsman et al., 2009; Jameson and Cairns, 2012; Prada et al., 2014). *Porites compressa* (Dana 1846) is a branching coral that dominates in shallow sheltered lagoons, while *P. lobata* (Dana 1846) forms mounds or encrusts, and dominates in reefs exposed to higher wave energy (Storlazzi et al., 2004). However, the two species can also be found together in intermediate habitats, and their spawning times are variable and overlap (Richmond and Hunter, 1990). *P. lobata* and *P. compressa* can appear strikingly different when growing side by side in the same habitat (Fig. 1), which seems to rule out phenotypic plasticity as a primary explanation for the morphological variation. Vaughan (1907) recognized and named a variety of forms and subforms of *P. lobata* and *P. compressa* in Hawai'i (Fig. 2), indicating that there are many intermediates between the branching and mounding extremes, consistent with hybridization, or a single polymorphic species. The two species have do not have entirely overlapping geographic ranges, which may be more consistent with hybridization than a single polymorphic species: *Porites lobata* occurs in the Eastern Pacific, but branching morphotypes (*P. compressa* and *P. cylindrica*) do not (Veron and Stafford-Smith, 2000).

Despite striking colony-level morphological differences, microskeletal features appear highly similar although there are as yet no detailed studies of skeletal landmarks (e.g. Jameson, 1997; Forsman et al., 2015) between the two species. Previous genetic work using ribosomal and mitochondrial markers have failed to distinguish named species in the '*P. lobata* species complex', which includes *P. lobata*, *P. compressa*, *P. cylindrica*, *P. duerdeni*, *P. solida*, and *P. annae* (Forsman et al., 2009). The closest sister species to the '*P. lobata* species complex' is the mounding coral *P. evermanni*, which in Hawai'i can clearly be differentiated by microskeletal landmarks and ribosomal and mitochondrial markers (Forsman et al., 2009, 2015), however a recent study using coalescent analysis of multiple genetic makers has suggested that *P. evermanni* may interbreed with *P. lobata* in the Eastern Pacific but not in Hawai'i (Hellberg et al., 2016). The lack of resolution into the *P. lobata* species complex and the apparent geographic variation in permeability between the *P. lobata* complex and *P. evermanni* could be simply an artefact of limited resolution afforded by a few noisy molecular markers, or alternatively these patterns could be providing new insights into the nature of variation within coral species and permeability of boundaries between species. Our goal in this study was apply more powerful genomic methods of identifying and comparing coral loci to provide further insights into the nature of variation within and between these coral species.

We used RADseq data to more closely examine relationships between *Porites* morphospecies in Hawai'i, focusing on *P. lobata*, *P. compressa*, and *P. evermanni* relative to outgroup taxa. We examined RAD libraries from twenty-one *Porites* samples representing six morphologically defined species with the broad goal of resolving species relationships. In addition to examining the entire holobiont dataset, we used several strategies to parse and more confidently identify coral loci; (1) we used reference mapping against previously published complete coral mitochondrial genomes to acquire and compare nearly complete mitochondrial genomes from each library; (2) we binned the data into reads that map to previously published coral and *Symbiodinium* transcriptomic data sets for *de novo* assembly and comparison to the complete holobiont dataset; (3) we mapped reads to transcripomic reference sequences to examine genetic structure from Single Nucleotide Polymorphisms (SNPs); (4) we used *de novo* assemblies and the BLAST suite of tools (Altschul et al., 1997; McGinnis and Madden, 2004) to further characterize loci.

2. Methods

2.1. Sample collection

Porites lobata samples were collected from both windward and leeward coasts of O'ahu, Hawai'i (Fig. 2, Table 1). *Porites compressa* was collected from two locations on the windward coast; from patch reefs within Kāne'ohe Bay where *P. lobata* does not co-occur, and from reefs off Lanikai Beach, where *P. lobata* and *P. compressa* co-occur (Table 1). All *P. compressa* colonies had clear and distinct branches; only easily identifiable colonies were selected avoiding intermediate morphologies. Samples were collected under the State of Hawai'i Special Activity Permit (SAP2013 and SAP2013-26). Additional taxa were selected as outgroups (Fig. 1), based on previous systematic work (Forsman et al., 2009). These

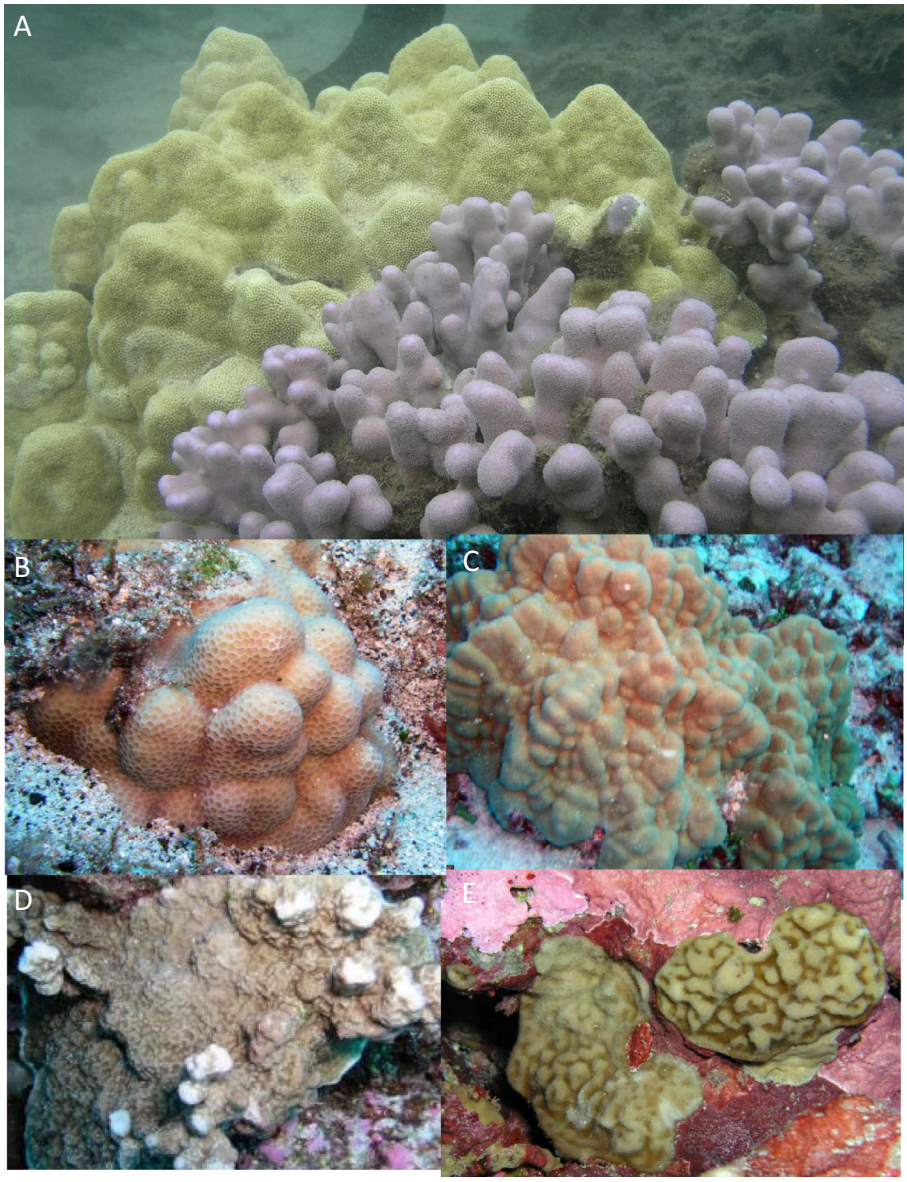

Fig. 1. Example *in-situ* photographs of *Porites* species sampled for this study. (A) *Porites lobata* (yellow colony) adjacent to a *P. compressa* colony (purple colony). (B) *P. c.f. brighami* (C) *P. evermanni* (D) *P. rus* (E) *P. superfusa*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outgroup samples were identified in the field by the regional taxonomic expert Dr. James E Maragos. Sample *P. cf. brighami* appeared to be similar to *P. brighami* in the field however upon closer examination under the microscope it appeared within the range of variation of *P. lobata* (ZHF personal opinion). All extracted samples were either stored in salt saturated DMSO buffer (Gaither et al., 2011), >95% ethanol, collected fresh, or flash frozen with liquid nitrogen on site and stored at −80 °C. Samples PLob1, Plob02, PLob3, PCom1, PCom2, and PCom3 were extracted and sequenced as described previously (Toonen et al., 2013).

2.2. DNA extraction and quantification

Fresh coral samples consistently yielded relatively low quality DNA, presumably because of excess mucus, and so these samples were stored in DMSO for at least 24 h, which ameliorated this issue. DMSO yielded higher molecular weight DNA than ethanol as with previous studies (Gaither et al., 2011). Surgical bone cutters were used to fragment the corals and to remove excess calcium carbonate skeleton. The resulting fragments were between 0.3 cm^3 and 0.5 cm^3 and consisted of mostly the top tissue layer. The fragment was placed on a clean piece of Kimwipe paper for 5–15 min to remove residual buffer. The sample was then crushed in aluminium foil, and placed into a 1.5 ml tube with ATL or TL lysis buffer and proteinase K for ~3 h.

All samples were extracted using the Qiagen (DNeasy® Blood & Tissue) and Omega DNA extraction kits, with 4 µl of 20 mg/ml RNase A added after tissue lysis. The elution step was modified since the largest proportion of high molecular weight DNA can vary from sample to sample. Instead of the recommended 1 × 200 µl or 2 × 100 µl elutions, we used multiple small volume elutions. The first elution was 35 µl and typically yielded low molecular weight DNA, although occasionally this first elution yielded high molecular weight at a concentration suitable for digestion. The second elution was 50 µl, and if necessary a final round of 2 × 50 µl elutions (100 µl total) was performed. Heated (70 °C) HPLC grade water was used for all elutions instead of the supplied elution buffer. All extractions were inspected on a 1% agarose gel. Samples were

