Phylogenomics of *Porites* from the Arabian Peninsula

Tullia I. Terraneo [a,b,*], Francesca Benzoni [a], Roberto Arrigoni [c,d,a], Andrew H. Baird [b],
Kiruthiga G. Mariappan [a], Zac H. Forsman [e], Michael K. Wooster [a], Jessica Bouwmeester [f],
Alyssa Marshell [g], Michael L. Berumen [a]

[a] *Red Sea Research Centre, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*
[b] *ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville 4811, QLD, Australia*
[c] *European Commission, Joint Research Centre (JRC), Ispra, Italy*
[d] *Department of Biology and Evolution of Marine Organisms (BEOM), Stazione Zoologica Anton Dohrn Napoli, Villa Comunale, 80121 Napoli, Italy*
[e] *Hawaii Institute of Marine Biology, Kaneohe 96744, HI, USA*
[f] *Smithsonian Conservation Biology Institute, Front Royal 22630, VA, USA*
[g] *Department of Marine Science and Fisheries, College of Agricultural and Marine Sciences, Sultan Qaboos University, Muscat, Oman*

A R T I C L E   I N F O

A B S T R A C T

The advent of high throughput sequencing technologies provides an opportunity to resolve phylogenetic relationships among closely related species. By incorporating hundreds to thousands of unlinked loci and single nucleotide polymorphisms (SNPs), phylogenomic analyses have a far greater potential to resolve species boundaries than approaches that rely on only a few markers. Scleractinian taxa have proved challenging to identify using traditional morphological approaches and many groups lack an adequate set of molecular markers to investigate their phylogenies. Here, we examine the potential of Restriction-site Associated DNA sequencing (RADseq) to investigate phylogenetic relationships and species limits within the scleractinian coral genus *Porites*. A total of 126 colonies were collected from 16 localities in the seas surrounding the Arabian Peninsula and ascribed to 12 nominal and two unknown species based on their morphology. Reference mapping was used to retrieve and compare nearly complete mitochondrial genomes, ribosomal DNA, and histone loci. *De novo* assembly and reference mapping to the *P. lobata* coral transcriptome were compared and used to obtain thousands of genome-wide loci and SNPs. A suite of species discovery methods (phylogenetic, ordination, and clustering analyses) and species delimitation approaches (coalescent-based, species tree, and Bayesian Factor delimitation) suggested the presence of eight molecular lineages, one of which included six morphospecies. Our phylogenomic approach provided a fully supported phylogeny of *Porites* from the Arabian Peninsula, suggesting the power of RADseq data to solve the species delineation problem in this speciose coral genus.

1. Introduction

Understanding of species boundaries and evolutionary relationships among organisms is a key goal in biology. Recent advances in molecular and computational techniques have revolutionized our understanding of the systematics of numerous organisms (Faircloth et al., 2012; Puritz et al., 2014). Restriction-sites-associated fragmentation of genomic DNA (RADseq) is an effective method for harnessing the power of high throughput sequencing technologies (NGS) (Baird et al., 2008), providing genomic-wide data and a large number of homologous markers for non-model organisms (Pante et al., 2015). RADseq is currently the most widely used genomic approach for high-throughput single nucleotide polymorphism (SNP) discovery and genotyping in non-model organisms (Pante et al., 2015; Forsman et al., 2017). It allows for the simultaneous discovery and genotyping of thousands of polymorphic loci throughout the genome, without requiring any prior genomic resources for the study taxon (Baxter et al., 2011). Closely related species share orthologous restriction sites, thus RADseq is generally used to infer recent evolutionary history (Harvey et al., 2016; Gottscho et al., 2017). However, it has also been used to clarify more

distant evolutionary relatedness going back to the Paleocene (Rubin et al., 2012; Eaton and Ree, 2013; Cariou et al., 2013; Hipp et al., 2014).

Anthozoans are an ancient and ubiquitous group of benthic marine invertebrates, for which high levels of morphological variation, phenotypic plasticity, and few available orthologous conserved markers, have hindered a clear understanding of evolutionary history (Prada et al., 2008; Paz-García et al., 2015; Herrera and Shank, 2016; Quattrini et al., 2018). The systematics of the class has historically been based primarily on morphology, which is known to be highly variable and phenotypically plastic (Todd, 2008). Molecular studies have uncovered widespread homoplasy and convergent evolution of morphological characters within the subclasses Hexacorallia, Octocorallia, and Ceriantharia (Fukami et al., 2004; Stampar et al., 2014; Ament-Velásquez et al., 2016). The use of molecular barcoding has often proved unsuccessful because of a slow rate of evolution of mitochondrial DNA (Hellberg, 2006; Huang et al., 2008), the presence of divergent paralogous copies in the nuclear ribosomal DNA (Odorico and Miller, 1997; Sánchez and Dorado, 2008), and the presence of few phylogenetically informative nuclear genes discovered so far (Concepion et al., 2008; McFadden et al., 2010). Incomplete lineage sorting, hybridization, and topology discordance between gene and species trees in a plethora of cnidarians have also been hypothesized to affect the use of molecular markers to infer meaningful phylogenies (Mcfadden and Hutchinson, 2004; Ament-Velásquez et al., 2016; Terraneo et al., 2016; Pratlong et al., 2017). Recently, RADseq has been successfully applied to clarify species boundaries and identify hybridization of octocoral genera *Chrysogorgia* Duchassaing and Michelotti, 1864, *Paragorgia* Milne Edwards, 1857*, and *Sinularia* May 1898 (Pante et al., 2015; Herrera and Shank, 2016; McFadden et al., 2017; Quattrini et al., 2019), and in attempts to clarify species boundaries within the scleractinians *Galaxea* Oken, 1815, *Montipora* Blanville, 1830, *Leptastrea* Milne Edwards and Haime, 1849, *Pocillopora* Lamarck, 1816, and *Porites* Link, 1807 (Combosch and Vollmer, 2015; Forsman et al., 2017, Dimond et al., 2017; Johnston et al., 2017; Cunha et al., 2019; Arrigoni et al., 2020; Forsman et al., 2020).

The scleractinian genus *Porites* with 190 nominal species is the second most speciose hermatypic coral (Hoeksema and Cairns, 2019) and represents a major component of coral communities worldwide (Bellwood et al., 2004). Nevertheless, species boundaries and evolutionary relationships within *Porites* remain mostly unresolved (Forsman et al., 2009, 2017; Terraneo et al., 2018a,b). Several of the morphological traits traditionally used to separate species in *Porites* have been proved to be affected by stasis and convergent evolution, and informative morphological synapomorphies have yet to be evaluated on the whole genus (Smith et al., 2007; Forsman et al., 2015; Tisthammer and Richmond, 2018). So far, multi-locus phylogenetic reconstructions have revealed the presence of undescribed species but also identify unresolved species complexes (Forsman and Birkeland, 2009; Forsman et al., 2009; Benzoni and Stefani, 2012; Prada et al., 2014; Hellberg et al., 2016; Terraneo et al., 2019a). Patterns of introgression have also been discovered among different species in the Eastern Pacific and Hawai'i (Hellberg et al., 2016; Forsman et al., 2017), highlighting gaps in our understanding of the evolution and biogeography of the genus.

The seas around the Arabian Peninsula, comprising the Red Sea, the Gulf of Aden, the Gulf of Oman, and the Arabian Gulf, are hypothesized to be a biodiversity hotspot for *Porites* (Sheppard and Sheppard, 1991; Veron et al., 2015). Indeed, based on morphological identifications, 26 species of *Porites* have been reported around Arabian Peninsula (Veron, 2000; Terraneo et al., 2019a; Berumen et al., 2019), many more than in similarly sized coral-rich regions of the West Pacific such as the Great Barrier Reef and Japan (Veron, 2000). However, recent work integrating two molecular loci and corallite level micromorphology, showed that in the Red Sea and Gulf of Aden 10 morphologically defined species of *Porites* actually belong to six genetic lineages (Terraneo et al., 2019a,b).

In this study, we provide the phylogeny of *Porites* from the Arabian Peninsula. We reconstruct molecular phylogenies from 54,108 SNPs

generated with *de novo* assembly, 96,986 SNPs mapped to *Porites* transcriptome, nearly complete mitochondrial genomes, nuclear ribosomal DNA, and histone regions. We apply analyses of genetic clustering and ordination and coalescent-based species delimitations. Finally, we discuss evolutionary and biogeographical hypothesis of *Porites* in this region.

2. Materials and methods

2.1. Collection and identification

A total of 126 *Porites* colonies were collected from 16 sites in the seas around the Arabian Peninsula, between 2013 and 2017 (Table S1 and Figs. S1, S2). Each coral colony was imaged underwater and a 5 × 5 cm fragment was collected with hammer and chisel (Fig. S2). Tissue samples (<1 cm) from the surface of each colony were preserved in 98% ethanol or CHAOS solution and stored for genomic analyses. The specimens were then bleached with sodium hypochlorite for 24 h and air dried for morphological examination (see below). Specimens collected in Saudi Arabia are deposited at King Abdullah University of Science and Technology (KAUST), Saudi Arabia. Specimens collected in Qatar are deposited at Qatar University (Qatar), while specimens collected in Oman are deposited at Sultan Qaboos University (Oman). Examined material from the other localities is housed at University of Milano-Bicocca (UNIMIB), Italy. Specimens were imaged with a Leica M80 microscope equipped with a Leica IC80HD camera, and assigned to species following a comparison with original descriptions and type material. The following features were considered: corallite diameter, wall thickness, fusion pattern of the ventral triplet, number of pali, number of denticles, presence or absence of the columella, and presence or absence of the coenosteum (for glossary see Budd et al., 2012).

2.2. DNA extraction and quantification

Genomic DNA was extracted using DNeasy® Blood & Tissue Kit (Qiagen, Hilden, Germany) for samples stored in ethanol or using a phenol-chloroform-based method for samples stored in CHAOS solution. Extracted DNA was quantified with the Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, Waltham, MA, USA) using Qubit® fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.3. PCR amplification and sequencing

The mitochondrial Control Region (mtCR) was amplified using Polymerase Chain Reaction (PCR) and the primers zpsRNSf (5′ – AGC AGA CGC GGT GAA ACT TA – 3′) and zpCOX3r (5′ – GCC AAA GTA ACA GTA CCC CC – 3′) (Terraneo et al., 2019b). Amplifications were conducted in a 15 μl PCR volume, composed of 0.2 μM each primer, 1X Multiplex PCR Master Mix (Qiagen, Hilden, Germany), and <5 ng DNA. PCR products were purified by adding 1.5 μl Illustra ExoStar (GE Healthcare, Buckinghamshire, UK), incubated at 37 °C for 60 min, followed by 85 °C for 15 min, and directly sequenced in forward and reverse directions using an ABI 3730xl DNA Analyzer (Applied Biosystems, Carlsbad, CA, USA). Forward and reverse sequences were assembled using Geneious® v.10.1.3 (Biomatters Ltd. Auckland, New Zealand). The newly produced sequences were integrated with mtCR sequences from Terraneo et al. (2019a, b). Multiple alignment was performed using the E-INS-i option in MAFFT v.7 (Katoh and Standley, 2013). Newly produced mtCR sequences were deposited in GenBank (https://www.ncbi.nml.nih.gov/genbank/) (Accession numbers: from MW412256 to MW412330).

