**A peer-reviewed version of this preprint was published in PeerJ on 21 November 2017.**

View the peer-reviewed version (peerj.com/articles/4077), which is the preferred citable publication unless you specifically need to cite this preprint.

Devlin-Durante MK, Baums IB. 2017. Genome-wide survey of single-nucleotide polymorphisms reveals fine-scale population structure and signs of selection in the threatened Caribbean elkhorn coral, *Acropora palmata*. PeerJ 5:e4077 https://doi.org/10.7717/peerj.4077

# Genome-wide survey of single-nucleotide polymorphisms reveals fine-scale population structure and signs of selection in the threatened Caribbean elkhorn coral, *Acropora palmata*

**Meghann D Durante** [1], **Iliana B Baums** [Corresp. 1]

[1] Biology, Pennsylvania State University, University Park, PA, United States

Corresponding Author: Iliana B Baums
Email address: baums@psu.edu

The advent of next-generation sequencing tools has made it possible to conduct fine-scale surveys of population differentiation and genome-wide scans for signatures of selection in non-model organisms. Such surveys are of particular importance in sharply declining coral species, since knowledge of population boundaries and signs of local adaptation can inform restoration and conservation efforts. Here, genome-wide surveys of single-nucleotide polymorphisms in the threatened Caribbean elkhorn coral, *Acropora palmata*, reveal fine-scale population structure and place the major barrier to gene flow that separates the eastern and western Caribbean populations between the Bahamas and Puerto Rico. The exact location of this break had been subject to discussion because two previous studies based on microsatellite data had come to differing conclusions. We investigate this contradiction by analyzing an extended set of 12 microsatellite markers including the five previously employed and discovered that one of the original microsatellite loci is apparently under selection. Exclusion of this locus reconciles the results from the SNP and the microsatellite datasets. Scans for outlier loci in the SNP data detected 12 candidate loci under positive selection. Together, these results suggest that restoration of populations should use local sources and utilize existing functional variation among populations in *ex situ* crossing experiments to improve stress resistance of this species.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3043v1 | CC BY 4.0 Open Access | rec: 22 Jun 2017, publ: 22 Jun 2017

# Genome-wide survey of single-nucleotide polymorphisms reveals fine-scale population structure and signs of selection in the threatened Caribbean elkhorn coral, *Acropora palmata*

Meghann Devlin-Durante[1], Iliana B Baums[1]

[1]Department of Biology, Pennsylvania State University, University Park, PA, USA

Corresponding author:
Iliana B. Baums

Email address: baums@psu.edu

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3043v1 | CC BY 4.0 Open Access | rec: 22 Jun 2017, publ: 22 Jun 2017

Abstract

The advent of next-generation sequencing tools has made it possible to conduct fine-scale surveys of population differentiation and genome-wide scans for signatures of selection in non-model organisms. Such surveys are of particular importance in sharply declining coral species, since knowledge of population boundaries and signs of local adaptation can inform restoration and conservation efforts. Here, genome-wide surveys of single-nucleotide polymorphisms in the threatened Caribbean elkhorn coral, *Acropora palmata*, reveal fine-scale population structure and place the major barrier to gene flow that separates the eastern and western Caribbean populations between the Bahamas and Puerto Rico. The exact location of this break had been subject to discussion because two previous studies based on microsatellite data had come to differing conclusions. We investigate this contradiction by analyzing an extended set of 12 microsatellite markers including the five previously employed and discovered that one of the original microsatellite loci is apparently under selection. Exclusion of this locus reconciles the results from the SNP and the microsatellite datasets. Scans for outlier loci in the SNP data detected 12 candidate loci under positive selection. Together, these results suggest that restoration of populations should use local sources and utilize existing functional variation among populations in *ex situ* crossing experiments to improve stress resistance of this species.