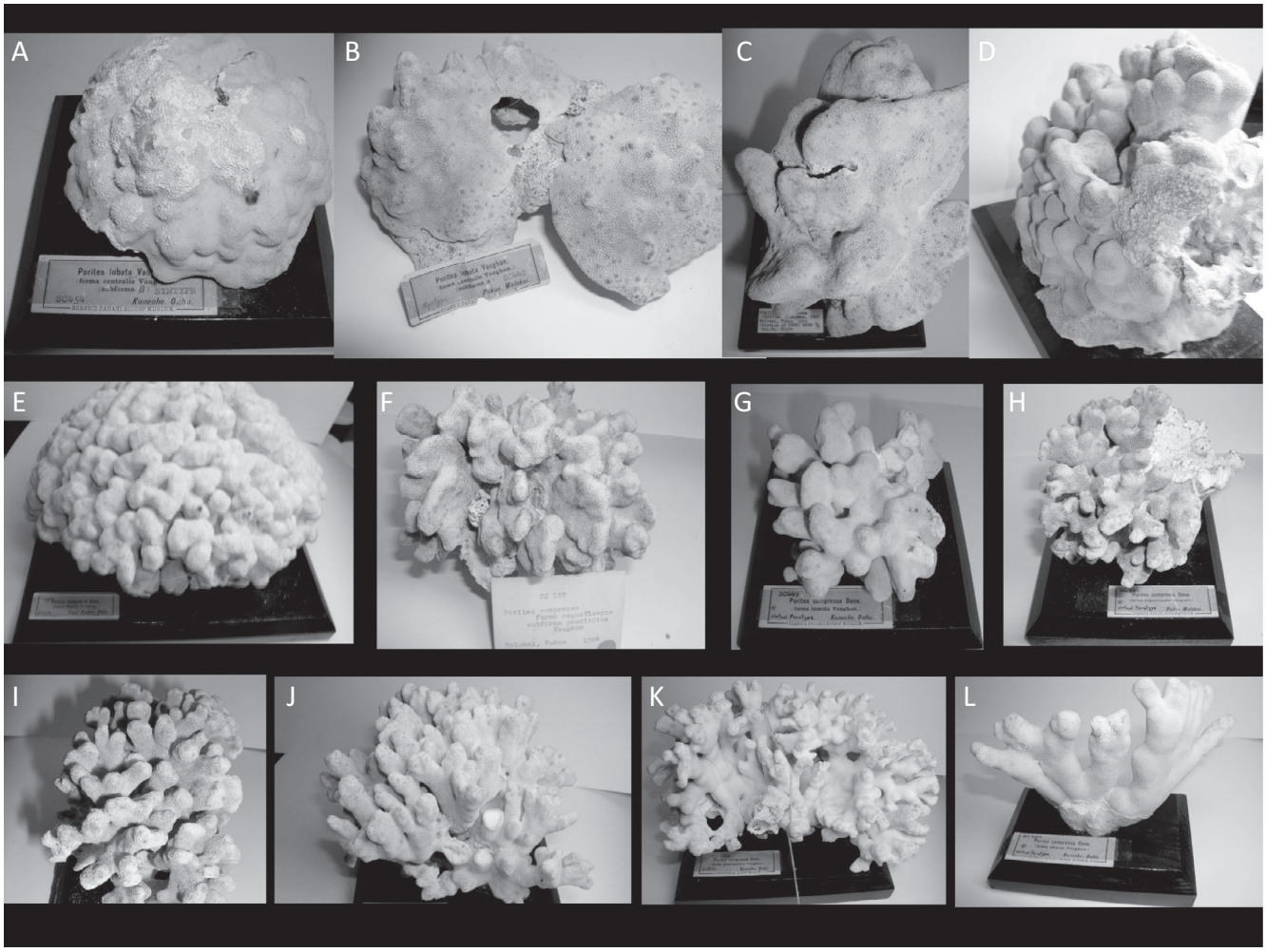

Fig. 2. Type specimens of various forms of *Porites lobata* and *P. compressa*. (A) *P. lobata* (forma centralis, subforma β), Vaughan, 1907, Kāneʻohe, Oʻahu, Syntype SC454; (B) *P. lobata* (forma centralis subforma alpha), Vaughan, 1907, Pukoo, Molokaʻi, Syntype SC442; (C) *P. lobata* (forma parvicalyx), Vaughan, 1907, Molokaʻi, Pukoo, portion of type SC574; (D) *P. lobata* Dana (forma infundibulum Vaughan) type SC435 (E) *P. compressa* Dana (forma fragilis Vaughan), Pearl Harbor, Oʻahu, type SC456; (F) *P. compressa* Dana (forma angustisepta subforma paucispina), Molokaʻi, Pukoo, Duerden 1904, portion of type SC108; (G) *P. compressa* Dana (forma tumida Vaughan) Kāneʻohe, Oʻahu, virtual paratype SC449; (H) *P. compressa* Dana (forma angustisepta Vaughan), Pukoo, Molokaʻi, virtual paratype SC441; (I) *P. compressa* Dana (forma densimurata Vaughan), Pukoo, Molokaʻi, syntype SC440; (J) *P. compressa* Dana (forma abacus) Vaughan, Kāneʻohe, Oʻahu, virtual paratype SC452; (K) *P. compressa* Dana (forma granimurata Vaughan) Kāneʻohe, Oʻahu, syntype SC451; (L) *P. compressa* Dana (forma abacus Vaughan), Kāneʻohe, Oʻahu, virtual paratype SC450.

considered acceptable if there was a high band or a smear with at least half of the DNA in the sample above 2500 bp. Samples failing to meet this criterion were re-extracted. Extractions were quantified with the AccuBlue™ High Sensitivity dsDNA quantitation kit (Biotium, Inc.) with 8 standards by measuring absorbance at $\lambda_{Ex}/\lambda_{Em}$ 485/530 nm using a SpectraMax M2 microplate reader (Molecular Devices, LLC) or a Qubit® fluorometer (Thermo Fisher Scientific, Inc.). To reduce the amount of extraction used per quantification, we diluted 1 µl of DNA with 9 µl HPLC grade water on parafilm to obtain the 10 µl needed for quantification. All extractions and standards were well mixed by flicking and spun-down before they were digested.

2.3. Restriction enzyme digestion

All samples were adjusted by dilution or evaporation with a speed-vac (at room temperature) to a final concentration ≈ 1 µg of DNA in 25 µl prior to digestion. The samples were cleaved using NEB (New England BioLabs) frequent cutter restriction enzymes MboI and Sau3AI to cleave sequences at GATC cut sites (Toonen et al., 2013). Digestions were performed in 50 µl reactions consisting of 18 µl HPLC grade water, 5 µl Cutsmart Buffer, 1 µl MboI (1 unit), 1 µl Sau3AI (1 unit) and 25 µl dsDNA (≈1 µg) with the following thermocycler profile: 37 °C for 3 h, then 65 °C for 20 mins and hold at 15 °C. The digested samples were then cleaned using Beckman Coulter Agencourt AMPure XP PCR purification beads at a 1:1.8 (DNA:beads) ratio following the standard protocol. The digests were then run on a 1% agarose gel (as above) and were considered properly digested when there was a smear with little to no DNA above 2500 bp.

2.4. Library preparation

The libraries were created following either the Illumina TruSeq® Sample Prep v2 Low Throughput (LT) protocol including gel excision and PCR, or using the Illumina TruSeq® Nano DNA Library Preparation Kit without PCR. All libraries were size selected for 300–500 bp by gel excision or using magnetic beads, and passed two quality control steps (bioanalyzer and qPCR) in the Hawaiʻi Institute of Marine Biology (HIMB) Genetics Core Lab before sequencing. Each library consisted of a single sample with a ~300–500 bp insert ligated to a unique Illumna barcode and

Table 1
Collected sample information, reads and lengths after trimming, reads mapped to reference sequences and short read archive number (SRI#). Abbreviations; mt = mitochondrial, K.Bay = Kāneʻohe Bay, NWHI = North West Hawaiian Islands, FFS = French Frigate Sholes, T_reads = trimmed reads; acov = mean coverage, sdcov = standard deviation of coverage, refseq = percentage of reference sequence covered.

							mt genome				rDNA				Histones				Transcriptome				
Code	Label	Genus	Species	Location	T_reads	length	reads	avcov	sdcov	refseq	reads	avcov	sdcov	refseq	reads	avcov	sdcov	refseq	reads	avcov	sdcov	refseq	SRA #
PlLob1	PL1W	*Porites*	*lobata*	Oahu, La	3,286,926	72	2,384	12.9	13.6	98%	36,476	368	429	97%	5,521	80	130	96%	774,472	1.4	183	19%	SAMN06648849
Plob02	PL2W	*Porites*	*lobata*	Oahu, La	4,442,626	78	2,519	13.6	14.0	98%	21,647	235	277	97%	11,179	173	251	96%	1,085,049	1.9	537	23%	SAMN06648850
Plob3	PL3W	*Porites*	*lobata*	Oahu, La	6,533,302	44	1,095	5.9	7.2	85%	31,912	321	399	97%	7,153	104	133	96%	539,333	0.8	168	15%	SAMN06648851
Coral1	PL1L	*Porites*	*lobata*	Oahu, MCW	2,619,180	128	765	5.1	8.0	81%	17,705	318	316	97%	4,364	108	119	96%	387,584	1	11	15%	SAMN06648852
Coral2	PL2L	*Porites*	*lobata*	Oahu, MCW	2,365,636	155	1,812	16.0	11.5	99%	15,195	284	142	99%	6,772	195	112	97%	360,565	1.1	11	20%	SAMN06648853
Coral5	PL5L	*Porites*	*lobata*	Oahu, K	1,338,310	172	1,160	10.9	12.7	94%	8,072	184	95	88%	3,324	106	74	64%	174,504	0.6	8	14%	SAMN06648854
Coral6	PL6L	*Porites*	*lobata*	Oahu, MN	2,117,954	178	1,232	12.1	10.8	97%	10,341	229	154	98%	7,199	224	166	94%	279,835	0.9	10	19%	SAMN06648855
Coral7	PL7L	*Porites*	*lobata*	Oahu, MN	1,551,034	141	502	4.8	5.9	83%	18,650	407	344	99%	8,619	276	258	95%	191,216	0.6	9	13%	SAMN06648856
Coral8	PL8L	*Porites*	*lobata*	Oahu, MN	2,586,262	170	2,929	28.2	20.7	99%	14,324	287	287	99%	1,853	57	39	93%	352,555	1.1	10	22%	SAMN06648857
Coral9	PL9L	*Porites*	*lobata*	Oahu, MN	2,569,830	182	1,156	11.6	11.1	97%	17,568	403	274	99%	5,228	162	138	97%	349,602	1.2	12	22%	SAMN06648858
Coral10	PL10L	*Porites*	*lobata*	Oahu, MN	4,851,770	184	1,957	19.4	21.1	99%	38,231	855	898	98%	20,174	662	737	95%	709,588	2.5	29	27%	SAMN06648859
Pbrigl24	PBNWHI	*Porites*	*c.f. brighami*	NWHI; FFS	4,736,444	84	3,190	22.1	22.2	100%	25,430	269	280	98%	7,398	114	119	97%	719,764	1.6	17	27%	SAMN06648860
Pcom1	PC1W	*Porites*	*compressa*	Oahu, La	4,391,642	72	2,396	12.9	17.8	93%	28,115	354	531	100%	10,520	151	292	97%	716,985	1.4	27	21%	SAMN06648861
Pcom2	PC2W	*Porites*	*compressa*	Oahu, La	3,970,170	57	795	4.5	11.9	55%	22,030	216	488	96%	15,238	223	605	95%	649,038	1.2	241	11%	SAMN06648862
Pcom3	PC3W	*Porites*	*compressa*	Oahu, KB	2,662,612	62	834	4.5	9.3	85%	28,517	280	568	95%	13,905	202	529	93%	508,249	0.9	142	11%	SAMN06648863
PcomL28	PC4W	*Porites*	*compressa*	Oahu, KB	5,889,650	97	5,073	35.7	82.2	90%	14,794	179	399	97%	20,975	351	635	97%	1,068,923	2.6	37	24%	SAMN06648864
BLL62	PE1W	*Porites*	*evermanni*	Oahu, La	4,204,302	90	4,290	29.9	28.1	100%	17,265	191	200	99%	14,691	236	283	98%	679,109	1.5	14	27%	SAMN06648865
PeveR2	PE2W	*Porites*	*evermanni*	Oahu, La	8,114,324	117	1,783	14.6	40.2	68%	241,795	4,182	8,360	100%	38,018	133	1,489	100%	1,067,831	2.8	93	12%	SAMN06648866
Coral4	PE4L	*Porites*	*evermanni*	Oahu, K	1,780,300	166	1,249	11.8	9.2	99%	10,114	213	173	98%	2,432	69	67	94%	201,528	0.6	6	15%	SAMN06648867
PrusR10	Prus	*Porites*	*rus*	Hawaii, Kona	3,978,198	87	1,950	13.7	37.8	88%	7,731	82	236	94%	14,732	235	557	99%	776,113	1.8	44	14%	SAMN06648868
PsupL25	Psup	*Porites*	*superfusa*	Palmyra Atoll	5,337,350	88	2,728	19.2	31.2	98%	25,329	270	585	91%	3,819	58	142	94%	760,722	1.8	48	16%	SAMN06648869
				Mean	3,777,515	115	1,990	15	20	91%	31,011	482	735	97%	10,624	187	327	94%	588,217	1.4	79	18%	

Abbreviations:
T_reads, trimed reads.
Avcov, mean coverage.
Sdcov, standard deviation of coverage.
FFS, French Frigate Shoals.
KB, Kaneohe Bay.
K, Kewalo.
La, Lanikai.
MCW, Maunalua Bay (China Walls).
MN, Maunalua Bay (siteN).
NWHI, North West Hawaiian Islands.

sequencing adaptors and eight to twelve libraries were sequenced per lane. The samples were sequenced on a MiSeq® (Illumina, Inc.) at the HIMB Genetics Core Lab, with one sample PeveR2 (PE2W) was sequenced at the University of Texas at Arlington, Genomics Core Facility on a MiSeq® (Illumina, Inc.). The sequenced lengths, number of reads and sample information for each library are presented in Table 1.

2.5. Mitochondrial reference assemblies

Raw Ilumina reads were sorted by barcodes. Lists of paired reads (expected distance = 500) were trimmed on both 5′ and 3′ ends of adapter sequence (allowing no mismatch and a minimum overlap of 8 bp). Low quality bases (with more than a 0.1% chance of error) were removed using Geneious v8.0.2 (Biomatters Inc.). The whole mitochondrial genome of *Porites okinawensis* (NC015644) was used as a reference sequence, using *P. panamensis* (NC024182) reference sequence resulted in identical alignments and trees (data not shown). Each library was assembled to the mitochondrial reference sequence using the default parameters (low sensitivity with no fine tuning and the fast/read mapping settings). These assemblies were visually inspected which revealed that they were of very high quality, with very low levels of polymorphism. Consensus sequences were then calculated from each library (not including the reference sequence) using the 0% majority option and N's were called if coverage was less than 3X. Manual inspection and editing of the contigs produced identical results. Additional mitochondrial genomes from the family Poritidae (NC024182 *P. panamensis,* NC008166 *P. porites,* NC015643 *Goniopora columna*) were included to provide a broader phylogenetic framework. Multiple sequence alignments were constructed using MUSCLE (Edgar, 2004) with 8 iterations and phylogenetic trees were constructed with PHYML v2.2.0 (Guindon et al., 2009) using the HKY model with 1000 bootstrap iterations (and otherwise default settings for each program).