2.4. Restriction enzyme digestion and ezRAD libraries preparation

We followed protocols by Toonen et al. (2013) and Knapp et al. (2016) for DNA digestion and ezRAD library preparation. In detail, each sample was digested using frequent cutter restriction enzymes MboI and

Sau3AI (New England BioLabs, Ipswich, MA, USA) to cleave sequences at GATC cut sites (Toonen et al., 2013). Digestions were performed in a 50 μl reaction volume consisting of 43 μl dsDNA (about 1.2–1.3 μg), 5 μl of Cutsmart Buffer (New England BioLabs, Ipswich, MA, USA), and 1 μl of each undiluted restriction enzyme, under the following thermocycler profile: 37 °C for 3 h followed by 65 °C for 20 min. Digested samples were cleaned using Agencourt AMPure XP beads (Beckmann Coulter, Danvers, MA, USA) at a 1:1.8 (DNA:beads) ratio following the standard protocol. The concentration of cleaned digests was checked with Qubit® Fluorometer 3.0 (Thermo Fisher Scientific, Waltham, MA, USA). A total amount of 200 ng of each digested DNA sample was used for the library preparation using the TruSeq® Nano DNA Library prep kit (Illumina, San Diego, CA, USA), following the manufacture protocol. Libraries were size-selected at 350 bp following the manufacture's protocol and the protocol by Knapp et al. (2016), and passed through two quality control steps, *i.e.* bioanalyzer and qPCR, to check size and concentration, respectively. Finally, ezRAD libraries were normalized and combined to two pools of 65 libraries each. Each libraries pool was run in a single 150 bp paired-end lane on Illumina HiSeq 4000 System at KAUST Genomics Core Lab (Thuwal, Saudi Arabia). Sample information with sequenced lengths and number of reads are presented in Table S1.

2.5. ezRAD data processing

The Illumina raw data consisted of ~339 million 150 bp reads. The raw data has been submitted to NCBI SRA (https://www.ncbi.nlm.nih.gov/sra), under the project number PRJNA714198. Samples were de-multiplexed using their unique barcode and adapter sequences under the Illumina pipeline bcl2fastq/2.17.1.14, effectively removing reads that lacked identifiable barcode pairs. An average of 2.6 million reads per individual (N = 126) were trimmed, assembled, and genotyped using dDocent v.2.25 (Puritz et al., 2014) (Table S2).

Two assembly strategies were used and compared: *de novo* assembly and reference-based assembly. For the *de novo* assembly dataset, ~339 million reads were placed in a folder as *.F.fq.gz and *.R.fq.gz and trimmed reads were placed as *.R1.fq.gz and *.R2.fq.gz, respectively. Reads were merged using PEAR v.0.9.6 (Zhang et al., 2013) and assembled using BWA v.0.7.15 (Li and Durbin, 2009). SNPs were identified using FreeBayes (Garrison and Marth, 2012) with settings mentioned in Forsman et al., (2017). Since no public genome of *Porites* was availble when these analyses were initally performed, for the reference-based assembly dataset, the trimmed reads were first mapped to the transcriptome of *P. lobata* obtained from Bhattacharya et al. (2016) using Bowtie v.2 2.3.4 (Langmead and Salzberg, 2012). Subsequently, the mapped reads were converted to bam format using SAMtools v.1.6 (Li et al., 2009) and then converted to fastq using BEDtools v.2.26.0 (Quinlan and Hall, 2010). These binned files were then copied to a separate folder and genotyped using dDocent v.2.25 (Puritz et al., 2014). In short, the reads were trimmed using Trimmomatic v.0.36 (Bolger et al., 2014), merged using PEAR v.0.9.6 (Zhang et al., 2013) and aligned to the reference transcriptome again using BWA v.0.7.15 (Li and Durbin, 2009) under the settings -t 16 -a -M -T 10 –R. SNPs were finally identified using FreeBayes (Garrison and Marth, 2012), as mentioned in Forsman et al. (2017). The unfiltered *de novo* assembly dataset included 3102 loci, while the coral transcriptome had 21,062 loci. Thus, the chance of reads mapping and the detection of SNPs was higher in the binned case. Moreover, we compared the dataset generated with *de novo* assembly strategy to BLAST searches to coral and Symbiodiniaceae genomes and transcriptomes in order to identify the loci composition of this dataset. We used the online tool SequenceServer v.2.0.0.rc7 (Priyam et al., 2019) available at reefgenomics.org/blast/ and mapped the dataset to the genome of *Porites lutea* and several Symbiodiniaceae genome and transcriptomes with a threshold of e values lower than 1e-22. The two resulting VCF files were further filtered using VCFtools v.0.1.16 (Danecek et al., 2011). To examine the sensitivity of the phylogenetic inference to the filtering process, we generated two filtered supermatrices for the two datasets. We obtained the "refbased-max" and the "denovo-max" supermatrices using the following filter options: mean depth = 3, max missing data = 50%, and minimum distance between SNPs = 5. Conversely, we generated the "refbased-min" and the "denovo-min" supermatrices under mean depth = 10, max missing data = 5%, and minimum distance between SNPs = 10 (for a complete overview regarding the filtering options available using VCFtools v.0.1.16 please refer to http://vcftools.sourceforge.net/man_latest.html also available at https://vcftools.github.io/man_0112a.html). Haplotypes were then called and filtered for complex loci, potential paralogs, missing data, and sequencing errors using the rad_haplotyper v.1.1.8 pipeline (https://github.com/chollenbeck/rad_haplotyper; Willis et al., 2017). PGDspider v.2.1.1.5 (Lischer and Excoffier, 2011) was used to convert the dataset to the required file types for further analysis. The "refbased-max" supermatrix contained 6452 loci and 96,986 SNPs, the "denovo-max" 2041 loci and 54,108 SNPs, the "refbased-min" 367 loci and 4140 SNPs, the "denovo-min" 719 loci and 10,918 SNPs. Each of the resulting four concatenated loci supermatrix was analyzed in RAxML-HPC2 v8.0 (Stamatakis, 2014) for maximum likelihood (ML) phylogenetic inference. We applied the GTR + GAMMA substitution model and the branch support was assessed by 1000 bootstrap replicates. ML analyses were run on the CIPRES Science Gateway (Miller et al., 2010).

2.6. Reference assemblies and phylogenetic analyses of mitochondrial genomes, histone, and rDNA regions

One of the main advantages of ezRAD among the other RADseq techniques is that it provides a mix of breadth and depth of coverage (Toonen et al., 2013; Stobie et al., 2019). While depth of coverage is important to accurately genotype SNPs, breadth of coverage can result in very long contigs, resulting in the resolution of the complete or a large percentage of the mitochondrial genomes and other multicopy gene regions such as histones and ribosomes. Therefore, we used reference mapping against previously published reference sequences to acquire and compare from each library nearly complete mitochondrial genome (mtGenome), histone region (histone), and nuclear ribosomal DNA array (rDNA, including 18S, ITS1, 5.8S, ITS2, and 28S regions). We used the complete mtGenome of *P. lobata* (NC030186, 18,647 bp) and the nearly complete histone (5301 bp) and rDNA (6629 bp) sequences of *P. superfusa* obtained by Forsman et al. (2017) as reference. Trimmed reads were aligned to the three reference sequences using Bowtie v.2.3.4 (Langmead and Salzberg, 2012) in –fast-local mode. Aligned reads were converted to bam and indexed using SAMtools v.1.6 (Li et al., 2009), and the consensus sequences were identified using SAMtools mpileup combined with Vcfutils.pl.

We aligned mtGenome, histone, and rDNA sequences using MAFFT v.7 (Katoh and Standley, 2013) (all alignment data are available upon request to the corresponding author). We determined the optimal among-gene partitioning scheme and model choice in PartitionFinder v.2 (Lanfear et al., 2012) under the Bayesian Information Criterion (BIC). The mtGenomes were partitioned according to the genes and considering all intergenic regions as a single partition, with genes further partitioned according to the codon position. The rDNA dataset was partitioned in five partitions (18S, ITS1, 5.8S, ITS2, and 28S), the histone dataset was partitioned by genes and codon position. Phylogenetic relationships based on these three datasets were inferred using ML. *Porites superfusa* Gardiner, 1898, a basal and highly divergent species from the Central Pacific, was selected as outgroup (Forsman et al., 2017). ML trees were inferred with RAxML-HPC2 v8.0 (Stamatakis, 2014), using the GTR + GAMMA model of nucleotide substitution. Node support was assessed using 1000 bootstrap replicates. Analyses were run on the CIPRES Science Gateway (Miller et al., 2010). The mtCR dataset was analyzed following the same criteria and methods used for mtGenome, histone, and rDNA datasets. Finally, the ITS region (ITS1, 5.8S, ITS2) was extracted from the rDNA alignment, in order to compare the

newly produced data with Forsman et al. (2009). Sequence data from Terraneo et al. (2019a, b) was used when the we could not extract high quality sequences from the rDNA reads mapped to the *P. superfusa* rDNA, as the same samples were analyzed in the current project, as well as in Terraneo et al. (2019a, b). Phylogenetic relationships among species were assessed comparing a subset of 30 ITS regions sequences from our dataset, 17 sequences downloaded from GenBank from Terraneo et al. (2019a, b) and 192 sequences downloaded from GenBank from Forsman et al. (2009). Sequence were aligned using MAFFT as previously described. The ML tree was inferred using RAxML-HPC2 v.8.0 (Stamataki, 2014), using the GTRCAT model of nucleotide substitution. Node support was assessed using 1000 bootstrap replicates. Analyses were run on the CIPRES Science Gateway (Miller et al., 2010).

2.7. Ordination and clustering analyses

The VCF file including refbased SNPs obtained from the dDocent v.2.25 pipeline (Puritz et al., 2014) was further filtered using VCFtools v.0.1.16 (Danecek et al., 2011) to create additional filtered datasets that contained different number of SNPs and various levels of filtering options. In particular, SNPs were filtered based on different values of mean depth (–min-meanDP as 3, 5, and 10), missing data (–max-missing-count as 5%, 20%, and 50%), and minimum distance between SNPs (–thin as 5, 10, and 300), generating a total of 27 different reference-based assembly SNPs filtered datasets. The most "relaxed" dataset (–min-meanDP 3, –max-missing-count 50%, –thin 5) included a total of 96,986 SNPs, whereas the most "stringent" dataset (–min-meanDP 10, –max-missing-count 5%, –thin 300) contained a total of only 343 SNPs.