## Introduction

There is an ongoing debate about the importance of local recruitment and barriers to gene flow in marine species. Many marine species reproduce via planktonic larvae and strong ocean currents have the potential to carry propagules over long distances. However, genetic evidence has revealed a high degree of self-recruitment in a range of species with planktonic larval duration being a poor predictor of genetic structure (Selkoe & Toonen 2011). The development of cheap genome-scale genotyping is poised to open a new chapter in this discussion (Peterson et al. 2012a; Toonen et al. 2013; Wang et al. 2012). American eels for example show panmixia in their central breeding ground in the North Atlantic but single nucleotide polymorphism (SNP) genotyping of adults along the Eastern seaboard revealed local differentiation (Gagnaire et al. 2012). Thus, a well-mixed pool of larvae sorted into environmental niches resulting in a structured adult population.

Resolving significant population genetic structure indicative of predominantly local recruitment is of particular significance for *in situ* restoration efforts targeting declining populations of reef-building corals (Baums 2008 - 2011; Epstein et al. 2001; Griffin et al. 2015; Griffin et al. 2012; Rinkevich 2006; Schopmeyer et al. 2012). Restoration genetic best practices suggest that propagules should not be moved among genetically distinct populations to avoid outbreeding depression (Baums 2008 - 2011) and this approach has been adopted by some permitting agencies (e.g. the Florida Keys National Marine Sanctuary). On the other hand, crosses between genotypes from different populations may show heterosis with respect to environmental stressors as seen in the hybrid *Acropora prolifera* (Fogarty 2012) and therefore might be worth exploring in an *ex situ* setting.

Diversity within functional regions of the genome that may be under selection (those regions that code for proteins or regulate transcription of genes), are not commonly surveyed even though it is these regions of the genome that are of interest to conservation managers who want to understand how much capacity there is in a species to adapt to changing conditions (Becks et al. 2010). Statistical methods have been developed that allow scanning of SNP loci for signatures of selection. Despite the risk of generating false positive results (Vilas et al. 2012), these methods yield candidate loci that should be substantiated by further testing to be of functional significance. The same methods can be used to scan microsatellite loci for signatures of selection, however, power is often limited by the small number of assayed loci.

SNPs are ubiquitous throughout the genome, located in coding and non-coding regions, and each locus has a maximum of four alleles (the four bases). This is in contrast to microsatellite loci that consist of tandem repeats, in which allelic variation is determined by the number of tandem repeats and thus can be large. The limited number of alleles at each SNP locus requires a larger number of loci to be assayed to achieve the same power of detecting population genetic structure as a panel of microsatellite loci (Morin et al. 2009; Ryman et al. 2006). The advent of reduced representation sequencing methods have made it possible to develop and assay a large number of SNP loci at a reasonable cost (Altshuler et al. 2000; Hoffberg et al. 2016). Recently, Genotyping by Sequencing (GBS) data including 4,764 SNPs in *A cervicornis* identified population structure within the Florida Reef tract (Drury et al. 2016b; Willing et al. 2012). Other flavors of reduced representation sequencing methods (Toonen et al. 2013; Wang et al. 2012) have yielded information on population structure, loci under selection and genetic diversity in reef building corals (Drury et al. 2016b; Howells et al. 2016a)

*Acropora palmata* is one of a few Caribbean coral species whose population genetic structure has been thoroughly investigated on local and range-wide scales (Baums et al. 2014a; Baums et al. 2005b; Baums et al. 2006a). A range-wide survey of *A. palmata* population genetic structure using five coral specific polymorphic microsatellite markers showed that *A. palmata* stands are structured into two long-separated populations (Baums et al. 2005a). While most reefs are self-recruiting, *A. palmata* stands are not inbred and harbor high genetic diversity at the microsatellite loci (Baums et al. 2005b). Bio-physical modeling identified a transient feature in the Mona Passage important in restricting present-day gene flow between the eastern and western population (Baums et al. 2006b). However, it is unclear whether the eastern and western populations differentiated initially due to selection. Subsequent denser sampling of *A. palmata* along the Antilles Island Arc raised the possibility of a hybrid zone across Puerto Rico rather than a clear-cut break between the eastern and western Caribbean at the Mona Passage (Mège et al. 2014).