2.6. Binning data for de novo assembly

In order to place reads into bins of putative coral and symbiont loci, all reads were mapped against transcriptomic reference datasets. Mapping genomic data to transcriptomic data is imperfect due to differences in intron/exon boundaries and mRNA splicing, nevertheless this method should allow reads to be binned into organismal categories. The *P. lobata* transcriptomic reference sequences (n = 21,062) were downloaded from http://comparative.reefgenomics.org/ on April 1, 2016. These sequences represent putative orthologous protein-coding sequences from coral genomes (Bhattacharya et al., 2016). These reference sequences were sorted by size and concatenated, together with a 200 bp segment of N's separating each transcript. All *Porites* libraries (cleaned raw reads) were then mapped to this 'transcriptomic' reference sequence using the Geneious v8.1.4 Mapper set to medium/fast sensitivity with up to 5 iterations. The resulting reads were exported as fastq files and this subset of the data was referred to as the 'coral transcriptomic' data subset for *de novo* assembly and phylogenomic analysis. Putative *Symbiodinum* protein coding sequences from the *P. australensis* holobiont Shinzato et al. (2014) were sorted by size and concatenated together with a 200 bp segment of N's separating each transcript. All *Porites* libraries were then mapped to this '*Symbiodinium* transcriptomic' reference sequence using the Genieous v8.1.4 Mapper set to medium/fast sensitivity with up to 5 iterations. The resulting reads were exported as fastq files and this subset of the data was referred to as the '*Symbiodinium* transcriptomic' data subset.

2.7. Phylogenomic analysis

The program pyRAD v.3.0.2 (www.dereneaton.com/software) was used to examine the holobiont metagenomic dataset (the entire dataset), as well as subsets of the reads that map to the coral and *Symbiodinium* transcriptomes. All reads were filtered and merged using PEAR, an Illumina Paired-End reAd mergeR (Zhang et al., 2014) using the following settings: –u 0.06, -n 36, -q 20, -j 6, -p 0.05, -t 36, as recommended by the pyRAD v.3.0 documentation (Eaton, 2014). The merged and unmerged reads were then combined into a single file for further analysis. The program was invoked using the following parameters: (6) restriction site = GATC; (8) min depth = 8; (9) NQual = 6; (10) Wclust = 0.85; (11) Datatype = GBS; (12) MinCov = 4; (13) MaxSH = 3; (26) maxSNP = 20; (29) trim = 1,1; (31) call maj. = 2; (35) Hierarchical = 1; (outgroups 1 R10prus, L25psup, L62bl, PeveR2 ingroups 1 PCom3, PCom2, PCom1, PLob1, PLob02, PLob3, L28pcom oddgroup 1 L24pbrig). These 'relaxed' parameter settings were chosen after several preliminary runs retrieved few loci and closer inspection of filtered loci indicated a high proportion of divergent alleles. The resulting phylogenetic trees were constructed using RAxML (raxmlHPC-PTHREADS-SSE3); (Stamatakis, 2006), invoked with the following parameters: –f a T 10 –m GTRGAMA –x 1234 –# 500 –p 1234.

2.8. Reference mapping and SNP analysis

Single nucleotide polymorphisms (SNPs) were analyzed from both the coral transcriptomic and *Symbiodinium* transcriptomic data subsets using modified settings from programs in the dDocent v2.25 (Puritz et al., 2014a) pipeline. Trimmed reads were mapped to the *P. lobata* transcriptome reference sequence and the *Symbiodinium* transcriptomic reference sequence using BWA (Li, 2013) with the following settings: -t 16 -a -M -T 10 –R. Each BAM file was sorted using Samtools (Li et al., 2009), and BAM files from each library were merged together using Bamtools (Barnett et al., 2011). INDEL positions were realigned using Genome Analysis Tool Kit (McKenna et al., 2010), and variants were called with FreeBayes (Garrison and Marth, 2012) using the following settings: -0 -E 3 -G 5 -z 0.1 -X -u -n 4 --min-coverage 5 --min-repeat-entropy 1 -V –b. The resulting concatenated BAM file was visually inspected using Tablet (Milne et al., 2013), and spot checks of the assemblies appeared very high quality (i.e. well aligned with relatively few bi-allelic polymorphisms). The resulting VCF files were further filtered using VCF tools (Danecek et al., 2011) to create additional data subsets filtered to contain varying numbers of SNPs and levels of missing data. The data subset of the coral transcriptome 'coralmax' was generated to allow up to 5 missing taxa per locus, with each SNP thinned to be no closer than 10 bp, the data subset 'coralmin' allowed no missing data, and included SNPs that were at least 300 bp apart; eg. (--min-meanDP 10 --remove-indels --max-missing-count 0 --thin 300 --recode --recode-INFO-all). Similarly, a symbiomax data set consisted of variants with a minimum coverage of 5x of reads that mapped to *Symbiodinium*.

The resulting VCF files were converted to STRUCTURE (Pickrell and Pritchard, 2012), EIGENSOFT (Price et al., 2006), and nexus format for the SNAPP/BEAST2 (Bouckaert et al., 2013) package using PGDSpider (Lischer and Excoffier, 2012). The SNAPP plugin was implemented in the BEUti/BEAST2 programs, with mutation rates and priors estimated during the MCMC chains. The mutation rate was estimated from the data, only polymorphic sites were included in the dataset, otherwise all other settings were set as default and the MCMC run was sampled every 1000 generations and run for 1,000,000 generations. Simulations were conducted in STRUCTURE V.2.3 for 5 replicate runs with K ranging from 1 to 9, assuming the Admixture model with 100,000 MCMC generations after a burn-in

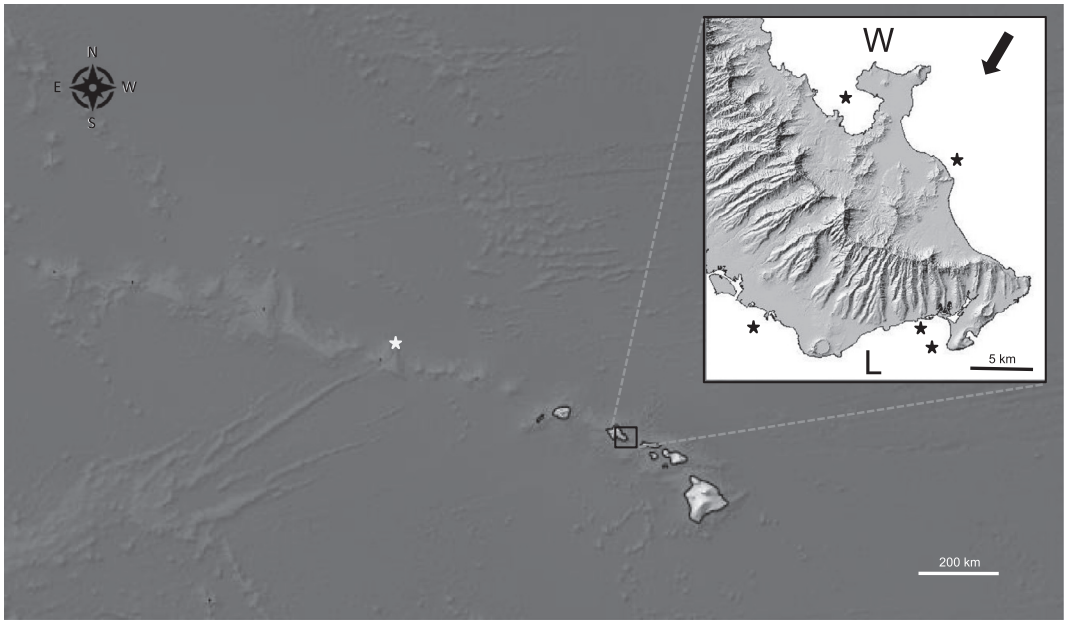

Fig. 3. Collection locations of *Porites* samples from Oʻahu, Hawaiʻi for this study. * = collection locations (see Table 1). The arrow represents prevailing direction of trade winds, W = windward, L = leeward.

of 10,000 generations. The resulting simulations were examined with STRUCTURE HARVESTER (Earl and vonHoldt, 2012), using the method from Evanno et al. (2005). EIGENSOFT formatted files were visualized with the package smartPCA.

2.9. BLAST of de novo assembled contigs

De-novo assembly and BLAST searches were used to identify additional loci of interest and to further examine components of the coral holobiont. Library PcomL28 (PC4W) was one of the highest quality libraries representing *P. lobata* or *P. compressa* and was therefore selected for additional *de novo* assemblies to examine larger contigs. The reads were quality trimmed as above; reads below 20 bp were excluded. Assemblies were conducted using the *de novo* low sensitivity/fast settings of the Geneious v 8.1.4 assembler. Contigs >200 bp (357,879) were compared against the NCBI nr nucleotide database to identify loci with a threshold of e values lower than 1e-22 with hits over at least 50% of the contig. To examine the effect of contig length on taxonomic composition, the contigs were sorted by length and divided into bins (>1 kb, 500 bp–1 kb, 200–500 bp) and searched against a local version of the National Center for Biological Information (NCBI) Genbank nt database, which was downloaded on 4/13/2015 using Megablast (Altschul et al., 1997; McGinnis and Madden, 2004). The contigs were sorted by e-scores and two of the longest contigs with high coverage and long BLAST hits (a 7.8 kb ribosomal contig and a 5.8 kb long histone contig), were selected for further reference mapping against all libraries. In addition, these contigs were confirmed by designing PCR primers and Sanger sequencing. Although typical RAD data results in short non-overlapping loci, closer inspection of these large contigs indicated they were rich in GATC cut sites and areas of high coverage were often bridged by areas of lower coverage, perhaps due to incomplete digestion or degraded DNA (example Fig. 3). All libraries were assembled to the consensus sequence of these contigs using the default parameters (low sensitivity with no fine tuning and the fast/read mapping settings) in Geneious v8.1.4. Consensus sequences were constructed from each library (excluding the reference sequence) using the 75% majority option and N's were called if coverage was less than 3x (manual inspection and editing of low quality portions of the contigs produced the same results). Multiple sequence alignments were constructed using MUSCLE (Edgar, 2004) with eight iterations and phylogenetic trees were constructed with PHYML v2.2.0 (Guindon et al., 2009) using the HKY model with 1000 bootstrap iterations.

3. Results

3.1. Mitochondrial reference assemblies

Mapping paired reads in Geneious v8.1.4 to the mitochondrial reference genome (GenBank # NC015644) resulted in a mean of 1990 reads covering of 91% of the reference sequence at a mean depth of 15 ± 20 (mean ± sd) per library. The aligned mitochondrial consensus sequences revealed low divergence across samples with 91.9% of the positions conserved across the full alignment (Table 1, Fig. 4). Fig. 4 provides a graphic illustration of paired reads mapping to a single library (sample L24) with mean coverage of 22× over 100% of the reference sequence, illustrating both stacks of loci (vertical coverage) and coverage across the reference sequence (horizontal coverage). Of the ~18.8 kb alignment, only 1.3% (256 bp) of the positions were variable and only 0.5% (98 bp) were parsimony informative (Table 2). The mean distance between *P. lobata* and *P. compressa* was 28.4 bp ± 2.2 bp (mean ± sd), which is comparable to the within group distance for *P. lobata* (23.3 bp ± 2.3 bp) and *P. compressa* (19.7 bp ± 3.1 bp). Several *P. lobata* and *P. compressa* individuals differed by only one or two nucleotides over the entire 18.8 kb alignment and interestingly, the sample from GenBank identified as *P. okinawensis* differs from *P. compressa* (sample PC1W), by only 8 nucleotides. By comparison the mean genetic distance between *P. evermanni* and the *P. lobata/compressa* group was 56.1 bp ± 4.9 bp. A RAxML tree of the aligned consensus sequences from each library showed strong bootstrap support across the majority of nodes (Fig. 5). The resulting tree was consistent with previous phylogenetic work (Forsman et al., 2009), showing strong bootstrap support for outgroup taxa, but several nominal species with variable skeletal morphology formed