These datasets were analyzed by means of ordination and clustering analyses without any *a priori* hypotheses about individual assignment in order to evaluate the effects of filters and to guide subsequent species delimitation analyses. First, a Principal Components Analysis (PCA) was conducted using Plink v.1.9 (Purcell et al., 2007). The main benefit of PCA is its ability to detect data structure without the computational burden of Bayesian clustering algorithms and the absence of assumptions about the underlying population genetic model. Second, Admixture v.1.23 (Alexander et al., 2009) was used to detect the genetic structure among the analyzed samples. This maximum likelihood-based program implements an underlying population genetic model similar to Structure (Pritchard et al., 2000). While both programs assign individuals into clusters using population allele frequencies and ancestry proportions, Admixture has the added benefit of a fast-numerical optimization algorithm to decrease computational time while avoiding problems with MCMC convergence. We used the cross-validation procedure to select the optimal K value (Alexander et al., 2009), testing values (K) ranging from 2 to 12. Moreover, following Gowen et al. (2014), starting from the optimal K recovered, we included subsequent Admixture analyses on smaller clusters of individuals, until the analyses did not reveal further clustering.

2.8. Species delimitation and species tree inference

We used Bayes Factor Delimitation (BFD*) to rank species delimitation models in a multispecies coalescent framework (Leaché et al., 2014). Briefly, BFD* consists of running SNAPP analyses (Bryant et al., 2012) on models with different numbers of species and assignments of individuals to species, estimating the marginal likelihood of each model, and ranking model fit among runs by comparing Bayes factors (BF). The BFD* approach uses path sampling to estimate the marginal likelihood (MLE) of a population divergence model directly from SNPs data (without integrating over gene trees) and has been shown to be robust to a relatively large amount of missing data (Leaché et al., 2014), being especially suited for RADseq data. We tested the following five models: (A) one species; (B) current morphology-based taxonomy, including *P. annae* Crossland, 1952, *P. columnaris* Klunzinger, 1879, *Porites* sp2, *P. farasani* Benzoni and Terraneo, 2019, *P. fontanesii* Benzoni and

Stefani, 2012, *P*. *hadramauti* Benzoni and Terraneo, 2019, *P. harrisoni* Veron, 2000, *P. lobata* Dana, 1846, *P. lutea* Milne Edwards and Haime, 1851, *P. monticulosa* Dana, 1846, *P. somaliensis* Gravier, 1910, *P. solida* (Forskål, 1775), *P. rus* (Forskål, 1775) and *Porites* sp (a total of 12 species and two unknown morphology); (C) many species as the number of molecular clades recovered in the concatenation-based phylogenies (a total of eight species); (D) lumped clade V (*P. annae*, *Porite* sp2, *P. harrisoni*, *P. lobata*, *P. lutea*, *P. solida*), clade VII (*Porites* sp), and clade VIII (*P. somaliensis*), split remaining molecular clades (a total of six species); (E) partitions inferred by Admixture with the optimal K (a total of five species). We performed the BFD* analysis using the SNAPP package (Bryant et al., 2012) implemented in BEAST v.2.5.2 (Bouckaert et al., 2014). We estimated MLE of each model by running path sampling with 48 independent steps (chain length of 100,000 MCMCs with a pre-burnin of 10,000 steps). Model convergence was assessed by monitoring the ESS for the likelihoods of each path using Tracer v.1.6 (Rambaut and Drummond, 2007). We ranked the alternative species delimitation models by their MLE and calculated the corresponding BF to compare the models. The strength of support from BF (2 * [MLE$_{best}$ − MLE$_{alternative}$]) comparisons of competing models was evaluated using the framework of Kass and Raftery (1995).

To investigate phylogenetic relationships among the Arabian *Porites* species, we used the coalescent-based species tree approach implemented in SNAPP (Bryant et al., 2012) with BEAST v.2.5.2 (Bouckaert et al., 2014). The method calculates species tree likelihoods directly from the data by estimating the probability of allele frequency change across nodes, thus bypassing the inference of individual gene trees. Two separated analyses were run using the two best models from BFD* as *a priori* taxa assignments (Liu et al., 2009). In order to reduce the complexity in species tree estimation and increase parameter convergence probability, we sampled one or two individuals for each clade/species since calculations do not benefit from adding extra individuals over large number of loci (Drummond and Bouckaert, 2015). We used VCFtools v.0.1.16 (Danecek et al., 2011) to generate a supermatrix of 1107 unlinked biallelic SNPs with 0% missing data for the model C (8 molecular clades) dataset and 965 SNPs for the model B (14 morpho-species) dataset. The MCMCs were run for 10 million generations with mutation rate and priors estimated during the chains and all the other settings were set as default. We monitored the traces for convergence using Tracer v.1.6 (Rambaut and Drummond, 2007). We concluded the analyses when ESSs for all parameters were large (>200) and the traces have reached stationarity, and discarded the first 10% of trees as burn-in. Densitree v.2.5.2 (Bouckaert et al., 2014) was used to visualize the posterior distributions of topologies as cloudograms, hence allowing for a clear depiction of uncertainty in the topology.

3. Results

3.1. Morphological identification of <u>Porites</u>

Based on the morphological examination, the 126 collected colonies were assigned to 12 nominal species currently considered valid and two undescribed morphology which are hereafter referred as *Porites* sp and P. sp2 (Table S1, Fig. S2).

3.2. mtCR, mtGenomes, histone, and rDNA phylogenetic analyses

The final mtCR alignment consisted of 1287 bp, with 40 variable sites, 10 of which were singleton sites and 30 parsimony informative. Reads mapping to the *P. lobata* mtGenome resulted in a mean of 2737 reads, covering 88% of the reference sequence, at a mean depth of 31 ± 44 standard deviation (s.d). Mapping paired end reads to the *P. superfusa* histone and rDNA resulted in a mean of 7216 and 14,816 reads, covering 92% (mean depth 328 ± 321 s.d) and 97% (mean depth 617 ± 600 s.d) of the reference sequences (Table S2). The mtGenome alignment (n = 124) consisted of 18,647 bp, with 57 variable sites, of which 16 were

singleton and 41 parsimony informative. The histone (n = 119) and rDNA (n = 124) alignments were 5464 bp and 6675 bp long, respectively. The histone alignment contained 11 variable sites, of which five singleton sites and six parsimony informative sites. A total of 76 variable sites, with 40 singleton and 36 parsimony informative sites were found in the rDNA. The ML topologies from the four datasets were mostly congruent, with the *Porites* samples clustered into 8 well supported clades (clades I to VIII – Fig. 1, Fig. S3). The notable exceptions were the mtCR tree where clade VIII was not resolved (Fig. 1, Fig. S3), and the histone tree where no sequences of representatives of clade III were obtained by reference mapping analysis (Fig. 1, Fig. S3). Five clades uniquely consisted of multiple specimens from a single morphologically defined species, and their monophyly was highly supported: clade I = *P. fontanesii*; clade II = *P. columnaris*; clade III = *P. farasani*; clade VI = *P. hadramauti*; clade VIII = *P. somaliensis*. Specimens of *P. rus* and *P. monticulosa* clustered together within clade IV. Specimens identified as *P. annae, Porites* sp2, *P. harrisoni, P. lobata, P. lutea,* and *P. solida* clustered within clade V, without any meaningful genetic structure. Finally, clade VII included all *Porites* sp samples. Clade I was identified as the basal clade in these phylogenies, with the exception of the histone tree. The mtCR, mtGenome, and rDNA topologies highlighted sister relationships between clade III and VI. Similarly, the mtGenome, histone, and rDNA highlighted sister relationships between clade IV and VII. The phylogenetic position of clade II, V, and VIII varied among the different reconstructions (Fig. 1, Fig. S3).

The final alignment of newly produced data with previous published ITS sequences from Forsman et al. (2009), and Terraneo et al. (2019a, b), consisted of 836 bp. The ML topology recovered all the above-mentioned clades (I to VIII), and the 12 clades from Forsman et al. (2009) (clades I to XII *sensu* Forsman et al., 2009). In particular, three clades were overlapping among these datasets: Clade IV sequences were included with Clade III *sensu* Forsman et al. (2009): Clade V was overlapping with Clade I *sensu* Forsman et al. (2009); Clade VIII was overlapping with Clade V *sensu* Forsman et al. (2009). Clades I, II, III, IV, VI, and VII only comprised sequences from the seas around the Arabian Peninsula. The clades II, IV, VI, VII, VIII, IX, X, XI, XII *sensu* Forsman et al. (2009) only nested sequences from the Pacific and Atlantic Oceans (Fig. S4).

3.3. Phylogenomic analyses

The dataset generated with the *de novo* assembly strategy was composed of 3102 loci. The BLAST searches revealed that 2654 loci had hits to coral genome while only 73 loci were mapped to Symbiodiniaceae genome and transcriptomes. The "denovo-max" and "refbased-max" topologies were almost identical, resulting in two well-supported trees (Fig. 2). Samples were clustered in eight clades in both topologies, in agreement with those in Fig. 1. *Porites fontanesii* (clade I)*, P. columnaris* (clade II), *P. farasani* (clade III), *P. hadramauti* (clade VI), *Porites* sp (clade VII), and *P. somaliensis* (clade VIII) were monophyletic, while *P. rus* and *P. monticulosa* consistently merged within clade IV. *Porites annae, Porites* sp2*, P. harrisoni, P. lobata, P. lutea,* and *P. solida* were indistinguishable within clade V (Fig. 2). The two trees showed congruent relationships among clades, with the exception of clade II, whose phylogenetic position differed in the two trees. The analyses consistently recovered sister relationships between clade III and VI, and between clade IV and VII. Similarly, the "denovo-min" and the "refbased-min" analyses defined the same clades, with the exception of the merging of clades IV and VIII in the "refbased-min" tree (Fig. S5).

3.4. Ordination and clustering analyses

The PCA results for the "refbased-max" dataset are shown in Fig. 3a-b. Specimens belonging to clades I, II, III, V, and VI were clearly separated in groups along the first two principal components (PC1 and PC2), while representatives of the remaining clades IV, VII, and VIII were

mixed. The PC3 and PC4 plot isolated clades II, III, IV, V, VI, VII, and VIII. Results from 27 different combinations of filtering options yielded similar results (Fig. S6), indicating that the outcomes are independent from the filtering process.

The Admixture analysis of 96,986 refbased SNPs supported K = 5 as the optimal model (Fig. 3c). The resulting plot showed three groups corresponding to clades I, II, and V, a fourth group including clades III, VI, and VII, a fifth group composed of clade IV, while clade VII was admixed between the latter two groups. Results from 27 combinations of filtering options yielded similar results (Figs. S7-S8), indicating that the outcomes are independent of the filtering process. Following Gowen et al. (2014) we then analyzed the unresolved clusters of samples (III, IV, VI, VII, and VIII) until not further clustering was recovered. The Admixture analyses supported K = 3 as the optimal model (Fig. 3d), with the plot showing three groups corresponding to clades IV and VIII, plus a third group comprising clades III, VI, VII. Subsequentially, we analyzed this latter unresolved group (III, VI, VII). The analyses supported K = 2 as the optimal model (Fig. 3e). The two groups were composed of clade III and clades VI and VII. Finally, we analyzed the cluster comprising clades VI and VII, and we recovered K = 2 as the optimal model (Fig. 3f), with the subsequent plot showing clade VI and clade VII.