The east-west population divide or possible finer scales of population differentiation were tested in this study by developing a large number of SNP markers to obtain a more comprehensive estimate of genetic differentiation across the genome and compare them to a set of microsatellite loci (Baums et al. 2009). The second goal was to screen SNP loci for signatures of selection. We developed genome-wide SNPs and assayed them in archived samples from two

regions in the western *A. palmata* population (Bahamas and Florida) and two regions in the eastern population (Puerto Rico and the U.S. Virgin Islands (USVI)). We then compared the results to population structure derived from ten and eleven microsatellite loci.

## Materials & Methods

### SAMPLE COLLECTION

Colonies of *A. palmata* were collected between 2002 and 2010 and previously genotyped (Baums et al. 2014a; Baums et al. 2005b). Unique genets were selected from our database for a total of 24 samples from each of four regions; the Bahamas, Florida, Puerto Rico and the US Virgin Islands (USVI). The goal was to have eight samples from three different reefs within each region, however this was not always possible either due to a small sample sizes from a particular reef or low clonal diversity of a reef. In those cases, we selected additional unique genets from nearby reefs. See Table 1A for detailed sample information.

We used an extended set of samples to compare the population genetic structure ascertained via microsatellite genotyping to the SNP results. This extended set of samples included 260 samples from six regions; Belize, Florida, Puerto Rico, the USVI, and Curacao (Table 1B).

### LIBRARY PREPARATION

Coral tissue samples were extracted from ethanol preserved samples using DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) with the following modifications. Time of incubation in the extraction buffer was increased to 16-20 hours and two 100 µl elutions were performed, the second of which was kept for library production as this fraction contained the high molecular weight DNA. Extracted DNA was then treated with 0.01 mg of RNase A (10 mg/ml, Amresco Solon, OH). Extraction concentrations ranging from 500 ng to 6 µg were double-digested with 10 units of each of the restriction enzymes MluCI (^AATT) and NlaIII (CATG^) (New England Biolabs, Ipswich, MA) following the protocol described by Peterson et al. (2012a). Digestions were purified using 1.5X Ampure beads (Beckman Coulter Inc, Brea, CA) and quantified on a Qubit® fluorometer (Life Technologies, Carlsbad, CA). Digested DNA was standardized to 100 ng for each sample before adaptor ligation. Samples were identified with eight 6-bp indices on the NlaIII (rare-cutter) P1 adapter (Supplementary Table 1). Samples were pooled into 12

libraries and then size selected in the range of 200-800 bp on a Pippin-Prep (Sage Science, Beverly, MA). Next, Illumina flow-cell annealing sequences, unique multiplexing indices and sequencing primer annealing regions were added through PCR amplification to the MluCL cut end (See (Peterson et al. 2012b), Protocol S1, Figure 1). The libraries were enriched with 12 amplification cycles in four separate PCR reactions for each library containing 10 µl of Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs, Ipswich, MA), 2 µl of each amplification primer, 1 µl of library DNA and 5 µl of water (total 20 µl). Samples were pooled into four libraries each containing 24 samples (Table 2). Each library was sequenced on one lane of Illumina HiSeq 2000 sequencer (paired-end, 2x150 bp) at the Pennsylvania State Genomics Core Facility. There were two libraries sequenced on each chip.

## RAW SEQUENCE FILTERING

Raw sequence reads were filtered using the process_radtags in the pipeline STACKS 1.21 (Catchen et al. 2013; Catchen et al. 2011). Barcodes and the RAD-Tag cut sites were identified to de-multiplex the pooled data into individual samples. Reads were discarded that had low quality (with an average raw phred score <10 within a 15-base pair sliding-window), adapter contamination, and uncalled bases. Since all indices differed by at least 2 bp, it was possible to correct and retain any index that differed by a single bp from an expected index.