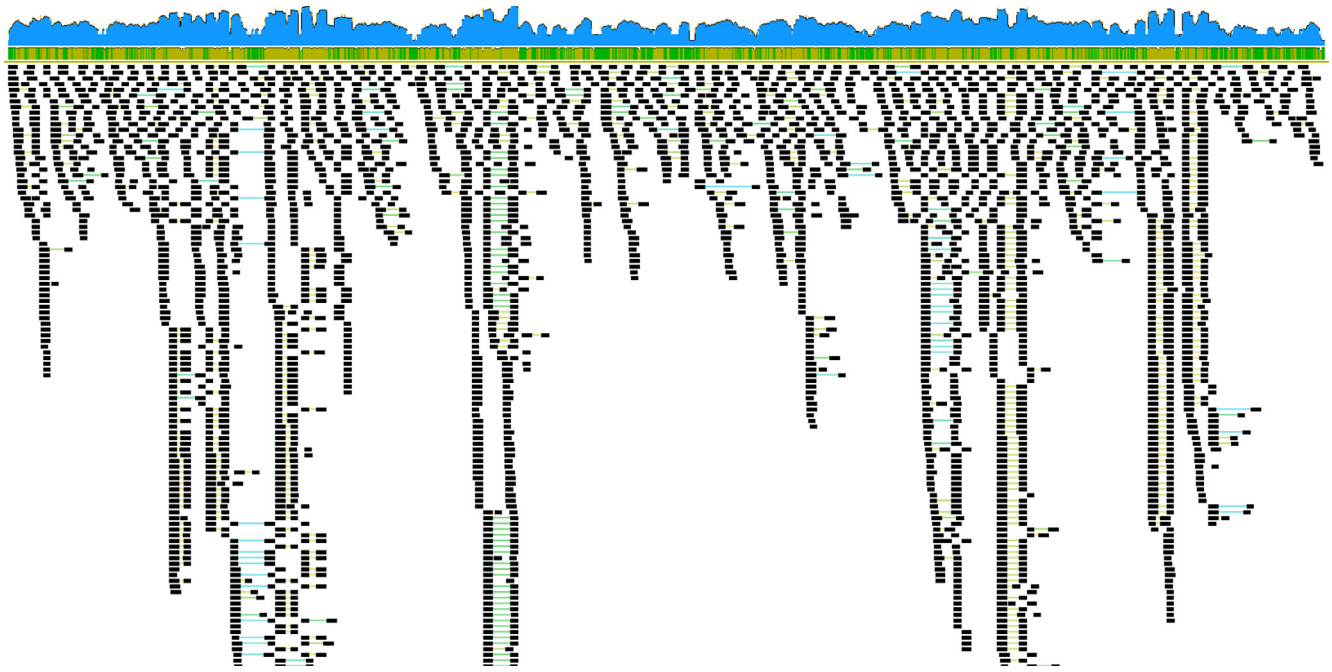

Fig. 4. Example of coverage across a mitochondrial genome (sample L24). Stacks of reads with higher depth of vertical coverage are associated with the GATC restriction enzyme cut site. Coverage on the log scale is indicated by the blue graph, reads mean 130 bp (±19 stdev) with a mean coverage of 22× (±22stdev), covering 100% of the 18 Kb reference sequence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Multiple sequence alignment properties for reference mapping of *Porites* species libraries where: Pi = Parsimony informative sites and nt = nucleotides.

Alignment	Length	Conserved		Variable		Pi		Singletons		Missing nt	
mtgenome	20,092	18,468	(91.9%)	256	(1.3%)	98	(0.5%)	158	(0.8%)	49,264	(20.4%)
rDNA	7848	7322	(93.3%)	307	(3.9%)	102	(1.2%)	197	(2.5%)	m3257	(3.5%)
Histone	5830	5386	(92.4%)	330	(5.7%)	125	(2.1%)	202	(3.5%)	1785	(2.6%)

a large unresolved species complex (*P. lobata*, *P. compressa, P. okinawensis*, and *P.cf. brighami*, Fig. 5).

3.2. Phylogenomic analysis

Each library averaged ~3.8 M reads after filtering low quality reads, with a mean length of 115 bp (Table 1). After merging overlapping reads with PEAR and discarding orphaned reads shorter than 50 bp, this resulted in a mean of ~2.5 m reads per library for the total 'holobiont' data set, with ~1.2 m reads passing initial quality filters in pyRAD v.3.2 (Table S1). These reads clustered into ~ 594,000 loci with a mean coverage depth of 1.5 ± 3.9 (mean ± standard deviation) per library, which after quality filtering resulted in ~116,000 loci with a mean depth of 17× coverage per library (Table S1). After paralog filtering the loci were further reduced to a mean of ~18,000 loci per library (Table S1). Clustering across the 21 libraries resulted in 80,219 total loci after final filtering, which was concatenated into a 10.6 million bp nucleotide alignment.

The RAxML tree of the complete alignment resulted in a highly-resolved tree with strong bootstrap support at every node (Fig. 6A). The taxa clustered consistently with previous studies based on nuclear and mitochondrial markers (Forsman et al., 2009). *Porites evermanni* was strongly supported as reciprocally monophyletic, however there was no resolution between the mounding coral *P. lobata* and the branching coral *P. compressa*. Unexpectedly, *P. lobata* and *P. compressa* samples clustered geographically based on windward or leeward sides of the island, except for a single sample (PL1L; Fig. 6A). The 'coral transcriptomic' data averaged ~386,000 reads passing filtering per library, forming clusters of ~100,000 loci with 1.6 ± 59.8 (mean ± sd) coverage (Table S1). Quality filtering of these loci resulted in ~15,000 with a mean depth of 34 ± 439 (mean ± sd) coverage, and after paralog filtering this resulted in 2055 loci per library (Table S1). Clustering across libraries resulted in 8627 total loci after final filtering, which were concatenated into a 1,016,476 million bp nucleotide alignment. The RAxML tree of the coral transcriptomic alignment was nearly identical to the holobiont tree, although with fewer completely resolved nodes and a rearrangement of two taxa (PL1L, and PC1W; Fig. 6B). As with the holobiont tree, all *P. lobata* and *P. compressa* samples collected from the windward side of the island clustered together and all samples collected from the leeward side of the island clustered together with the exception of a single sample (PL1L; Fig. 6B). In comparison to the mitochondrial tree, the *de novo* assembly trees were more resolved with more strongly supported clades. The mitochondrial tree showed no concordance with collection location as is seen in the holobiont and coral transcriptomic trees. Furthermore, comparison of the mitochondrial genome tree to the holobiont tree indicated mito-nuclear discordance between clades within the *P. lobata* species complex (Fig. 7A and B).

The '*Symbiodinium* transcriptomic' subset of the data had a mean of ~133,000 reads per library, clustering into a mean of ~7000 loci per library with a mean coverage of 6.1 ± 245.9 (Table S1). Further quality filtering of these loci resulted in a mean of 1844 per library with a depth of 108 ± 1104 (mean ± sd) resulting in 146 loci per library after paralog filtering in pyRAD (Table S1). Clustering across all libraries resulted in 590 total loci,

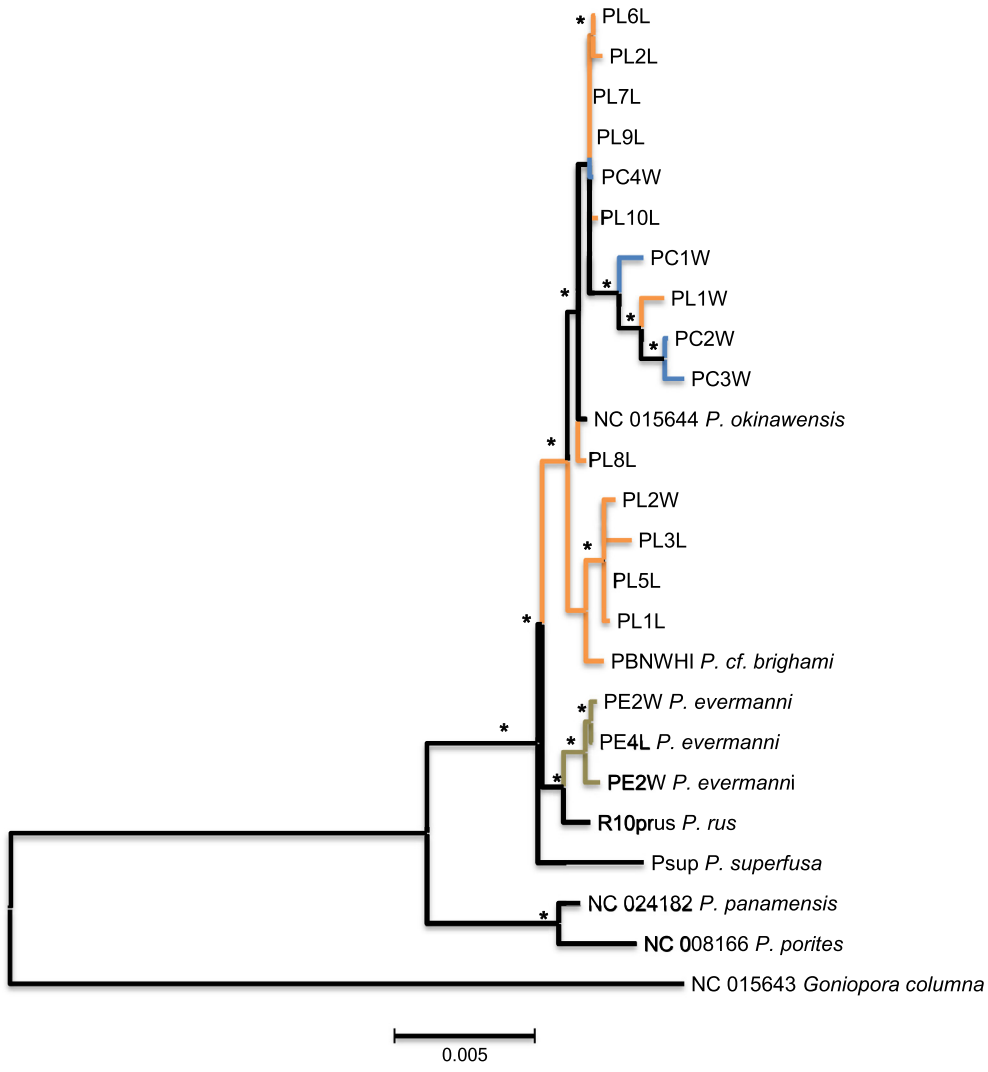

Fig. 5. RAxML tree of the mitochondrial genome consensus sequences with outgroups from GenBank (samples with NC prefix). Branch colors represent morphospecies (blue = *P. compressa*, orange = *P. lobata*, green = *P. evermanni*). W = windward, L = leeward side of O'ahu, Hawai'i (See Fig. 3) and * = maximum likelihood support values higher than 80%. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which were concatenated into a 50,507 bp sequence alignment (Table S1). The RAxML tree of the *Symbiodinium* transcriptomic tree was poorly resolved, with few nodes providing strong bootstrap support, no resolution of *P. lobata* and *P. compressa*, and no resolution of geographic patterns within *P. lobata* (Fig. S1). Nevertheless, *P. lobata* and *P. compressa* clustered together as separate from *P. evermanni* and *P. rus*, with *P. superfusa* as a distant outgroup.

3.3. Reference mapping and SNP analysis

Three alternative methods were used to examine patterns of SNP variation among the samples within the *P. lobata* species complex: SNAPP, STRUCTURE and Principle Components Analysis (PCA). In each case, several data subsets were examined to determine the effects of missing data. The 'coralmax' dataset consisted of reads that map to the *P. lobata* transcriptome with a minimum of 10x coverage and up to 5 missing taxa per locus, this yielded a mean of ~25,000 SNPs per library with a mean depth of 33 reads, the 'coralmin' dataset allowed no missing data, had a meandepth of 34 reads, and each SNP was thinned to be at least 300 bp apart (Table 3). The 'symbiomax' dataset consisted of ~10,000 SNPs per library with a mean depth of 19 reads each. For each dataset, SNAPP resolved differences between *P. evermanni* and the *P. lobata* species complex, but no genetic structure within the *P. lobata* species complex (all results similar to Fig. S2 of the coralmax dataset).