3.5. Species delimitation analyses

The BFD* analysis suggested model C was the best supported model. Model C assigned the samples according to the eight molecular clades recovered by the phylogenomic reconstructions (MLE = −6424; BF = 0). The second-best model agreed with the current taxonomy recovering 14 species (model B, MLE = −6444; BF = 40). Model D (six species) ranked third (MLE = −7061; BF = 1274). Model E assigning the samples according to the optimal K from the Admixture analysis was the fourth best supported model (MLE = −7474; BF = 2100). Finally, model A lumping all samples as one single species, exhibited the lowest MLE value (MLE = −9485; BF = 6122) (Table 1).

The coalescent-based species tree approach based on the BFD* highest ranking (eight lineages) yielded a fully resolved phylogeny with the exception of clade II (*P. columnaris*), whose phylogenetic relationships remained unclear (Fig. 4). The species tree topology showed some differences with the trees presented in Fig. 2. In particular, it recovered the basal position of clade I (*P. fontanesii*) and the sister relationship between clades V and VIII (*P. somaliensis*). The analyses based on 14 species (Fig. S9) displayed an identical topology to the species tree based on eight lineages. Nevertheless, the six nominal species in clade V showed a very complex pattern, with mostly unresolved relationships.

4. Discussion

This work used RADseq to clarify evolutionary relationships among 14 morphologically defined species of *Porites* from the Arabian Peninsula. Although 28 species of *Porites* are reported to occur in the Arabian Peninsula (Veron, 2000; Berumen et al., 2019), we encountered half. This could either be the result of an insufficient sampling effort, for example we did not survey the western coasts of the Red Sea and the southern coasts of the Gulf of Aden, or of poor taxonomy in the existing reports on the genus diversity in the region (Sheppard and Sheppard, 1991; Veron, 2000). Indeed, the majority of the reported species have their type localities outside the Arabian Peninsula or the Indian Ocean, and their occurrence in the region has not been confirmed (Terraneo pers. comm).

We reconstructed the phylogeny of 126 corals from one mitochondrial marker (mtCR), nearly complete mtGenome, rDNA, and histone regions, and we compared SNPs generated with both reference-based and *de novo* assembly strategies. The trees inferred from the latter two datasets were highly congruent among the clades, with the exception of clade II (Fig. 2). The similarity between the two topologies was likely driven by the fact that the *de novo* assembly dataset was mainly
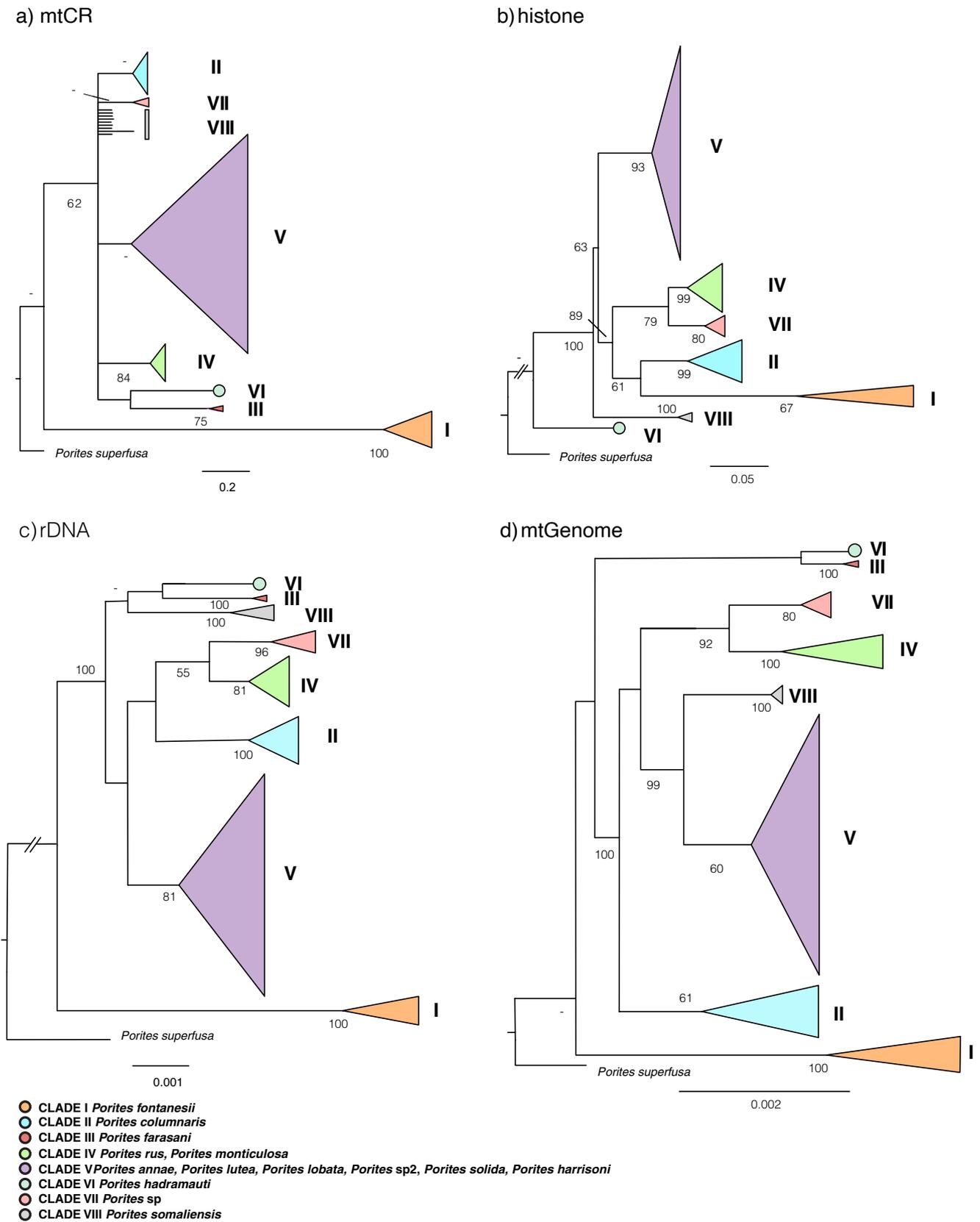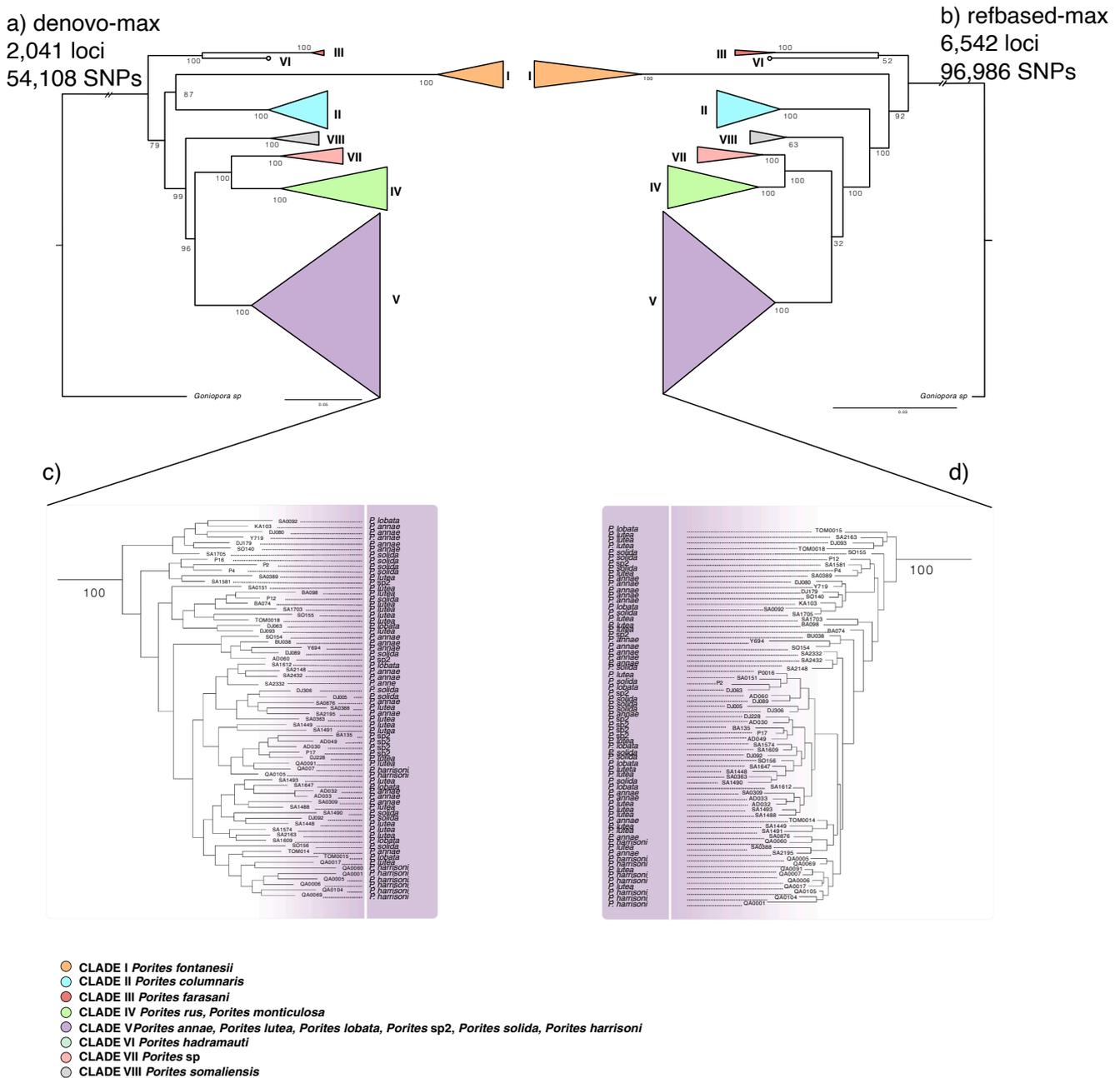
Fig. 1. . Comparison of phylogenetic trees of *Porites* corals based on: a) mtCR; b) histone; c) rDNA; d) mtGenome. Node values represent ML bootstrap supports. The dash (-) symbol refers to bootstrap values below 55. Roman numbers from I to VIII refer to the assigned clade numbers. The color codes and nominal species corresponding to the clades are explained in the legend.