## ASSEMBLY

Processed sequences were then aligned to the *Acropora digitifera* genome (V1.0) (Shinzato et al. 2011) with BOWTIE2 (Langmead & Salzberg 2012) within the GALAXY (Bedoya-Reina et al. 2013; Blankenberg et al. 2014) framework using end-end read alignment settings in order to remove symbiont and other associated microorganisms. After alignment, paired-end sequencing BAM files were assembled in the ref_map.pl pipeline in STACKS 1.30 with the following parameters. Each paired-end sequencing set was run separately through STACKS to compare results (designated Read1 and Read2). The number of raw reads required to report a stack was m=5. The number of mismatches allowed between loci when building the catalog was n=4. SNPs with a log-likelihood of less than -10 were removed as reads with poor log-likelihoods tend to have sequencing error and/or low coverage. Two of the barcodes (TCGAT and CGATC) had a very low amount of sequence reads across all four populations, all Illumina lanes were affected, and those samples were removed before assembly in STACKS.

### GENOME COVERAGE

BEDTOOLS (Quinlan & Hall 2010) was used to create a histogram of genome coverage for each sample from the BOWTIE2 BAM format alignment files. All positions with a depth of coverage greater to or equal to 20 were combined into a single bin in the histogram. Regions were averaged (excluding samples with barcodes TCGAT and CGATC) and a cumulative distribution of sequencing coverage was then plotted in SIGMAPLOT v12.

### POPULATION GENETIC STATISTICS

More stringent filtering was implemented in the Populations module in STACKS 1.30 where a locus had to be present in at least 60% of the individuals within a population and had to be present in all four populations in order to be processed for $F_{ST}$ calculations and outlier analysis. A minor allele frequency (MAF) cutoff of 0.05 was applied. A *p*-value correction was applied to $F_{ST}$ scores, so that if a $F_{ST}$ score is not significantly different from 0 (according to Fisher's Exact Test) the value was set to 0. Additionally, only one random SNP from any RAD locus was written to the STRUCTURE export file in order to prevent linked loci from being processed. Read 1 and Read 2 STRUCTURE export files were combined and duplicate loci removed randomly between reads. An Analysis of Molecular Variance (AMOVA) corrected $F_{ST}$ (*p*-value<0.05) was calculated within STACKS.

### CLUSTERING ANALYSES

Clustering analyses were performed in the program STRUCTURE 2.3.4 (Falush et al. 2003; Hubisz et al. 2009) using the admixture model with correlated allele frequencies. The analysis included the following parameters: 100,000 burn-in iterations and 1,000,000 Markov chain Monte Carlo repetitions, with and without a population prior, for a total of three replicates for each value of K. K values ranged from 2 to 5. The most likely value for K was determined by CLUMPAK (KOPELMAN ET AL. 2015B) BEST K which uses LN(PR(X|K) to identify the K for which PR(K=K) is the highest as described in STRUCTURE's manual section 5.1. Results of the three structure runs were merged with CLUMPAK (Kopelman et al. 2015a).

Previously genotyped samples at 10 and 11 (n=260) microsatellite markers (181, 182, 192, 207, 0585, 0513, 2637, 007, 9253, 5047, with and without locus 166) (Baums et al. 2009; Baums et al. 2005a) were also analyzed with STRUCTURE 2.3.4 (Falush et al. 2003; Hubisz et al. 2009) using the admixture model with correlated allele frequencies (See Table 1 for sample

information). The analysis included the following parameters; 100,000 burn-in iterations and 1,000,000 MCMC repetitions, with and without a population prior, for a total of 3 replicates for each value of K. K values tested ranged from 2 to 7.