The results from STRUCTURE also showed clear resolution of *P. evermanni* and the *P. lobata* species complex for all datasets (Fig. 8A). Excluding the clearly divergent *P. evermanni* samples and running STRUCTURE simulations on the *P. lobata* complex for various values of K, resulted in an optimal value of 2 as determined by the Evanno et al. (2005) method as implemented in the program STRUCTURE Harvester (Evanno et al., 2005; Earl and vonHoldt, 2012); see Fig. S4 for likelihood plots. The resulting STRUCTURE plots showed no resolution between branching and mounding morphospecies, but instead revealed two groups largely concordant with geographic sampling (Fig. 8B–D). These results were highly consistent with the results obtained from the holobiont and coral transcriptomic datasets. In each of these analyses, samples grouped by their geographical location (windward and leeward), with the exception of a single sample (PL1L), rather than morphospecies (*P. lobata* and *P. compressa*). The PCA analysis of both the 'coralmax' and 'coralmin' dataset analyses also supported this geographical pattern (Fig. 9). The putative *Symbiodinium* loci showed clear separation between *P. evermanni* and the *P. lobata*/*P. compressa* clusters, but for this dataset *P. lobata* and *P. compressa*

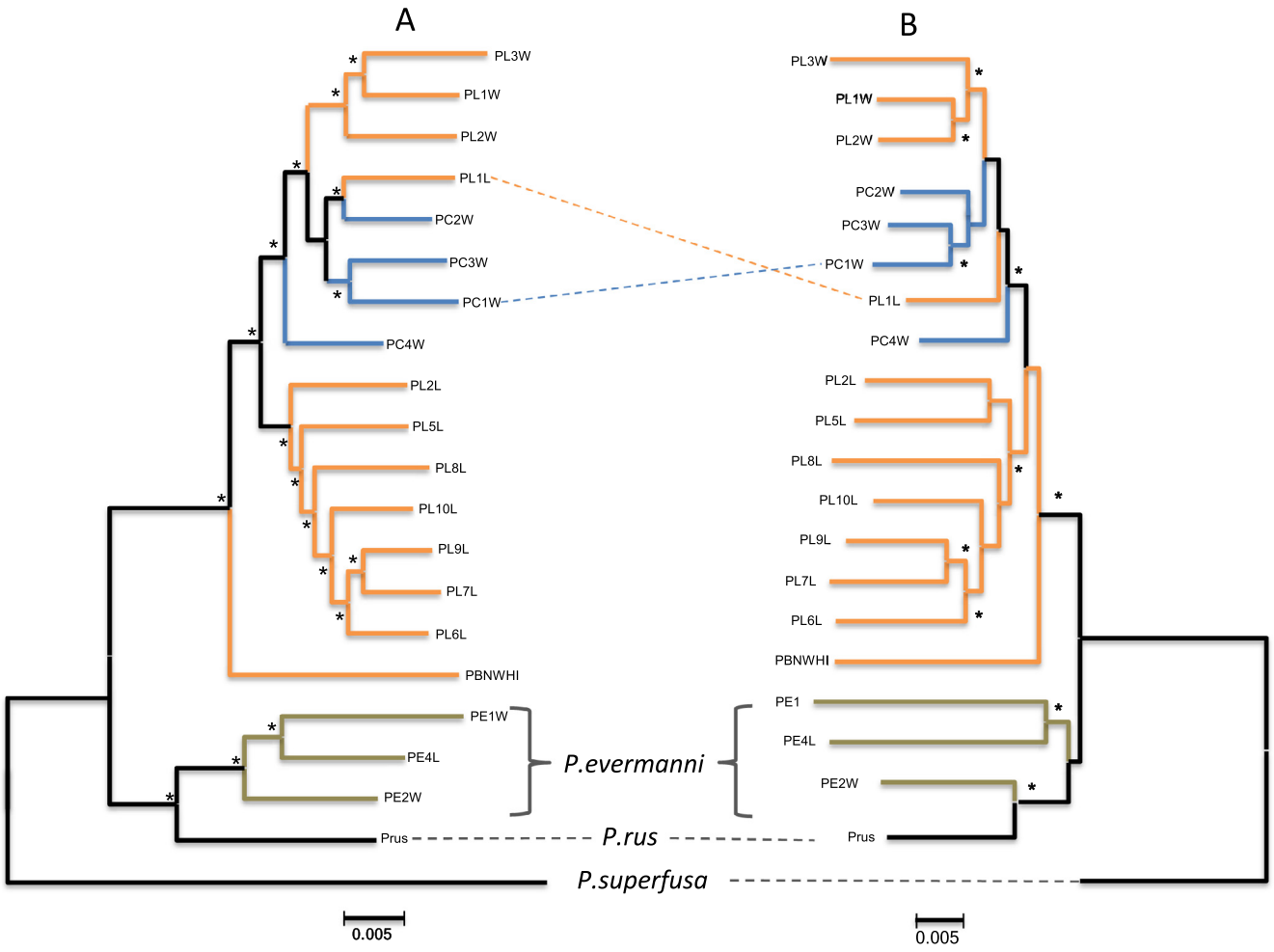

Fig. 6. Illustration of RAxML trees from *Porites* species *de novo* assembled loci. (A) Holobiont dataset, (B) coral transcriptomic dataset. Branch colors represent morphospecies (blue = *P. compressa*, orange = *P. lobata*, green = *P. evermanni*). W = windward, L = leeward side of Oʻahu, Hawaiʻi (See Fig. 3), and * = maximum likelihood support values higher than 80%. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

clustered more closely together than for the coral loci. These patterns also held true regardless of the level of missing data (Fig. 9C).

3.4. BLAST of de novo assembled contigs

A *de novo* assembly and BLAST search of one of the highest quality libraries (PC4W) was used to further characterize loci. The assembly resulted in 566,999 contigs, with an N50 of 500 bp. Of the 357,879 contigs that were over 200 bp, a BLASTn search of the *P. astreoides* transcriptome yielded 7546 (~2%) hits at a threshold of e^{-20} or less, while in contrast a local BLAST search of the same contigs against the *Symbiodinium minutum* genome at the same threshold yielded only 766 (~0.2%) hits, indicating that the contigs were likely to be predominantly coral. BLASTn searches against the NCBI nr database yielded similar results (Fig. S3), with 82% of hits below the e^{-22} threshold blasting to cnidarians and only 3–5% matching putative *Symbiodinum* loci. As the length of the contigs decreased, the proportional diversity of the BLAST hits increased, such that longer contigs (1–6 kb) tended to have a higher proportion of cnidarian hits, whereas shorter contigs (200–500 bp) tended to contain a higher proportion of hits to bacteria and other microbes. Two of the largest contigs with the highest overall coverage had highly significant BLASTn hits to coral ribosomal and histone gene regions and were examined in greater detail.

The ribosomal contig was 7.8 kb long, yielding multiple BLASTn hits to ribosomal genes of coral origin from the NCBI database, including the 18S, the 5.8S, and the 28S genes in the correct order. The strict consensus sequence of this contig was then used as a reference sequence for all 21 libraries in a reference assembly. On average, approximately 42 K reads mapped to this reference sequence at a mean depth of 624 ± 1182 (mean ± sd), with an mean of 97% of the reference sequence covered (Table 1). The resulting 75% majority rule consensus sequences were aligned with a length of 7848 bp, 93% of the sites were conserved with only 3.9% variable positions and 1.2% parsimony informative (Table 2). Only 3.5% of the nucleotide positions in the alignment were missing data, and 2.5% were singletons. The resulting maximum likelihood phylogenetic trees resolved the outgroup taxa with strong bootstrap support but provided no additional resolution of the *P. lobata* species complex (Fig. 7C). The branch lengths of the tips were long relative to the distance between individuals, indicating high variation between samples.

The putative histone region contig was 5.8 kb long with high continuous coverage (~14 K mean reads per library), with BLASTn hits including a cnidarian protein from *Nematostella vectensis* (XM_001640430), and a Histone H3 region with significant BLASTn hits to multiple taxa including stony corals. Reads mapped to this contig at mean depth of 177 ± 472 (±sd) covering an mean of 97% of the reference sequence (Table 1). This contig had 2.1% variable sites and all were parsimony informative

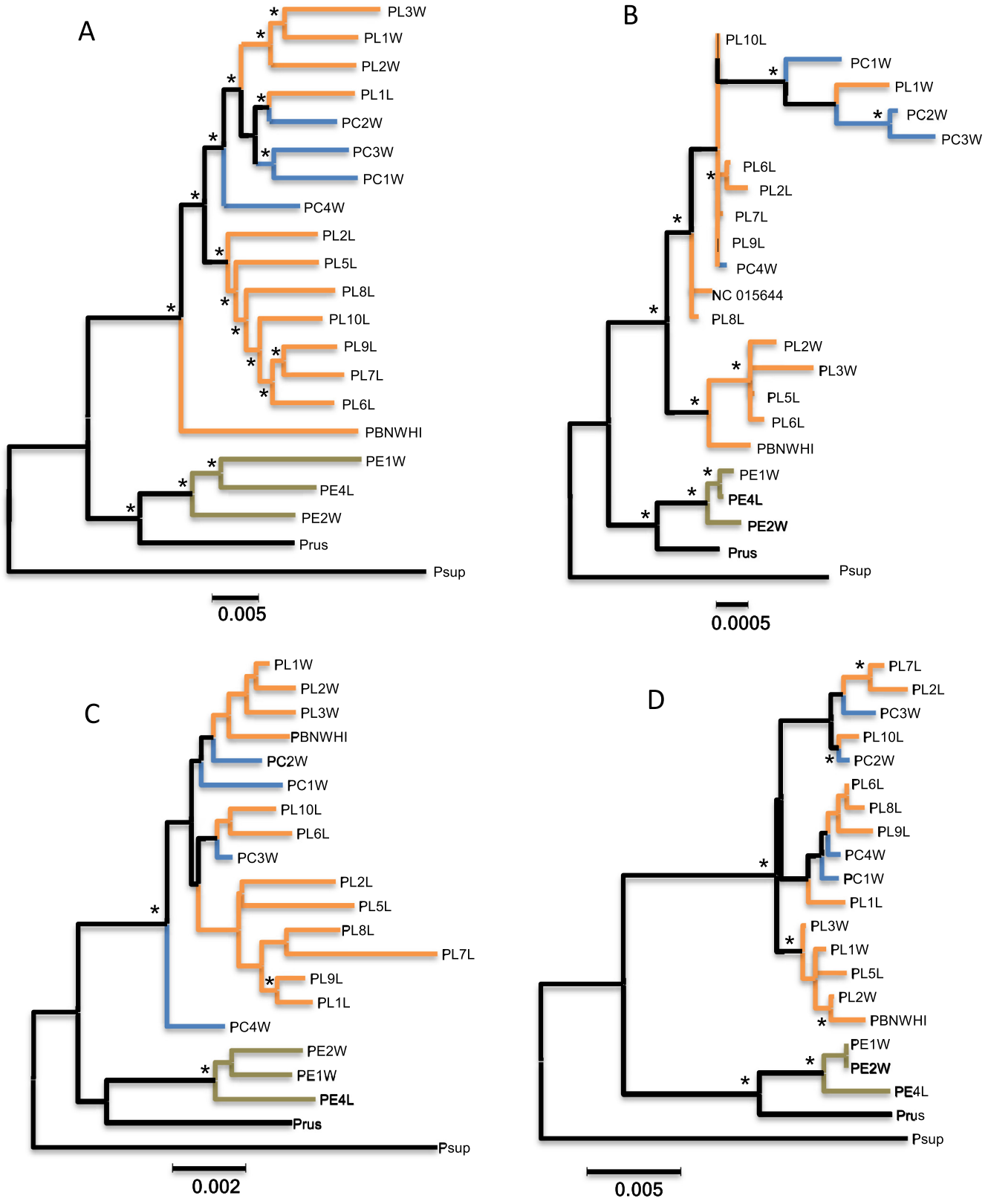

Fig. 7. Comparison of *Porites* species trees using: (A) Holobiont dataset, (B) mitochondrial genome (C) Coral ribosomal array and (D) Coral histone region. Branch colors represent morphospecies (blue = *P. compressa*, orange = *P. lobata*, green = *P. evermanni*), * = maximum likelihood support values higher than 80%, W = windward, and L = leeward side of Oʻahu, Hawaiʻi (See Fig. 3). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Table 2). As with the mitochondrial and ribosomal markers, the outgroup taxa were clearly resolved with strong bootstrap support (Fig. 7D), however, again there was no resolution within the *P. lobata* species complex and no clustering concordant with geographic sampling as was seen in the holobiont and coral transcriptomic trees.

Table 3
Number of SNPs per library and depth after assembly to the transcriptomic reference sequences and final filtering. The 'coralmax' dataset consisted of reads that mapped to the *Porites lobata* transcriptome with a minimum of 10× coverage, allowing 5 missing taxa per locus, with each SNP thinned to be at least 10 bp apart, while the 'coralmin' dataset allowed no missing data per locus resulting in 1537 loci shared among all samples with each SNP thinned to be at least 300 bp apart. The 'symbiomax' dataset included all variants with a minimum coverage of 5×.