Fig. 2. . Comparison of RAxML phylogenetic tree based on: a) "denovo-max" dataset (2041 loci and 54,108 SNPs); b) "refbased-max" dataset (6452 loci and 96,986 SNPs); c) zoom-in Clade V in the tree based on the "denovo-max" dataset; d) zoom-in Clade V in the tree based on the "refbased-max" dataset. Node values represent ML bootstrap supports. Roman numbers from I to VIII refer to the assigned molecular clade numbers. The color codes and nominal species corresponding to the clades are explained in the legend.

composed of coral loci. Indeed, the *de novo* assembly dataset was composed of 3102 loci and we showed that a major fraction of these loci (N = 2654) had BLAST hits to coral genome while only a minor fraction was mapped to Symbiodiniaceae genome and transcriptomes (N = 73). Together with ordination, clustering, and species delimitation analyses, with the only exception of the PCA analysis, our data identified eight separately evolving lineages of *Porites* in this region. Six of these lineages corresponded with five nominal species, *i.e. P. fontanesii* (clade I), *P. columnaris* (clade II), *P. farasani* (clade III), *P. hadramauti* (clade VI), and *P. somaliensis* (clade VIII), and one undescribed species (clade VII), thus suggesting the need for the formal description of a new species. In contrast, the two nominal species *P. rus* and *P. monticulosa* in clade IV, and the five nominal species *P. annae*, *P. harrisoni*, *P. lobata*, *P. lutea*, and *P. solida* in clade V, could not be distinguished based on our reduced-

genome approach and clustered into two complexes. The phylogenies in Figs. 1 and 2 resolved identical clades with slightly different topologies. Such discordance is common in phylogenetic analyses because different genes or genomic regions have different evolutionary histories (Bryant et al., 2012). Interestingly the mtCR reconstruction provided the same clade resolution (with the exception of clade VIII), yet slightly different topology, as the one obtained by analyzing almost complete rDNA, histone, and mtGenome and including thousands of SNPs. This study corroborates the finding from Terraneo et al. (2019 a, b) where reconstructions based on mtCR were provided, and highlights the utility of the mtCR region in delimiting species boundaries in *Porites*. Indeed, the use of a mitochondrial marker could avoid expensive NGS libraries preparation, and time-consuming data analyses. Nevertheless, the inclusion of more species from other localities might prove that a
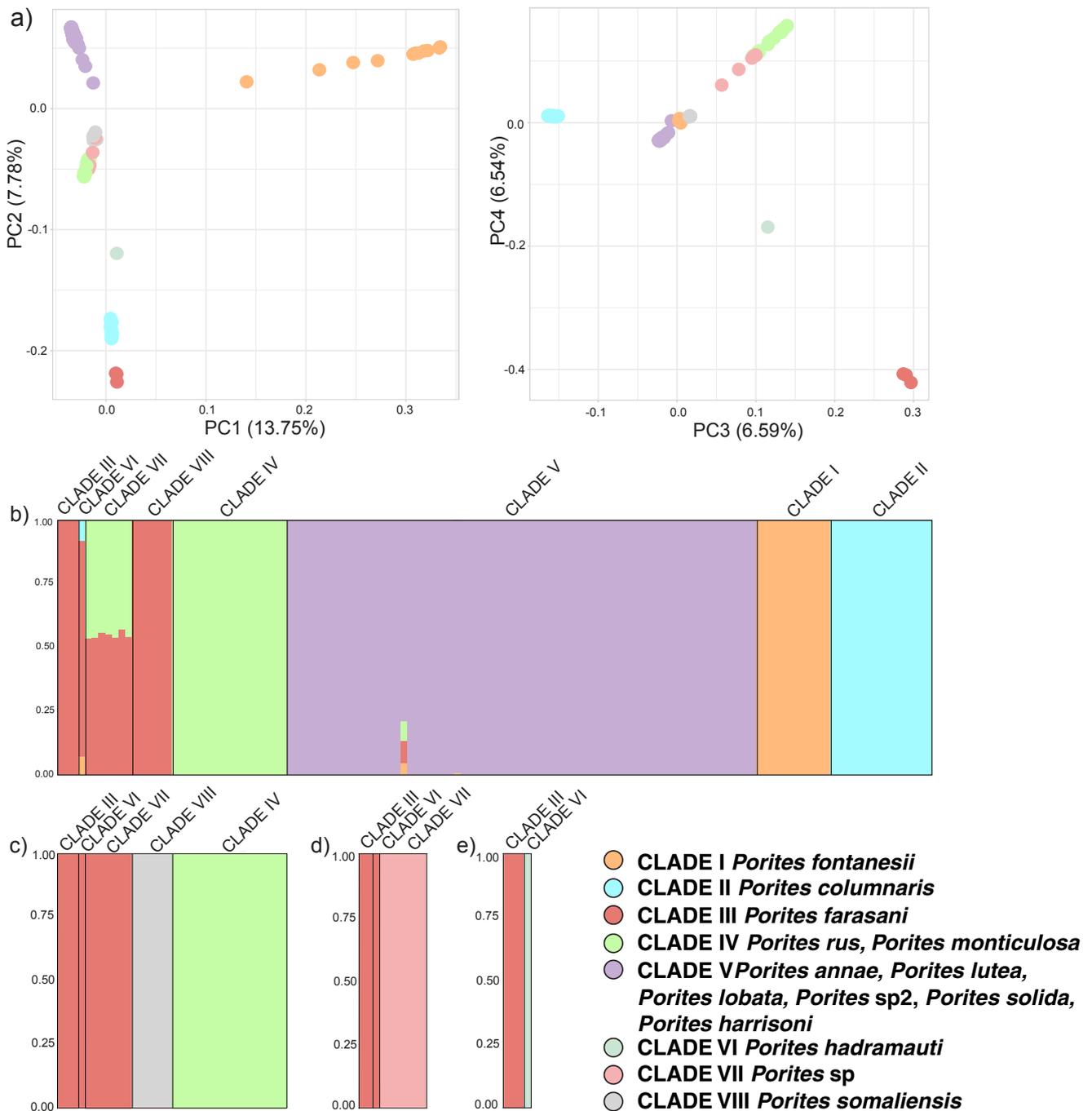
Fig. 3. . Principal Component Analysis (PCA) and Admixture results from the "refbased-max" dataset (96,986 SNPs): a) PCA results based on PC1 (x axis) and PC2 (y axis); b) PC3 (x axis) and PC4 (y axis); c) Admixture plot for K = 5 (best suggested model); d) Admixture plot for K = 4 (second most probable model); e) Admixture plot for K = 8 (third most probable model). Roman numbers from I to VIII refer to the assigned molecular clade numbers. Color codes are explained in the legend.

phylogenetic approach based on this region could not *be sufficient.* Finally, high concordance between the species tree (Fig. 4) and the mtGenome reconstruction (Fig. 1d) was recovered. In the case of our dataset, such result indicates an alternative methodology towards accurate phylogenetic reconstructions for the genus *Porites*. The use of complete mtGenome data in fact would bypass the need of time-consuming and expensive genomic data. Yet, the inclusion of further samples and species in the analyses might require a genomic approach as coral mitochondrial DNA is notorious for slow evolution rates that might mislead accurate evolutionary reconstructions.

Our data corroborated previously published results from Forsman et al. (2009), and recovered several new lineages of *Porites* from the seas

around the Arabian Peninsula. Clade IV sequences (P. *rus, P. monticulosa*) clustered together with Clade III sequences of *P. rus* and *P. monticulosa sensu* Forsman et al. (2009). Clade V sequences nested together with Clade I *sensu* Forsman et al. (2009), an unresolved clade comprised of *P. lutea, P. lobata, P. solida, P. annae, P. compressa, P. cylindrica*. With the exception of *P. compressa* that is an Hawaiian endemic species, and *P. cylindrica* that we did not encounter in the seas around the Arabian Peninsula, our study Clade V included the same nominal species (*P. lutea, P. lobata, P. annae, P. solida*), with the addition of *P.* sp2 (an unidentified morphology) and *P. harrisoni*, which only occurs in the Arabian Gulf. Clade VIII sequences nested together Clade V *sensu* Forsman et al. (2009). In the current work, Clade VIII included

Table 1

Bayes Factor delimitation (BFD*) results for each analysis using path sampling (PS) with SNAPP. The number of lineages represents the number of putative species included in each analysis. BF values are used to rank species models, relative to the species model with the lowest marginal likelihood. The model C with eight lineages corresponding to the eight molecular clades recovered in Fig. 2 was supported as the best fit model.

Model name	Model specifications	Number of lineages	MLE	BF	Rank
A	One single species	1	−9,485	6,121	5
B	Current taxonomy	14	−6,444	41	2
C	Molecular clades	8	−6,426	–	1
D	Lumped clades V, VII, VIII and all other clades	6	−7,061	637	3
E	Optimal K according to Admixture	5	−7,474	2,099	4

Note: MLE = Marginal likelihood (log$_e$); BF = Bayes factor (2 * [MLE$_{best}$ − MLE$_{alternative}$]).

only samples of *P. somaliensis*, while Forsman et al. (2009) recovered *P. lutea* and *P. lobata* within this clade. It remains to understand if the clustering of different nominal species within this clade is related with wrong species identification or if Clade VIII is also a complex of different nominal species. The possibility of misidentification might be the most likely explanation since *P. somaliensis*, *P. lutea* and *P. lobata* share a massive colony morphology and small corallites which might have hidden morphological variation among these nominal species.

4.1. Diversity of Porites from the Arabian Peninsula

We collected 12 nominal species of *Porites* out of the 28 reported from the region (Veron, 2000, Claereboudt, 2006, Benzoni and Stefani, 2012, Veron et al., 2015; Terraneo et al., 2019a) and discovered a genetically distinct undescribed species. Our sampling strategy aimed to sample all distinct morphological entities we encountered in every collection site. The success of this strategy was best demonstrated by the fact that we recently described three new species from this region, *i.e. P.*

fontanesii, *P. farasani* and *P. hadramauti* (Benzoni and Stefani, 2012, Terraneo et al., 2019a), and, moreover, in this study we collected two morphotypes that did not correspond to any described material, *i.e. Porites* sp and *Porites* sp2. This could be either the result of a) an insufficient sampling effort or b) of poor taxonomy in the existing reports on the genus diversity in the region (Sheppard and Sheppard, 1991; Veron, 2000). In fact, with regards to insufficient sampling, while the Red Sea, Gulf of Tadjoura, and Gulf of Aden were extensively surveyed and sampled for this study, an uneven sampling effort was produced for the Arabian Sea and the Gulf. Moreover, we did not collect specimens from the western coasts of the Red Sea and the southern coasts of the Gulf of Aden. Nevertheless, coalescent methods and species delimitation data suggested that *Porites* diversity from the seas around the Arabian Peninsula needs to be reconsidered. The 14 morphospecies we collected were placed into eight genetically defined lineages, one of which is identified for the first time.

The inclusion of about 96,000 SNPs generated with a reference-based assembly strategy and 54,000 SNPs obtained from a *de novo* assembly analysis questioned the validity of the many morphologically identified species included in clades IV and V, suggesting that these two species complexes represent either extreme phenotypic polymorphism, rapid incipient speciation, hybridization, or a mix of these possibilities. Further studies encompassing ecological, symbiont association, and reproductive data will be necessary to determine the potential presence of functional differences and reproductive isolation mechanisms among the morphology-based species nested within these two species complexes.