### MANTEL TESTS

Data on temperature, salinity, dissolved oxygen (ml/l), and phosphates was downloaded from the World Ocean Atlas 2013 (WOA13 V2). Silicates and nitrates were not used as there was not sufficient data for all locations. For the Bahamas, Puerto Rico, and the USVI the geographic center point among several sampling sites was used because reefs were further apart than in Florida. For all data, the statistical mean of the annual average of years 1955-2012 and depths of 0-10 m was used. Grid sizes were 1/4° for temperature and salinity, and 1° for dissolved oxygen (ml/l), and phosphates (µmol/l) (Supplemental Table 2). SPSS V22 was used to calculate a dissimilarity matrix expressed as the Euclidean distances between regions based on the above environmental data. GenAlEx v6.501 (Peakall & Smouse 2006) was used to calculate a pairwise geographic distance matrix between the four regions and to perform a Mantel multi-comparison test between the geographic distance matrix, $F_{ST}$ pairwise matrix between regions from STACKS, and the environmental dissimilarity matrix.

### OUTLIER ANALYSIS

Two independent methods were used to identify putative loci under selection. The first program used was LOSITAN (Antao et al. 2008) which utilizes the method of Beaumont and Nichols (1996) to identify loci under selection based on the joint distributions of expected heterozygosity and $F_{ST}$ under an island model of migration. The following settings were used for the SNP and the microsatellite datasets. The neutral mean setting was selected in which during an initial run (100,000 simulations), a candidate subset of selected loci (outside the 95 % confidence interval) were identified and removed. Then the distribution of neutral $F_{ST}$ was computed using 100,000 simulations and a bisection approximation algorithm (Antao et al. 2008), with the following options, force mean $F_{ST}$, infinite alleles mutation model, and a confidence interval 0.99. A FDR < 0.1 correction for multiple testing was applied. Loci outside the upper and lower confidence areas were identified as candidates affected by positive and balancing selection, respectively. All populations were analyzed together. The positive outlier loci were blasted against the NCBI nr, UniProt, and Trembl databases with parameters of

expected value = 0.00001, gap opening penalty = 11, gap extension penalty = 1, length of initial exact match (word size) = 6 and scoring matrix = BLOSUM62 using BLASTX 2.2.32+ (Altschul et al. 1997).

The STACKS exported GENEPOP dataset was also reformatted with PGDSPIDER version 2.0.5.2 (Lischer & Excoffier 2012) to a GESTE file. The method of Foll and Gaggiotti (2008) was performed using BAYESCAN 2.0 (http://www-leca.ujf-grenoble.fr/logiciels.html). For each locus, the probability of being under selection was inferred using the Bayes factor (BF). Based on Jeffreys' (1961) (Jeffreys 1961) scale of evidence, a log10 BF of 1.5–2.0 is interpreted as "strong evidence" of selection. For our analysis, the estimation of model parameters was set as 20 pilot runs of 5,000 iterations each, followed by 50,000 iterations.

## Results

### Summary statistics

Illumina sequencing of the RAD libraries generated 49.3 million reads per pool of eight samples, averaging 6.2 million 150 bp reads per individual prior to quality filtering. After quality filtering, 4.99 million (81%) reads per individual were retained on average (Table 2). Pools had similar numbers of reads after processing (mean = 39.9 million per pool, SD = 4.95 million, one-way ANOVA, $F = 2.638$, $p > 0.1$). The average % GC content for Read 1 and 2 was 41.7 and 39.6, respectively. The percentage of polymorphic sites per region varied little among populations, from 0.150 to 0.173 % (Table 3). The observed heterozygosity in variant sites was 22% on average. Overall $F_{IS}$ values, when considering all sites with a minor allele frequency cutoff a $\geq 0.05$, were close to 0 and hence provided no evidence of inbreeding (Table 3). However, when only considering variant positions within the region of Florida, $F_{IS}$ values were negative ($F_{is} = -0.0086$), indicating an excess of heterozygosity. Using the two paired-end read sets as replicates, a one-way ANOVA was performed for each variable (Table 4). Populations were found to be similar for all summary statistics. Alignment of *A. palmata* SNPs to the published *A. digitifera* genome indicated that on average, 2.5% percent of the *A. digitifera* genome had sequence coverage at a stack depth of 5 (Fig 1). All four populations produced similar sequence coverage.