	Coralmax		Coralmin		Symbiomax	
Sample	Sites	Mean depth	Sites	Mean depth	Sites	Mean depth
PBNWHI	26,657	45	1537	48	12,885	17
PC1W	26,106	24	1537	24	10,771	15
PC2W	22,283	20	1537	17	7,656	15
PC3W	22,668	16	1537	14	7,607	13
PC4W	26,630	55	1537	58	12,111	26
PE1W	25,766	23	1537	24	11,966	19
PE2W	22,923	68	1537	62	8,230	59
PE4L	24,246	21	1537	20	9,249	11
PL10L	26,621	89	1537	92	12,867	40
PL1L	26,502	32	1537	33	9,564	14
PL1W	26,287	20	1537	21	10,793	16
PL2L	26,651	34	1537	34	10,534	14
PL2W	26,451	23	1537	24	11,230	14
PL3W	25,230	12	1537	12	9,247	10
PL5L	25,750	21	1537	21	9,031	10
PL6L	26,273	34	1537	34	10,710	14
PL7L	25,351	21	1537	22	8,780	13
PL8L	26,334	37	1537	36	11,042	18
PL9L	26,272	43	1537	42	10,926	24
Mean	25,526	33	1537	34	10,274	19

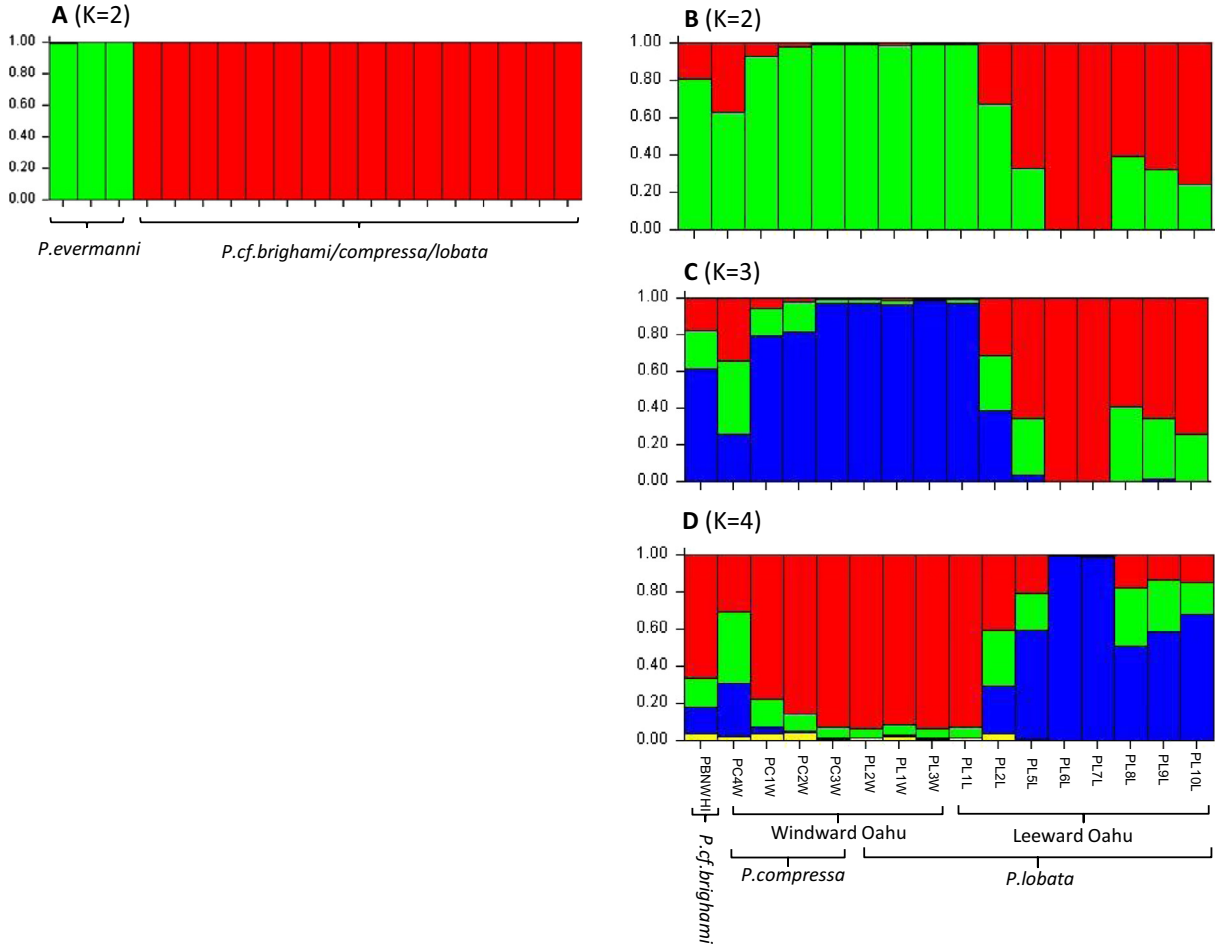

Fig. 8. STRUCTURE results: (A) among species (*Porites evermanni*, *P. cf. brighami*, *P. compressa*, *P. lobata*) comparisons, the optimal value for K was 2 as determined by the Evanno et al. (2005) method implemented by STRUCTURE Harvester; after excluding *P. evermanni*, geographic structure can be seen within the *P. lobata* species complex (B–D). K = 2 was considered optimal as determined by the Evanno et al. (2005) method implemented by STRUCTURE Harvester. Results from 5 separate runs were highly similar to results presented here (B) K = 2, (C) K = 3, (D) K = 4.

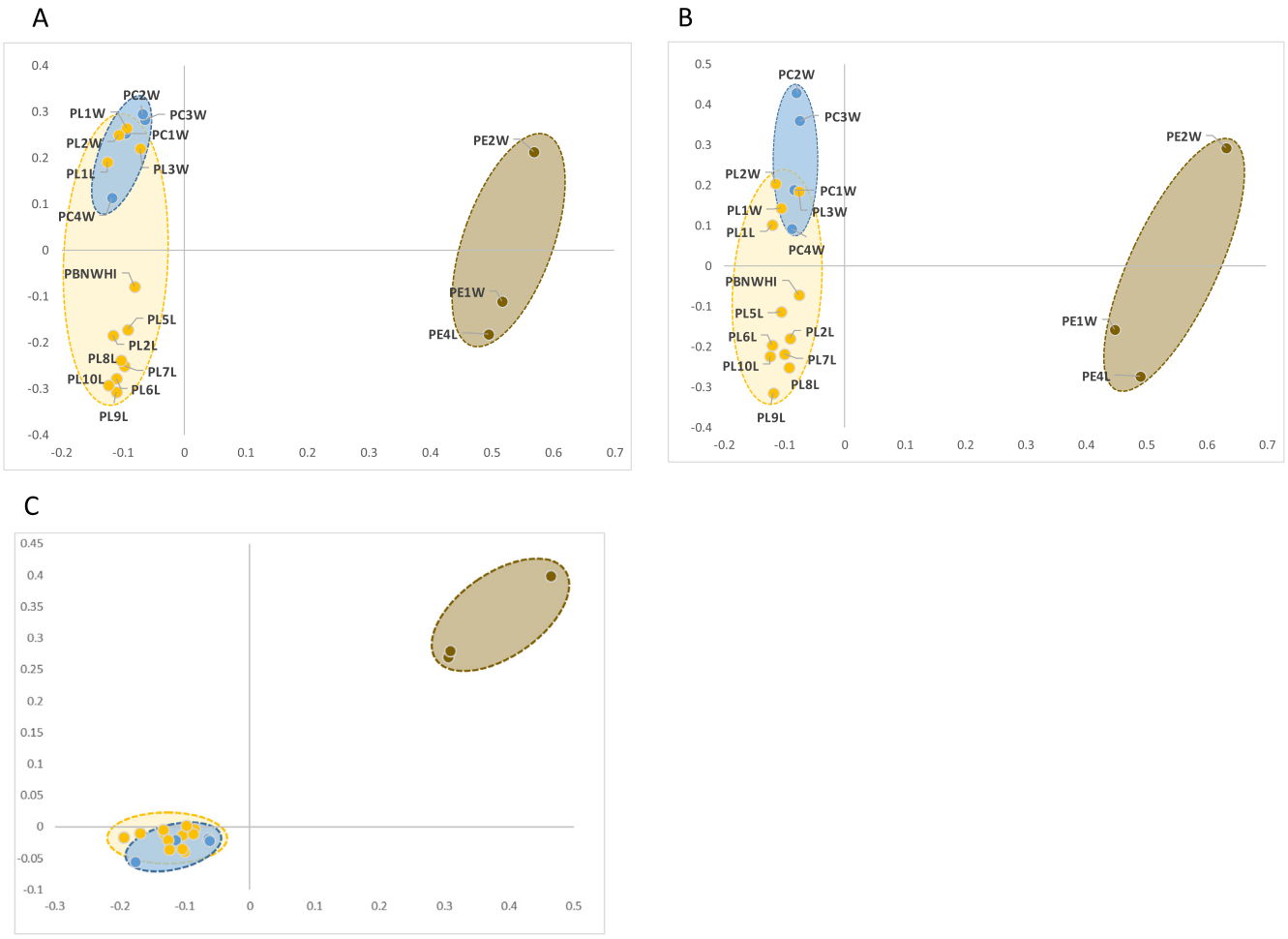

Fig. 9. Principal Components Analysis (PCA) of putative coral and *Symbiodinium* protein coding loci found in this study. (A) the 'coralmax' dataset allowed up to 5 missing taxa per locus, consisted of ~20 k SNPs that mapped to the *Porites lobata* transcriptome; (B) the 'coralmin' dataset allowed no missing data and consisted of ~1.5 k SNPs that map to *P. lobata* transcriptome. (C) ~17 K SNPs that map to putative *Symbiodinium* loci. (Y axis = PCA2, x axis = PCA1. Ellipsoids were drawn to surround samples from each species to illustrate overlap; Brown = *P. evermanni*, Yellow = *P. lobata*, blue = *P. compressa*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

To provide insights into the nature of genetic variation in this key reef-building coral species complex, we explored several strategies for examining both well-characterized and anonymous loci from the coral holobiont. Nearly complete coral mitochondrial genomes, large ribosomal and histone contigs, thousands of loci representing putative coral genes, as well as anonymous SNPs from the entire holobiont dataset revealed no evidence for genetic isolation between *P. lobata* and *P. compressa*. Unexpectedly, we found more genetic structure among sites than between the morphologically defined coral species. This finding is consistent with two explanations; either colony morphology is a polymorphic trait or there is ongoing hybridization and introgression between species. If there are fewer barriers to gene flow between species than between opposite sides of the island of Oahu, then it is not clear how these morphological differences could be maintained in sympatry.

Phenotypic plasticity between branching and plating colony morphology in response to light has been observed previously in a different *Porites* species (*P. sillimaniani*) as determined by reciprocal transplantation experiments (Muko et al., 2000). Muko et al. (2000) proposed that branching or plating across a depth gradient is a mechanism for mediating the total light available to the colony. The branching morphospecies *P. compressa* does indeed have a distinct habitat preference for shallow areas with limited wave exposure, while *P. lobata* typically dominates deeper and more wave exposed environments (Storlazzi et al., 2004). However, the two species' ecological distributions also overlap significantly and they often occur side-by-side in the same environment (e.g. Fig. 1), therefore phenotypic plasticity in this case remains an unsatisfactory explanation.

Phenotypic plasticity and population-level polymorphism have been proposed to be related in a theoretical framework as a response to selection (Pigliucci, 2005; Forsman, 2014). If colony-level morphology is a polymorphic trait that exhibits alternative phenotypes in contrasting environments, then this would be similar to color polymorphisms in a variety of animals (Gray and McKinnon, 2007; Forsman et al., 2008) or settlement preferences of larval invertebrates (Toonen and Pawlik, 1994, 2001). A stable polymorphism appears to be consistent with this study and would have major implications for the understanding of species-level morphological variation in corals. However, this explanation is not entirely satisfactory, because *P. lobata* has a larger geographic range than its branching counterparts: *Porites compressa* (endemic to Hawai'i) or *P. cylindrica* (Indo-Pacific). These branching species do not occur in the Eastern Pacific, whereas *P. lobata* is abundant in the Eastern Pacific. Further complicating the issue, recent work based on coalescent analysis of seven genes indicate that *P. evermanni* and *P. lobata* may hybridize in the Eastern Pacific, but not

in Hawai'i or American Samoa (Hellberg et al., 2016). Our results support the finding that *P. evermanni* and *P. lobata* are clearly resolved in Hawai'i, however samples from the Eastern Pacific were not examined for this study. Veron (1995) and Veron and Stafford-Smith (2000) proposed that coral species boundaries are semi-permeable, varying in space and time in response to changing ocean circulation patterns. Differentiating this hypothesis from alternative explanations such as polymorphism is a major challenge that would require large scale geographic sampling of all recognized morphotypes for a large number of known coral loci and even with semi-permeable species barriers it is difficult to understand how a trait such as colony morphology could remain distinct in the face of considerable gene flow. Further geographic sampling and or genome wide association studies should allow the alternative hypotheses of polymorphism or hybridization to be evaluated. Further colony level or corallite level morphometric comparisons between species may also provide insights since previous work has indicated that microskeletal landmarks can distinguish closely related species and are correlated with genetic distance in *P. lobata* across the Pacific (Forsman et al., 2015).