Porites rus and *P. monticulosa* consistently clustered into one lineage (clade IV, Figs. 1 and 2, Figs. S3-S4), a result corroborated by clustering and species delimitation analyses (Fig. 4, Table 1, Figs. S5-S9). Terraneo et al. (2019b) also failed to separate these two species in the Red Sea. Morphologically, this is hardly surprising. In fact, the two species are currently mainly told apart based on the colony growth form (Fig. S2d-e) (Veron, 2000 (3): 314). However, colony growth form is hardly an informative trait in scleractinian corals (Romano and Palumbi, 1996; Fukami et al., 2004; Huang et al., 2011; Budd et al., 2012), being subject to variability and environmental induced plasticity (Todd, 2008; Paz-
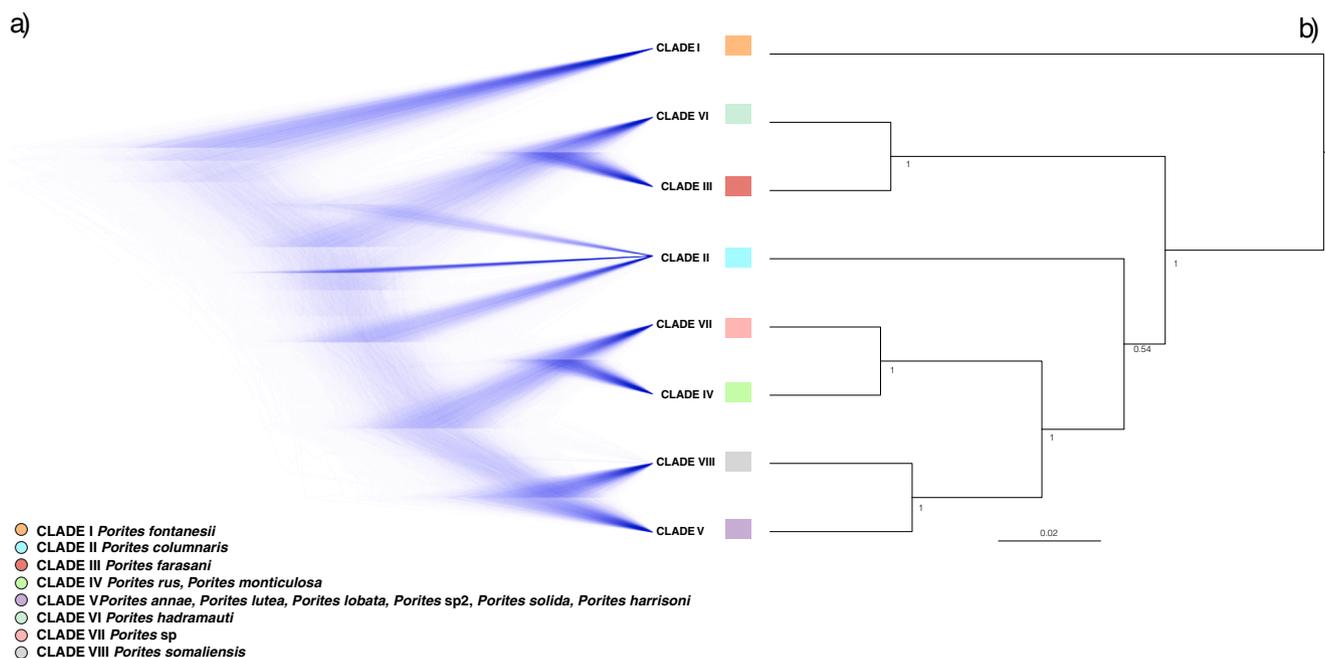

Fig. 4. . Species tree of *Porites* corals. a) The cloudogram represents the posterior distribution of species trees inferred with SNAPP, based on a priori imposition of eight taxa, 1107 unlinked biallelic SNPs and no missing data. High color densities are representative of high topology agreement in the species tree. Roman numbers from I to VIII refer to the assigned molecular clade numbers. b) Maximum credibility tree inferred with DensiTree. Node values represent posterior probabilities. The color codes and nominal species corresponding to the clades are explained in the legend.

García et al., 2015). For instance, light induced morphological plasticity has been documented in field transplantation experiments of *P. sillimaniana* in Japan, where different light intensity along the reef slope drove branching or mounding colony growth forms (Muko et al., 2000). Our reduced-genome approach on Arabian Region *Porites* suggested that material identified as *P. rus*, a species originally described from the Red Sea, and *P. monticulosa*, later named based on material from Fiji, represent a single molecular entity. Hence, the hypothesis that *P. monticulosa* is a junior synonym of *P. rus*, needs to be further investigated by including material from *P. monticulosa* type locality, as well as detailed morphological and morphometric examinations.

Clade V represents a different case. The genomic data indicated that clade V is a complex of six morphologically clearly defined species: *P. annae*, *Porites* sp2, *P. harrisoni*, *P. lobata*, *P. lutea*, and *P. solida*. Morphological variability can often result in apparent genetic polyphyly through erroneous assignment of alternative morphologies to different species (Arrigoni et al., 2016a,b; Benzoni et al., 2010; Terraneo et al., 2016; Cunha et al., 2019). From a morphological point of view, colonies of *P. lobata, P. lutea,* and *P. solida* share a similar massive growth form, while *P. annae* and *P. harrisoni* have a columnar morphology, and *Porites* sp2 forms mainly encrusting colonies. However, the six species present consistent and well-defined corallite level differences (summarized and illustrated in Terraneo et al., 2019a), based on which they were described as different species. A possible scenario is that clade V consists of a single remarkably morphologically variable species, characterized by different and often sympatric corallite phenotypes. An alternative hypothesis is that incomplete lineage sorting and weak genetic drift led to a misleading phylogeny reconstruction (de Queiroz, 1998, 2007). Under this scenario, the polyphyly of species found in clade V may be explained by rapid diversification or recent speciation of the clustered lineages (Funk and Omland, 2003). Furthermore, phylogenetic signals may be hidden by gene transfer among divergent lineages undergoing hybridization and introgression (van Oppen et al., 2000, 2002; Frade et al., 2010: Combosch and Vollmer, 2015; Forsman et al., 2017). *In vitro* trials suggested that hybridization in hard corals may be common (Willis et al., 2006), and has been reported in many genera such as *Acropora* Oken, 1815, *Platygyra* Ehrenberg, 1834, *Pocillopora*, and *Stylophora* Schweigger, 1820 (Richards et al., 2008; Richards and Hobbs, 2015). In the Caribbean, hybridization has been reported between the species *A. cervicornis* (Lamarck, 1816) and *A. palmata* (Lamarck, 1816), and backcrossing of the hybrid *A. prolifera* (Lamarck, 1816) with the parental species seems to occur at low frequencies too (Vollmer and Palumbi, 2002). Combosch et al. (2008) first reported hybridization among *Pocillopora damicornis* (Linnaeus, 1758), *P. eydouxi* Milne Edwards, 1860, and *P. elegans* Dana, 1846 in the Eastern Pacific, and RADseq recently confirmed one-way introgression among these species (Combosch and Vollmer, 2015). Our genome-wide data show that the six morphologies in clade V belong to a single lineage. Nevertheless, further analyses are necessary to exclude signatures consistent with introgressive hybridization or incomplete lineage sorting. Species complexes in scleractinian corals are common, yet our understanding of these is still vague (Frade et al., 2010; Arrigoni et al., 2016b; Cunha et al., 2019). Further analyses including traits from additional sources, such as coral reproduction biology and algal symbiont association, might allow to better evaluate these hypotheses and the biological nature of this lineage.

4.2. Geographic distribution of Porites in the Arabian Peninsula

Understanding how species distributions are historically shaped remains a central topic in evolutionary biology (Wiens and Donoghue, 2004; Bowen et al., 2013). Our results showed that *Porites* molecular lineages display a peculiar geographical distribution in the seas around Arabia, with some lineages widespread around the peninsula, and others apparently restricted to a given region (Fig. 5).

Five of the species examined in this study are Arabian endemics, each

with a distinct distribution. *Porites fontanesii* and *P. columnaris* are widely distributed in the Red Sea and the Gulf of Aden, but not recorded in the Arabian Gulf. *Porites hadramauti* is restricted to the Gulf of Aden, *Porites* sp is found in the Gulf of Tadjoura, Gulf of Aden, Gulf of Oman, and *P. farasani* is a southern Red Sea endemic hard coral (Terraneo et al., 2019a). High rates of endemism in several marine groups are typical of the seas around the Arabian Peninsula. This region has been recognized as an endemism hotspot in the Indian Ocean (Obura, 2012, 2016; DiBattista et al., 2016a), and recent estimates suggested that 11% of scleractinian corals in the basins around the Arabian Peninsula are endemic (Berumen et al., 2019). The evolutionary processes that led to the origin of endemism hotspots in peripheral areas remain elusive. However, the diversity of the habitats and environments, and the complex geological and paleoclimatic history of the seas around the Arabian Peninsula might have played a key role in shaping the current biodiversity patterns (Sheppard et al., 1992; Bosworth et al., 2005; DiBattista et al., 2016a; Siddall et al., 2003). The Bab Al Mandeb Strait is the only present connection between the Red Sea and the Gulf of Aden. Limited water exchange seasonally driven by the Indian Ocean monsoon system occurs through this shallow and narrow channel, creating a potential barrier to genetic exchange between the Red Sea and the rest of the Indian Ocean (DiBattista et al., 2016a, 2016b). Moreover, a monsoon-driven upwelling system causes major fluctuations in the summer water temperature and nutrients in the Gulf of Aden, limiting reef development in this region, as opposed to the oligotrophic biodiverse waters of the Red Sea, and limiting the persistence of only some well adapted species in this region (Vénec-Peyré and Caulet, 2000; Benzoni et al., 2003).

Only one out of the examined eight clades of *Porites* is widespread throughout the Arabian Peninsula, the enigmatic and morphologically diverse clade V. This might indicate the presence of species of recent divergence that still need to acquire fixed genomic signatures.

Only 16% of the scleractinian corals that inhabit the Arabian basins are found in the Arabian Gulf (Veron, 2000; Coles, 2003; Veron et al., 2015; Berumen et al., 2019). This region is one of the most extreme habitats for hermatypic corals with high nutrient input from the Gulf of Oman while water temperatures can vary up to 20 °C during the year (Coles, 2003). Moreover, it is a young basin, originated 14 Ka, and mainly constituted by shallow-water environments (Ross et al., 1986). The combination of these factors limits the chance for hard corals to settle and survive and, therefore, it is not surprising that only one lineage (clade V) out of the eight molecular groups is found in the Arabian Gulf (Fig. 5).