In this study, complete or nearly complete mitochondrial genomes were assembled from each of the genomic libraries by reference assembly to previously sequenced whole mitochondrial genomes (Medina et al., 2006; Lin et al., 2011). Coral mitochondrial genomes evolve at a rate approximately 10–20 times slower than vertebrates mitochondrial genomes (Shearer et al., 2002; Hellberg, 2006), and therefore longer sequences are needed for equivalent phylogenetic resolution at the species-level or population-level divergence. Complete mitochondrial genomes are usually obtained by isolation and purification of mitochondria, followed by shotgun sequencing or long polymerase chain reaction and molecular subcloning, which is labor intensive and costly (van Oppen et al., 1999; Fukami and Knowlton, 2005; Medina et al., 2006). Seventy-four universal PCR primers have recently been developed for Scleractinian corals (Lin et al., 2011), which reduces the cost of obtaining mitochondrial genomes considerably, however the present method is even more rapid and inexpensive (currently ~$30 to $200 USD per library including sequencing and labor costs depending on genome size, read length, and coverage) and has resulted in the publication of whole mitochondrial genomes for a variety of organisms (Capel et al., 2016; Price et al., 2016; Tisthammer et al., 2016). This approach generally requires a congeneric reference mitochondrial genome, but relatively distant relatives can be used to 'seed' the reference assembly (Price et al., 2016), and *de novo* assembly with no reference can often result in large mitochondrial contigs or even entire mitochondrial genomes.

As expected based on the slow rate in cnidarians, the mitochondrial genomes were remarkably conserved, with very few polymorphic and informative sites (Table 2). For example, a mounding coral from GenBank (collected from Okinawa Japan identified as *P. c.f. okinawensis*) differed by only 8 bp (0.0004%) from Hawaiian *P. compressa* and *P. lobata* samples. It is very likely that this sample was misidentified, which frequently occurs with *Porites*. This sample and several *P. lobata* and *P. compressa* samples had identical or nearly identical haplotypes over the entire mitochondrial genome, it further illustrates that the error rates from using the reference assembly/consensus approach are relatively low. The mitochondrial genome trees resolved outgroup taxa with strong bootstrap support, however they lacked any resolution within the *P. lobata*/*P. compressa* species complex. Further, several strongly supported clades within the species complex conflicted with the predominantly windward or leeward island clades observed in both the holobiont and coral transcriptomic data subsets (Fig. 7A and B). Discordance between nuclear and mitochondrial loci (mito-nuclear discordance) is often associated with hybridization in animals, particularly across hybridization zones as revealed by biogeographic surveys (e.g., Toews and Brelsford, 2012; Bowen et al., 2016). In this case, larger scale geographic sampling of several nuclear and mitochondrial markers would be needed to test this hypothesis.

The holobiont and 'coral transcriptomic' datasets, on the other hand, were remarkably concordant (Fig. 6). The coral transcriptomic dataset represents reads that map to putative protein coding genes that are orthologous among Scleractinains and other Anthozoans (Bhattacharya et al., 2016). The strong concordance between these holobiont and coral transcriptomic datasets likely indicates that the coral protein coding genes provide a major contribution to the phylogenetic signal in the dataset, and that this signal is not overcome by any potential conflict with non-coding regions or other components of the coral holobiont. Further study of this species complex may identify geographic hybridization zones, or identify specific genes that differentiate between distinctive branching and mounding morphologies, and provide insights into the mechanisms behind the maintenance of these distinct morphological forms in sympatry. Improved reference sequences (e.g., reference genomes for *Porites*, or for *Symbiodinium* clade C) would further improve the identification of individual loci. This study identified relatively few *Symbiodinium* loci, which may be due to a lack of closely related reference sequences, or because *Symbiodinium* DNA may be less abundant than coral DNA as a result of extraction methods. The few loci that mapped to *Symbiodinium* reference sequences provided some resolution between outgroup taxa which may indicate some degree of host/symbiont coevolution; however, there was no resolution between the *P. lobata*/*P. compressa* morphospecies complex which indicates that the morphological differences are unlikely to be due to divergent varieties of *Symbiodinium* (Fig. S1).

De-novo assembly and BLAST searches provided insights into the composition of the holobiont dataset, indicating that the majority of sequences were likely to be coral with rare sampling of diverse taxa that might be associated with the coral (Fig. S2). The *de novo* assembly/BLAST approach also yielded several large contigs (putative ribosomal and histone sequences), which were used for reference assemblies against all libraries. *De-novo* assembly is challenging and prone to artefacts, particularly for organisms with large and complex genomes (Schatz et al., 2012). Since RAD data is a reduced genomic approach, coverage can be highly variable, and therefore this approach was treated with caution and these regions were confirmed by PCR and Sanger sequencing. These extended gene regions resolved outgroup taxa with strong support, however they did not resolve patterns within the *P. lobata* species complex (Fig. 7C and D). The ability to recover the majority or entire ribosomal operon from multiple samples is beneficial for a variety of work. An increased proliferation of ribosomal reference sequences from non-modal organisms, such as those curated in the SILVA database (Quast et al., 2013), contribute to an improved understanding of the tree of life, they improve the accuracy of biodiversity and metagenomics studies, and provide insights into the evolution and function of ribosomes.

Although coral morphospecies complexes are nearly ubiquitous, they are also very poorly understood, representing a unique challenge for testing the limits of phylogenetic and population genetic approaches. The coral holobiont further complicates these approaches due to the mixture of multiple genomes present in virtually every sample. This study provides several approaches for parsing metagenomic data for a new level of phylogenetic resolution into a prominent morphospecies complex, providing a foundation for further work in corals. *Porites* is in urgent need of taxonomic revision since it is a ubiquitous and ecologically important reef building coral. Integrated work on *Porites* species boundaries has a wide range of immediate implications; for example: (i)

the identification of discrete morphological characters tied to genetic groups would greatly aid the interpretation of the fossil record (Zlatarski, 2010; Forsman et al., 2015); (ii) mounding *Porites* species are notoriously difficult to distinguish, yet they are a model organism for paleoclimate proxy records (Grottoli and Eakin, 2007); (iii) cryptic *Porites* species have contrasting ecological and reproductive strategies, and responses to bleaching or stress (Boulay et al., 2014); (iv) studies of recent and ongoing speciation contribute to understanding the ecological and evolutionary processes that create and maintain biodiversity (Carlon and Budd, 2002; Bongaerts et al., 2011; Budd et al., 2011); (v) delineating population level variability contributes to an understanding of the ability to adapt to future climate scenarios (Loya et al., 2001; Hoeke et al., 2011; Barshis et al., 2012; Jury and Jokiel, 2016); and (vi) the identification of isolated and rare groups allows for determining appropriate conservation strategies (Forsman et al., 2005; Richards et al., 2008; Brainard et al., 2011).

Author contributions

ZHF and RJT conceived of and designed the research, ZHF, ISSK, KT, MB performed research, RJT and DARE contributed reagents and analytical tools, ZHF and MB analyzed data, ZHF wrote the paper with major contributions, suggestions, and approval from all authors.

Data accessibility

- All raw data is available as NCBI BioProject PRJNA380807
- Final DNA sequence alignments are available as Supplementary material:
  - Mitochondrial genome alignment,
  - rDNA alignment
  - histone alignment
- A text file containing commands for all command-line programs and a list of all parameter files and settings.

Acknowledgements

The authors wish to thank the many colleagues who have contributed to this effort. We especially thank Jim Maragos and Gareth Williams for collecting many of the samples and photographs used in this study. We wish to thank Jon Whitney and Jon Puritz for help and advice with command line pipelines. We thank Mareike Sudek, Amy Eggers and the HIMB core facility for troubleshooting and improvement of library quality. We thank Michael E. Hellberg and two anonymous reviewers for offering editorial suggestions that greatly improved the manuscript. Our greatest gratitude goes to the Pauley foundation and the Edwin W. Pauley summer program for providing the generous support that enabled us to launch this project and for NSF-OA#1416889 and the Seaver Institute to enable us to complete it. This is HIMB contribution #1682 and SOEST #9990.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.03.023.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17, 81–92.

Andrews, K.R., Luikart, G., 2014. Recent novel approaches for population genomics data analysis. Mol. Ecol. 23, 1661–1667.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strmberg, M.P., Marth, G.T., 2011. Bamtools: A C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27, 1691–1692.

Barshis, D.J., Ladner, J.T., Oliver, T.A., Seneca, F.O., Traylor-knowles, N., Palumbi, S.R., 2012. Genomic basis for coral resilience to climate change. PNAS 2012, 1387–1392.

Bhattacharya, D., Agrawal, S., Aranda, M., Baumgarten, S., Belcaid, M., Drake, J.L., Erwin, D., Foret, S., Gates, R.D., Gruber, D.F., Kamel, B., Lesser, M.P., Levy, O., Liew, Y.J., MacManes, M., Mass, T., Medina, M., Mehr, S., Meyer, E., Price, D.C., Putnam, H.M., Qiu, H., Shinzato, C., Shoguchi, E., Stokes, A.J., Tambutté, S., Tchernov, D., Voolstra, C.R., Wagner, N., Walker, C.W., Weber, A.P., Weis, V., Zelzion, E., Zoccola, D., Falkowski, P.G., 2016. Comparative genomics explains the evolutionary success of reef-forming corals. Elife 5, 1–26.

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

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T.G., Wu, C.-H., Xie, D., Suchard, M.a., Rambaut, A., Drummond, A.J., 2013. Beast2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, 1003537.

Boulay, J.N., Hellberg, M.E., Cortés, J., Baums, I.B., Corte, J., Rica, C., Jose, S., 2014. Unrecognized coral species diversity masks differences in functional ecology. Proc. Biol. Sci. 281, 20131580.

Bowen, B.W., Gaither, M.R., DiBattista, J.D., Iacchei, M., Andrews, K.R., Grant, W.S., Toonen, R.J., Briggs, J.C., 2016. Comparative phylogeography of the ocean planet. Proc. Natl. Acad. Sci. 113, 201602404.

Brainard, R.E., Birkeland, C., Eakin, C.M., Mcelhany, P., Miller, M.W., Patterson, M., Piniak, G.A., 2011. Status Review Report of 82 Candidate Coral Species Petitioned Under the U. S. Endangered Species Act. Dep. Commer. NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-27, 530 p. +1 Append. 530.

Brakel, W.H., 1977. Corallite variation in*Porites* and the species problem in corals. 1, p. 457–462.

Budd, A.F., Nunes, F.L.D., Weil, E., Pandolfi, J.M., 2011. Polymorphism in a common Atlantic reef coral (Montastraea cavernosa) and its long-term evolutionary implications. Evol. Ecol., 1–26

Capel, K.C.C., Migotto, A.E., Zilberberg, C., Lin, M.F., Forsman, Z., Miller, D.J., Kitahara, M.V., 2016. Complete mitochondrial genome sequences of Atlantic representatives of the invasive Pacific coral species *Tubastraea coccinea* and *T. tagusensis* (Scleractinia, Dendrophylliidae): Implications for species identification. Gene 590 (2), 270–277.

Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 4, 846–852.

Carlon, D.B., Budd, A.F., 2002. Incipient speciation across a depth gradient in a scleractinian coral? Evolution 56, 2227–2242.

Danecek, P., Auton, A., Abecasis, G., Albers, C.a., Banks, E., DePristo M, M.a., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCF tools. Bioinformatics 27, 2156–2158.

Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., Blaxter, M.L., 2012. Special features of RAD Sequencing data: implications for genotyping. Mol. Ecol. 22, 3151–3164.

Earl, D.a., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361.

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844–1849.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Etter, P.D., Preston, J.L., Bassham, S., Cresko, W., Johnson, E., 2011. Local de novo assembly of RAD paired-end contigs using short sequencing reads. PLoS ONE 6, e18561.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14, 2611–2620.

Eytan, R.I., Hayes, M., Arbour-Reily, P., Miller, M., Hellberg, M.E., 2009. Nuclear sequences reveal mid-range isolation of an imperilled deep-water coral population. Mol. Ecol. 18, 2375–2389.

Flot, J.F., Magalon, H., Cruaud, C., Couloux, A., Tillier, S., 2008. Patterns of genetic structure among Hawaiian corals of the genus Pocillopora yield clusters of individuals that are compatible with morphology. Comptes Rendus – Biol. 331, 239–247.

Forsman, A., 2014. Rethinking phenotypic plasticity and its consequences for individuals, populations and species. Heredity (Edinb). 115, 1–9.

Forsman, A., Forsman, A., Ahnesjo, J., Ahnesjo, J., Caesar, S., Caesar, S., Karlsson, M., Karlsson, M., Ahnesjö, J., Caesar, S., Karlsson, M., 2008. A model of ecological and evolutionary consequences of color polymorphism. Ecology 89, 34–40.