5. Conclusions

Important gaps remain in the understanding of biodiversity, biogeography, and evolution of the hard coral genus *Porites*, and the present work demonstrated that there is an urgent need for a complete taxonomic revision. This work harnesses the power of NGS coupled with phylogenomics, ordination, clustering, and species delimitation methods, to clarify the diversity and evolutionary relationships of *Porites* in the seas around the Arabian Peninsula. Our results from different genomic resources fully demonstrated the presence of eight molecular lineages that are in agreement with morphology-based taxonomy with the exception of two species complexes, providing unprecedented resolution at the species level in *Porites*. The inclusion of about 96,000 and 54,000 SNPs generated from reference-based and *de novo* assembly strategies, respectively, questioned the validity of the many morphologically identified species included in clades IV and V, suggesting that these two species complexes represent either extreme phenotypic polymorphism, rapid incipient speciation, hybridization, or a mix of these possibilities. Further studies encompassing ecological, symbiont association, and reproductive data will be necessary to determine the potential presence of functional differences and reproductive isolation mechanisms among the morphology-based species nested within these
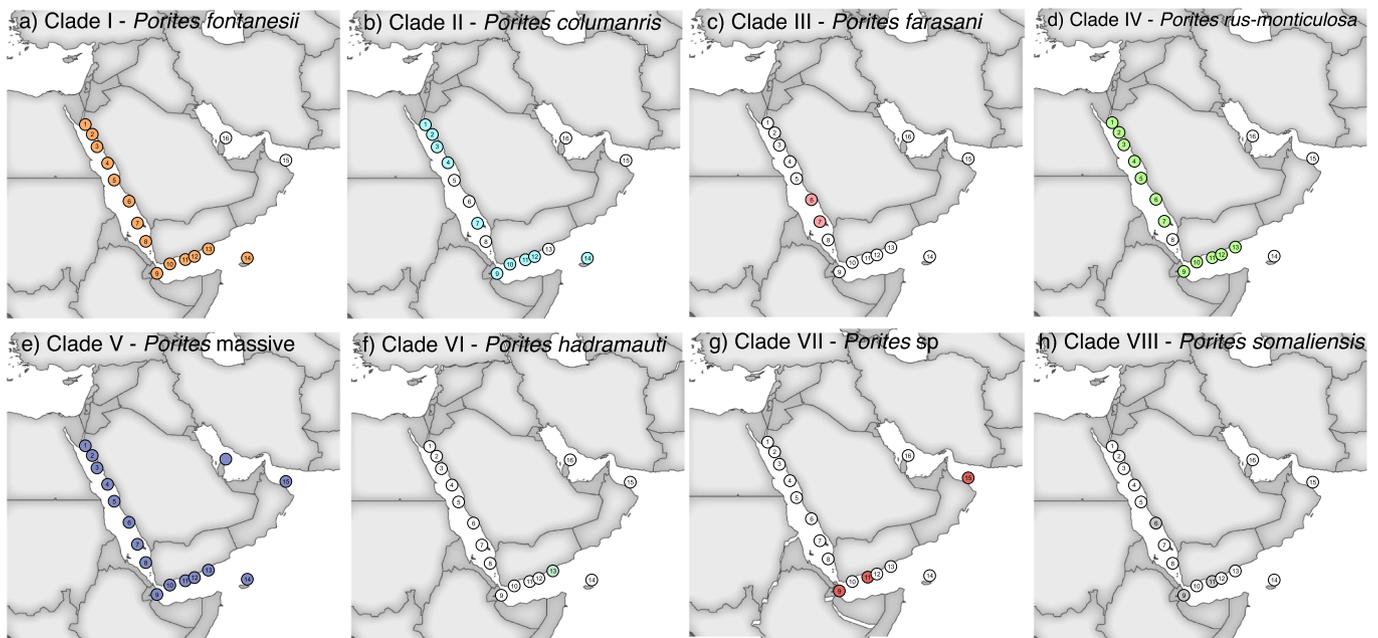
Fig. 5. Distribution maps and sampling localities at 16 sites around the Arabian Peninsula of *Porites* molecular lineages recovered in this study. a) *P. fontanesii* – clade I, b) *P. columnaris* – clade II, c) *Porites farasani* – clade III, d) *P. rus, P. monticulosa* – clade IV, e) *P. annae, Porites* sp2*, P. harrisoni, P. lobata, P. lutea,* and *P. solida* – clade V, f) *P. hadramauti* – clade VI, g) *Porites* sp – clade VII, h) *P. somaliensis* – clade VIII. Numbers from 1 to 16 refer to the sampling localities: 1 to 8 Red Sea; 9 to 14 Gulf of Aden and Socotra Island; 15 Gulf of Oman; 16 Arabian Gulf.

two species complexes.

CRediT authorship contribution statement

Tullia I. Terraneo: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Francesca Benzoni:** Supervision, Conceptualization, Investigation, Writing - review & editing. **Roberto Arrigoni:** Methodology, Formal analysis, Writing - review & editing. **Andrew H. Baird:** Supervision, Conceptualization, Writing - review & editing. **Kiruthiga G. Mariappan:** Software. **Zac H. Forsman:** Software, Writing - review & editing. **Michael K. Wooster:** Investigation. **Jessica Bouwmeester:** Investigation, Writing - review & editing. **Alyssa Marshell:** Investigation, Writing - review & editing. **Michael L. Berumen:** Supervision, Conceptualization, Writing - review & editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107173.

References

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Ament-Velásquez, S.L., Breedy, O., Cortés, J., Guzman, H.M., Wörheide, G., Vargas, S., 2016. Homoplasious colony morphology and mito-nuclear phylogenetic discordance among Eastern Pacific octocorals. Mol. Phylogenetics Evol. 98, 373–381.

Arrigoni, R., Benzoni, F., Terraneo, T.I., Caragnano, A., Berumen, M.L., 2016a. Recent origin and semi-permeable species boundaries in the scleractinian coral genus *Stylophora* from the Red Sea. Sci. Rep. 6, 34612.

Arrigoni, R., Berumen, M.L., Chen, C.A., Terraneo, T.I., Baird, A.H., Payri, C., Benzoni, F., 2016b. Species delimitation in the reef coral genera *Echinophyllia* and *Oxypora* (Scleractinia, Lobophylliidae) with a description of two new species. Mol. Phylogenetics Evol. 105, 146–159.

Arrigoni, R., Berumen, M.L., Mariappan, K.G., Beck, P.S., Hulver, A.M., Montano, S., Pichon, M., Strona, G., Terraneo, T.I., 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.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One. 3, e3376.

Baxter, S.W., Davey, J.W., Johnston, J.S., Shelton, A.M., Heckel, D.G., Jiggins, C.D., Blaxter, M.L., 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PloS One 6, e19315.

Bellwood, D.R., Hughes, T.P., Folke, C., Nyström, M., 2004. Confronting the coral reef crisis. Nature 429, 1532–1534.

Benzoni, F., Bianchi, C.N., Morri, C., 2003. Coral communities of the northwestern Gulf of Aden (Yemen): variation in framework building related to environmental factors and biotic conditions. Coral Reefs 22, 475–484.

Benzoni, F., Stefani, F., Pichon, M., Galli, P., 2010. The name game: morpho-molecular species boundaries in the genus *Psammocora* (Cnidaria, Scleractinia). Zool. J. Linnean Soc. 160, 421–456.

Benzoni, F., Stefani, F., 2012. *Porites fontanesii*, a new species of hard coral (Scleractinia, Poritidae) from the southern Red Sea, the Gulf of Tadjoura, and the Gulf of Aden. Zootaxa 3447, 56–68.

Berumen, M.L., Arrigoni, R., Bouwmeester, J., Terraneo, T.I., Benzoni, F., 2019. Biodiversity of Red Sea corals. In: Voolstra, C.R., Berumen, M.L. (Eds.), Coral Reefs of the Red Sea. Springer, Dordrecht (the Netherlands), pp. 123–155.

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., Falkowski, P.G., 2016. Comparative genomics explains the evolutionary success of reef-forming corals. Elife 5, e13288.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bosworth, W., Huchon, P., McClay, K., 2005. The Red Sea and Gulf of Aden basins. J. Afr. Earth Sci. 43, 334–378.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.

Bowen, B.W., Rocha, L.A., Toonen, R.J., Karl, S.A., 2013. The origins of tropical marine biodiversity. Trends Ecol. Evol. 28, 359–366.

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

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932.

Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3, 846–852.

Claereboudt, M.R., 2006. *Porites decasepta*: a new species of scleractinian coral (Scleractinia, Poritidae) from Oman. Zootaxa. 11, 55–62.

Coles, S.L., 2003. Coral species diversity and environmental factors in the Arabian Gulf and the Gulf of Oman: a comparison to the Indo-Pacific region. Atoll Res. Bull. 507, 1–21.

Combosch, D.J., Guzman, H.M., Schuhmacher, H., Vollmer, S.V., 2008. Interspecific hybridization and restricted trans-Pacific gene flow in the Tropical Eastern Pacific *Pocillopora*. Mol. Ecol. 17, 1304–1312.

Combosch, D.J., Vollmer, S.V., 2015. Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific *Pocillopora* corals. Mol. Phylogenetics Evol. 88, 154–162.

Concepion, G.T., Crepeau, M.W., Wagner, D., Kahng, S.E., Toonen, R.J., 2008. An alternative to ITS, a hypervariable, single-copy nuclear intron in corals, and its use in detecting cryptic species within the octocoral genus *Carijoa*. Coral Reefs 27, 323–336.

Cunha, R.L., Forsman, Z.H., Belderok, R., Knapp, I.S., Castilho, R., Toonen, R.J., 2019. Rare coral under the genomic microscope: timing and relationships among Hawaiian *Montipora*. BMC Evol. Biol. 19, 153.

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., McVean, G., McVean, G., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

DiBattista, J.D., Roberts, M.B., Bouwmeester, J., Bowen, B.W., Coker, D.J., Lozano-Cortés, D.F., Choat, J.H., Gaither, M.R., Hobbs, J.P.A., Khalil, M.T., Kochzius, M., Myers, R.F., Paulay, G., Robitzch, V.S.N., Saenz-Agudelo, P., Salas, E., Sinclair-Taylor, T.H., Toonen, R.J., Westneat, M.W., Williams, S.T., Berumen, M.L., 2016a. A review of contemporary patterns of endemism for shallow water reef fauna in the Red Sea. J. Biogeogr. 43, 423–439.

DiBattista, J.D., Choat, H.J., Gaither, M.R., Hobbs, J.P.A., Lozano-Cortés, D.F., Myers, R. F., Paulay, G., Rocha, L.A., Toonen, R.J., Westneat, M.W., Berumen, M.L., 2016b. On the origin of endemic species in the Red Sea. J. Biogeogr. 43, 13–30.

Dimond, J., Gamblewood, S., Roberts, S., 2017. Genetic and epigenetic insight into morphospecies in a reef coral. Mol. Ecol. 26, 5031–5042.

Drummond, A.J., Bouckaert, R.R., 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Press, Cambridge (UK).

Eaton, D.A., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 62, 689–706.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.

Forsman, Z.H., Birkeland, C., 2009. *Porites randalli*: a new coral species (Scleractinia, Poritidae) from American Samoa. Zootaxa 2244, 51–59.

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., 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. PeerJ. 3, e751.

Forsman, Z.H., Knapp, I., Tisthammer, K., 2017. Coral hybridization or phenotypic variation? Genomic data reveal gene flow between *Porites lobata* and *P. compressa*. Mol. Phylogenetics Evol. 111, 132–148.

Forsman, Z.H., Ritson-Williams, R., Tisthammer, K.H., Knapp, I.S.S., Toonen, R.J., 2020. Host-symbiont coevolution, cryptic structure, and bleaching susceptibility, in a coral species complex (Scleractinia; Poritidae). Sci. Rep. 10, 16995.

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

Fukami, H., Budd, A.F., Paulay, G., Solé-Cava, A., Chen, C.A., Iwao, K., Knowlton, N., 2004. Conventional taxonomy obscures deep divergence between Pacific and Atlantic corals. Nature 427, 832.

Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34, 397–423.

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv Prepr arXiv12073907:1–9.

Gottscho, A.D., Wood, D.A., Vandergast, A.G., Lemos-Espinal, J., Gatesy, J., Reeder, T. W., 2017. Lineage diversification of fringe-toed lizards (Phrynosomatidae: *Uma notata* complex) in the Colorado desert: delimiting species in the presence of gene flow. Mol. Phylogenetics Evol. 106, 103–117.

Gowen, F.C., Maley, J.M., Cicero, C., Peterson, A.T., Faircloth, B.C., Warr, T.C., McCormack, J.E., 2014. Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. BMC Evol. Biol. 14, 135.

Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., Brumfield, R.T., 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Syst. Biol. 65, 910–924.

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., 2016. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Mol. Phylogenetics Evol. 100, 70–79.

Hipp, A.L., Eaton, D.A., 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.

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.

Huang, D., Licuanan, W.Y., Baird, A.H., Fukami, H., 2011. Cleaning up the 'Bigmessidae': Molecular phylogeny of scleractinian corals from Faviidae, Merulinidae, Pectiniidae and Trachyphylliidae. BMC Evol. Biol. 11, 37.

Johnston, E.C., Forsman, Z.H., Flot, J.F., Schmidt-Roach, S., Pinzón, J.H., Knapp, I.S., Toonen, R.J., 2017. A genomic glance through the fog of plasticity and diversification in *Pocillopora*. Sci. Rep. 7, 5991.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

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

Knapp, I., Puritz, J., Bird, C., Whitney, J., Sudek, M., Forsman, Z., Toonen, R., 2016. ezRAD—an accessible next-generation RAD sequencing protocol suitable for non-model organisms_v3. 1. Protocols.io Life Sciences Protocol Repository.

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357.

Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63, 534–535.

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

Hoeksema, B.W., Cairns, S., 2019. World List of Scleractinia. Porites Link, 1807. Accessed through: World Register of Marine Species at: http://www.marinespecies.org/aphia. php?p=taxdetails&id=206485 on 2019-10-01.

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

Lischer, H.E., Excoffier, L., 2011. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298–299.

Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58, 468–477.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environ. Workshop 14, 1–8.

Mcfadden, C.S., Hutchinson, M.B., 2004. Molecular evidence for the hybrid origin of species in the soft coral genus *Alcyonium* (Cnidaria: Anthozoa: Octocorallia). Mol. Ecol. 13, 1495–1505.

McFadden, C.S., Sánchez, J.A., France, S.C., 2010. Molecular phylogenetic insights into the evolution of Octocorallia: a review. Integr. Comp. Biol. 50, 89–410.

McFadden, C.S., Haverkort-Yeh, R., Reynolds, A.M., Halàsz, A., Quattrini, A.M., Forsman, Z.H., Benayahu, Y., Toonen, R.J., 2017. Species boundaries in the absence of morphological, ecological or geographical differentiation in the Red Sea octocoral genus *Ovabunda* (Alcyonacea: Xeniidae). Mol. Phylogenetics Evol. 112, 174–184.

Muko, S., Kawasaki, K., Sakai, K., Takasu, F., Shigesada, N., 2000. Morphological plasticity in the coral *Porites silimaniani* and its adaptive significance. Bull. Mar. Sci. 66, 225–239.

Obura, D., 2012. The diversity and biogeography of Western Indian Ocean reef-building corals. PloS One. 7, e45013.

Obura, D., 2016. An Indian Ocean centre of origin revisited: Palaeogene and Neogene influences defining a biogeographic realm. J. Biogeogr. 43, 229–242.

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

van Oppen, M.V., Willis, B.L., Vugt, H.V., Miller, D.J., 2000. Examination of species boundaries in the *Acropora cervicornis* group (Scleractinia, Cnidaria) using nuclear DNA sequence analyses. Mol. Ecol. 9, 1363–1373.

van Oppen, M.J., Willis, B.L., Van Rheede, T., Miller, D.J., 2002. Spawning times, reproductive compatibilities and genetic structuring in the *Acropora aspera* group: evidence for natural hybridization and semi-permeable species boundaries in corals. Mol. Ecol. 11, 1363–1376.

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.C., Samadi, S., 2015. Use of RAD sequencing for delimiting species. Heredity 114, 450–459.

Paz-García, D.A., Hellberg, M.E., García-de-León, F.J., Balart, E.F., 2015. Switch between morphospecies of *Pocillopora* corals. Am. Nat. 186, 434–440.

Prada, C., Schizas, N.V., Yoshioka, P.M., 2008. Phenotypic plasticity or speciation? A case from a clonal marine organism. BMC Evol. Biol. 8, 47.

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.

Pratlong, M., Rancurel, C., Pontarotti, P., Aurelle, D., 2017. Monophyly of Anthozoa (Cnidaria): why do nuclear and mitochondrial phylogenies disagree? Zool. Scr. 46, 363–371.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Priyam, A., Woodcroft, B.J., Rai, V., Moghul, I., Munagala, A., Ter, F., Moon, H., 2019. Sequenceserver: a modern graphical user interface for custom BLAST databases. Mol. Biol. Evol. 36, 2922–2924.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

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

Quattrini, A.M., Faircloth, B.C., Duenas, L.F., Bridge, T.C., Brugler, M.R., Calixto-Botía, I. F., DeLeo, D.M., Forët, S., Herrera, S., Lee, S.M.Y., Miller, D.J., Prada, C., Radis-Baptista, G., Ramíz-Portilla, C., Sánchez, J.A., Rodríguez, E., McFadden, C.S., 2018. Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: new approaches to long-standing problems. Mol. Ecol. Resour. 18, 281–295.

Quattrini, A.M., Wu, T., Soong, K., Jeng, M.S., Benayahu, Y., McFadden, C.S., 2019. A next generation approach to species delimitation reveals the role of hybridization in a cryptic species complex of corals. BMC Evol. Biol. 19 (1), 116.

de Queiroz, K., 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: Howard, D.J., Berlocher, S.H. (Eds.), Endless forms: Species and speciation. Oxford University Press, New York (NY), pp. 57–75.

de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.

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

Rambaut, A., Drummond, A.J., 2007. Tracer v1. 6. http://beast.bio.ed.ac.uk.

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

Richards, Z.T., Hobbs, J.P.A., 2015. Hybridisation on coral reefs and the conservation of evolutionary novelty. Curr. Zool. 61, 132–145.

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

Ross, D.A., Uchupi, E., White, R.S., 1986. The geology of the Persian Gulf-Gulf of Oman region: a synthesis. Rev. Geophys. 24, 537–556.

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

Sánchez, J.A., Dorado, D., 2008. Intragenomic ITS2 variation in Caribbean seafans. Proc. 11th Int. Coral Reef Symp. 1383–1387.

Sheppard, C.R., Sheppard, A.S., 1991. Corals and Coral Communities of Arabia. Fauna of Saudi Arabia 12, 13–170.

Sheppard, C., Price, A., Roberts, C., 1992. Marine ecology of the Arabian region: patterns and processes in extreme tropical environments. Academic Press, London (UK).

Siddall, M., Rohling, E.J., Almogi-Labin, A., Hemleben, C., Meischner, D., Schmelzer, I., Smeed, D.A., 2003. Sea-level fluctuations during the last glacial cycle. Nature 423, 853.

Smith, L.W., Barshis, D., Birkeland, C., 2007. Phenotypic plasticity for skeletal growth, density and calcification of *Porites lobata* in response to habitat type. Coral Reefs 26, 559–567.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stampar, S.N., Maronna, M.M., Kitahara, M.V., Reimer, J.D., Morandini, A.C., 2014. Fast-evolving mitochondrial DNA in Ceriantharia: a reflection of Hexacorallia paraphyly? PLoS One 9, e86612.

Stobie, C.S., Cunningham, M.J., Oosthuizen, C.J., Bloomer, P., 2019. Finding stories in noise: mitochondrial portraits from RAD data. Mol. Ecol. Resour. 19, 191–205.

Terraneo, T.I., Benzoni, F., Arrigoni, R., Berumen, M.L., 2016. Species delimitation in the coral genus *Goniopora* (Scleractinia, Poritidae) from the Saudi Arabian Red Sea. Mol. Phylogenetics Evol. 102, 278–294.

Terraneo, T.I., Arrigoni, R., Benzoni, F., Forsman, Z.H., Berumen, M.L., 2018a. Using ezRAD to reconstruct the complete mitochondrial genome of *Porites fontanesii* (Cnidaria: Scleractinia). Mitochondrial DNA B 3, 173–174.

Terraneo, T.I., Arrigoni, R., Benzoni, F., Forsman, Z.H., Berumen, M.L., 2018b. The complete mitochondrial genome of *Porites harrisoni* (Cnidaria: Scleractinia) obtained using next-generation sequencing. Mitochondrial DNA B 3, 286–287.

Terraneo, T.I., Benzoni, F., Baird, A.H., Arrigoni, R., Berumen, M.L., 2019a. Morphology and molecules reveal two new species of *Porites* (Scleractinia, Poritidae) from the Red Sea and the Gulf of Aden. Syst. Biodivers. 17, 491–508.

Terraneo, T.I., Fusi, M., Hume, C.C., Arriogni, R., Voolstra, C.R., Benzoni, F., Forsman, Z. H., Berumen, M.L., 2019b. Environmental latitudinal gradients and host specificity shape Symbiodiniaceae distribution in Red Sea *Porites* corals. J. Biogeogr. 46, 2323–2335.

Tisthammer, K.H., Richmond, R.H., 2018. Corallite skeletal morphological variation in Hawaiian *Porites lobata*. Coral Reefs. 37, 445–456.

Todd, P.A., 2008. Morphological plasticity in scleractinian corals. Biol. Rev. 83, 315–337.

Toonen, R.J., Puritz, J.B., Forsman, Z.H., Whitney, J.L., Fernandez-Silva, I., Andrews, K. R., Bird, C.E., 2013. ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ. 1, e203.

Vénec-Peyré, M.T., Caulet, J.P., 2000. Paleoproductivity changes in the upwelling system of Socotra (Somali Basin, NW Indian Ocean) during the last 72,000 years: evidence from biological signatures. Mar. Micropal. 40, 321–344.

Veron, J.E.N., 2000. Corals of the World. Australian Institute of Marine Science, Townsville (Australia).

Veron, J., Stafford-Smith, M., DeVantier, L., Turak, E., 2015. Overview of distribution patterns of zooxanthellate Scleractinia. Front. Mar. Sci. 1, 81.

Vollmer, S.V., Palumbi, S.R., 2002. Hybridization and the evolution of reef coral diversity. Science 296, 2023–2025.

Wiens, J.J., Donoghue, M.J., 2004. Historical biogeography, ecology and species richness. Trends Ecol. Evol. 19, 639–644.

Willis, B.L., van Oppen, M.J., Miller, D.J., Vollmer, S.V., Ayre, D.J., 2006. The role of hybridization in the evolution of reef corals. Annu. Rev. Ecol. Evol. Syst. 37, 489–517.

Willis, S.C., Hollenbeck, C.M., Puritz, J.B., Gold, J.R., Portnoy, D.S., 2017. Haplotyping RAD loci: an efficient method to filter paralogs and account for physical linkage. Mol. Ecol. Resour. 17, 955–965.

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2013. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620.