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

Forsman, Z.H., Guzman, H.M., Chen, C., Fox, G.E., Wellington, G.M., 2005. An ITS region phylogeny of Siderastrea (Cnidaria: Anthozoa): is *S. glynni* endangered or introduced? Coral Reefs 24, 343–347.

Forsman, Z.H., Hunter, C., Fox, G., Wellington, G., 2006. Is the ITS region the solution to the "Species Problem" in Corals? Intragenomic variation and alignment permutation in porites, siderastrea and outgroup taxa. Proc 10th Int Coral Reef Symp, vol. 1, pp. 14–23.

Forsman, Z.H., Johnston, E.C., Brooks, A.J., Adam, T.C., Toonen, R.J., 2013. Genetic Evidence for Regional Isolation of Pocillopora Coral From Moorea. Oceanogr. Press, p. 69–71.

Forsman, Z.H., Wellington, G.M., Fox, G.E., Toonen, R.J., 2015. Clues to unraveling the coral species problem: distinguishing species from geographic variation in Porites across the Pacific with molecular markers and microskeletal traits. Peer J. 3, e751.

Fukami, H., Knowlton, N., 2005. Analysis of complete mitochondrial DNA sequences of three members of the Montastraea annularis coral species complex (Cnidaria, Anthozoa, Scleractinia). Coral Reefs 24, 410–417.

Gaither, M.R., Szabó, Z., Crepeau, M.W., Bird, C.E., Toonen, R.J., 2011. Preservation of corals in salt-saturated DMSO buffer is superior to ethanol for PCR experiments. Coral Reefs 30, 329–333.

Garrison, E., Marth, G., 2012. Haplotype-based Variant Detection from Short-Read Sequencing. arXiv Prepr. Available from: <1207.3907 9>.

Gray, S.M., McKinnon, J.S., 2007. Linking color polymorphism maintenance and speciation. Trends Ecol. Evol. 22, 71–79.

Grottoli, A.G., Eakin, C.M., 2007. A review of modern coral δ18O and Δ14C proxy records. Earth-Sci. Rev. 81, 67–91.

Guindon, S., Delsuc, F., Dufayard, J.F., Gascuel, O., 2009. Estimating maximum likelihood phylogenies with PhyML. Methods Mol. Biol. 537, 113–137.

Hellberg, M.E., 2006. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. BMC Evol. Biol. 6, 24.

Hellberg, M.E., Prada, C., Tan, M.H., Forsman, Z.H., Baums, I.B., 2016. Getting a grip at the edge: recolonization and introgression in eastern Pacific Porites corals. J. Biogeogr. 43, 2147–2159.

Herrera, S., Shank, T.M., 2015. RAD Sequencing Enables Unprecedented Phylogenetic Resolution and Objective Species Delimitation in Recalcitrant Divergent Taxa. bioRxiv 19745.

Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. PLoS ONE 9, e93975.

Hoeke, R.K., Jokiel, P.L., Buddemeier, R.W., Brainard, R.E., 2011. Projected changes to growth and mortality of Hawaiian corals over the next 100 years. PLoS ONE 6, e18038.

Huang, D., Meier, R., Todd, P.A., Chou, L.M., 2008. Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J. Mol. Evol. 66, 167–174.

Jameson, S.C., 1997. Morphometric analysis of the Poritidae (Anthozoa: Scleractinia) off Belize. 2. Proc 8th Int. Coral Reef Symp. Panama, vol. 2, pp. 1591–1596.

Jameson, S.C., Cairns, S.D., 2012. Neotypes for Porites porites (Pallas, 1766) and Porites divaricata Le Sueur, 1820 and remarks on other western Atlantic species of Porites (Anthozoa: Scleractinia). Proc. Biol. Soc. Washingt. 125, 189–207.

Jury, C.P., Jokiel, P.L., 2016. Climate Change, Ocean Chemistry, and the Evolution of Reefs Through Time. Coral Reefs at the Crossroads. Springer, Netherlands, pp. 197–223.

Keshavmurthy, S., Yang, S.-Y., Alamaru, A., Chuang, Y.-Y., Pichon, M., Obura, D., Fontana, S., De Palmas, S., Stefani, F., Benzoni, F., MacDonald, A., Noreen, A.M.E., Chen, C., Wallace, C.C., Pillay, R.M., Denis, V., Amri, A.Y., Reimer, J.D., Mezaki, T., Sheppard, C., Loya, Y., Abelson, A., Mohammed, M.S., Baker, A.C., Mostafavi, P.G., Suharsono, B.A., Chen, C.A., 2013. DNA barcoding reveals the coral "laboratory-rat", Stylophora pistillata encompasses multiple identities. Sci. Rep. 3, 1–7.

Li, H., 2013. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv Prepr. arXiv 0:3.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.Subgroup 1000 Genome Project Data Processing, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Lin, M., Luzon, K., Licuanan, Y., Ablan-Lagman, M., Chen, C., 2011. Seventy-four universal primers for characterizing the complete mitochondrial genomes of scleractinian corals (Cnidaria; Anthozoa). Zool. Stud. 50, 513–524.

Link, H., 1807. Bescheibung der Naturaleine. Sammlungen der Universaitat Rostock 3, 161–165.

Lischer, H.E.L., Excoffier, L., 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298–299.

Loya, Y., Sakai, K., Yamazato, Sambali., van Woesik, R., 2001. Coral bleaching: the winners and the losers. Ecol. Lett. 4, 122–131.

Luck, D., Forsman, Z.H., Toonen, R., 2013. Polyphyly and hidden species among Hawai'i's dominant mesophotic coral genera, Leptoseris and Pavona (Scleractinia: Agariciidae). PeerJ, e132.

McGinnis, S., Madden, T.L., 2004. BLAST: At the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res., 32

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

Medina, M., Collins, A.G., Takaoka, T.L., Kuehl, J.V., Boore, J.L., 2006. Naked corals: skeleton loss in Scleractinia. Proc. Natl. Acad. Sci. USA 103, 9096–9100.

Milne, I., Stephen, G., Bayer, M., Cock, P.J., Pritchard, L., Cardle, L., Shawand, P.D., Marshall, D., 2013. Using tablet for visual exploration of second-generation sequencing data. Brief. Bioinform. 14, 193–202.

Muko, S., Kawasaki, K., Sakai, K., Takasu, F., Shigesada, N., 2000. Morphological plasticity in the coral porites sillimaniani and its adaptive significance Soyoka Muko, Kohkichi Kawasaki, Kazuhiko Sakai. Bull. Mar. Sci. 66, 225–239.

Odorico, D.M., Miller, D.J., 1997. Variation in the ribosomal internal transcribed spacers and 5.8S rDNA among five species of Acropora (Cnidaria; Scleractinia): patterns of variation consistent with reticulate evolution. Mol. Biol. Evol. 14, 465–473.

van Oppen, M.J.H., Hislop, N.R., Hagerman, P., Miller, D.J., 1999. Gene content and organization in a segment of the mitochondrial genome of the scleractinian coral acropora tenuis: major differences in gene order within the anthozoan subclass Zoantharia. Mol. Biol. Evol. 16, 1812–1815.

van Oppen, M.J.H., Wörheide, G., Takabayashi, M., 2000. Nuclear Markers in Evolutionary and Population Genetic Studies of Scleractinian Corals and Sponges 1, pp. 131–138.

Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8, e1002967.

Pigliucci, M., 2005. Evolution of phenotypic plasticity: where are we going now? Trends Ecol. Evol. 20, 481–486.

Pinzón, J.H., Sampayo, E., Cox, E., Chauka, L.J., Chen, C.A., Voolstra, C.R., LaJeunesse, T. C., 2013. Blind to morphology: genetics identifies several widespread ecologically common species and few endemics among Indo-Pacific cauliflower corals (*Pocillopora*, Scleractinia). J. Biogeogr. 40, 1595–1608.

Prada, C., DeBiasse, M.B., Neigel, J.E., Yednock, B., Stake, J.L., Forsman, Z.H., Baums, I. B., Hellberg, M.E., 2014. Genetic species delineation among branching Caribbean Porites corals. Coral Reefs 33, 1019–1030.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909.

Price, M., Forsman, Z., Knapp, I., Hadfield, M., Toonen, R., 2016. The complete mitochondrial genome of *Achatinella mustelina* (Gastropoda: Pulmonata: Stylommatophora). Mitochondrial DNA 1, 183–185.

Puritz, J.B., Hollenbeck, C.M., Gold, J.R., 2014a. DDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2, e431.

Puritz, J.B., Matz, M.V., Toonen, R.J., Weber, J.N., Bolnick, D.I., Bird, C.E., 2014b. Demystifying the RAD fad. Mol. Ecol. 23, 5937–5942.

Purvis, A., 2008. Phylogenetic approaches to the study of extinction. Annu. Rev. Ecol. Evol. Syst. 39, 301–319.

Purvis, A., Gittleman, J.L., Cowlishaw, G., Mace, G.M., 2000. Predicting extinction risk in declining species. Proc. Biol. Sci. 267, 1947–1952.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D595.

Reitzel, A.M., Herrera, S., Layden, M.J., Martindale, M.Q., Shank, T.M., 2013. Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. Mol. Ecol. 22, 2953–2970.

Richards, Z.T., van Oppen, M.J.H., Wallace, C.C., Willis, B.L., Miller, D.J., 2008. Some rare Indo-Pacific coral species are probable hybrids. PLoS ONE 3, e3240.

Richmond, R.H., Hunter, C.L., 1990. Reproduction and recruitment of corals: comparisons among the Caribbean, the Tropical Pacific, and the Red Sea. Mar. Ecol. Prog. Ser. 60, 185–203.

Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7, e33394.

Rundell, R.J., Price, T.D., 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends Ecol. Evol. 24, 394–399.

Schatz, M.C., Witkowski, J., McCombie, W.R., 2012. Current challenges in de novo plant genome sequencing and assembly. Genome Biol. 13, 243.

Schmidt-Roach, S., Lundgren, P., Miller, K.J., Gerlach, G., Noreen, A.M.E., Andreakis, N., 2012. Assessing hidden species diversity in the coral *Pocillopora damicornis* from Eastern Australia. Coral Reefs 32, 161–172.

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

Shinzato, C., Inoue, M., Kusakabe, M., 2014. A snapshot of a coral "Holobiont": a transcriptome assembly of the scleractinian coral, porites, captures a wide variety of genes from both the host and symbiotic zooxanthellae. PLoS ONE 9, e85182.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Stat, M., Baker, A.C., Bourne, D.G., Correa, A.M.S., Forsman, Z., Huggett, M.J., Pochon, X., Skillings, D., Toonen, R.J., van Oppen, M.J.H., Gates, R.D., 2012. Molecular delineation of species in the coral holobiont. Adv. Mar. Biol. 63, 1–65.

Storlazzi, C.D., Brown, E.K., Field, M.E., Rodgers, K., Jokiel, P.L., 2004. A model for wave control on coral breakage and species distribution in the Hawaiian Islands. Coral Reefs 24, 43–55.

Takahashi, T., Nagata, N., Sota, T., 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. Mol. Phylogenet. Evol. 80, 137–144.

Tisthammer, K.K.H., Forsman, Z.Z.H., Sindorf, V.L., Massey, T.L., Bielecki, C.R., Toonen, R.J.R., 2016. The complete mitochondrial genome of the lobe coral Porites lobata (Anthozoa:Scleractinia) sequenced using ezRAD. Mitochondrial DNA, 1–3.

Toews, D.P.L., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21, 3907–3930.

Toonen, R., Puritz, J.J., Forsman, Z., Whitney, J., Fernandez-Silva, I., Andrews, K., Bird, C., 2013. EzRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ 1, e203.
Toonen, R.J., Pawlik, J.R., 1994. Foundations of gregariousness. Nature 370, 511–512.
Toonen, R.J., Pawlik, J.R., 2001. Foundations of gregariousness: a dispersal polymorphism among the planktonic larvae of a marine invertebrate. Evolution (NY) 55, 2439–2454.
Vaughan, T., 1907. Recent madreporaria of the Hawaiian Islands and Laysan. No. 59. Govt. print. off.
Veron, J., Stafford-Smith, M.G., 2000. Corals of the World. Australian Institute of Marine Science, Townsville.
Veron, J.E.N., 1995. Corals in Space and Time: The Biogeography and Evolution of the Scleractinia. Cornell University Press.
Vollmer, S., Palumbi, S.R., 2004. Testing the utility of internally transcribed spacer sequences in coral phylogenetics. Mol. Ecol. 13, 2763–2772.
Voolstra, C.R., Worheide, Gigacos, Lopez, J.V., 2017. Advancing genomics through the Global Invertebrate Genomics Alliance (GIGA). Invertebr. Syst. 31 (1), 1–7.
Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2012. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22, 787–798.
Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620.
Zlatarski, V.N., 2010. Palaeobiological perspectives on variability and taxonomy of scleractinian corals. Palaeoworld 19, 333–339.