Population genetics

A total of 390 SNPS were identified after filtering and including a minor allele frequency cutoff a ≥ 0.05 (Table 3). This included 219 for Read1 and 176 for Read2 from the paired-end sequencing (5 SNPs were identical between reads and only considered once). Analysis of Molecular Variance (AMOVA) revealed patterns of genetic differentiation among populations (Table 4). This was also evident when the 307 SNPs (analysis included only one SNP per 150 bp locus) after combining Read1 and Read2, were subjected to a multi-locus clustering analysis in STRUCTURE. Individuals from Florida clustered first, followed by the Bahamas at K=3. Puerto Rico and the USVI were not distinguishable until K=4, (Fig 2). CLUMPAK BEST K (Kopelman et al. 2015b) indicated that K=3 was the most likely K-value regardless of whether the sampling region was used as a prior.

To compare to the SNP analysis, microsatellite data from samples collected in six regions were analyzed in STRUCTURE using the sampling region as a prior. At K=2, a western (including Belize, Florida, Bahamas and Puerto Rico) and an eastern cluster (including the USVI and Curacao) was evident (Fig 3A). At K=3, an isolation-by-distance like pattern was apparent in the western cluster (Fig 3B). K=4 was the most likely K-value based on 11 microsatellite markers (Kopelman et al. 2015b) which grouped Florida and Belize as one cluster, and Puerto Rico and the Bahamas as the second, with the USVI as the third and Curacao as an admixed fourth cluster (Fig 3C).

According to the outlier analysis in Lositan, locus 166 was identified as a potential outlier and thus possibly under selection. It was therefore excluded from the analysis in STRUCTURE. This resulted in more comparable results to the SNP analysis with the most likely K-value being 3 (Kopelman et al. 2015b). Again, the first separation was between a western and an eastern cluster, however this time Puerto Rico assigned to the eastern cluster with an isolation-by-distance like pattern appearing between the west and east (Fig 3D). At the most likely K of 3, Curacao now formed a separate cluster. At K=4, the Bahamas started to separate from the remainder of the western regions similar to what was observed in the SNP clustering analysis (Fig 3E).

Environmental drivers of population structure

A Mantel test showed a significant positive relationship in the SNP dataset between pairwise $F_{ST}$ values and geographic distance ($R^2$= 0.65, $p$=0.05) consistent with the microsatellite results (10 loci) from the Florida, Bahamas, Puerto Rico, and Curacao samples only (Fig4C, Fig4D). Correlations between environmental factors including average temperature, salinity, dissolved oxygen, and pairwise $F_{ST}$ values or geographic distance were not significant (Fig4A, Fig4B). However, it should be noted that the environmental data had a resolution of ¼ to 1 degree latitude whereas the genetic data was collected on much smaller spatial scales (the reefs in each region are on average 66, 10, 106, and 68 km apart for the Bahamas, Florida, Puerto Rico and the U.S. Virgin Islands, respectively)(Supplemental Table 2). Therefore, landscape genetic approaches that may reveal environmental drivers of population differentiation (Manel et al. 2003) must await higher resolution environmental data.

Loci under selection

BAYESCAN and LOSITAN identified 2 and 13 SNPs (Supplementary Table 3) that showed signs of positive selection when including all four populations, one of which was identified by both programs (a total of 12 unique loci identified between both programs). Outliers accounted for 3.3% of the total SNPs, consistent with other studies in which $F_{ST}$ outlier loci have represented a substantial fraction of the total loci investigated (2-10%) (Nosil et al. 2009). Annotation of the candidate loci proved difficult as only 23% produced significant hits when queried against the NCBI NR database, Uniprot, and Trembl; with two of the hits being annotated as unconventional myosin-IXb isoform X7 and tyrosine-protein kinase transmembrane receptor ROR1-like. Screening of the microsatellite loci identified locus 166 as an outlier under positive selection, yet no annotation information of this locus is currently available.

## Discussion

### Comparison with previous *Acropora* gene flow studies

The previous range-wide survey of *A. palmata* population genetic structure using five, presumed neutrally evolving microsatellite markers showed that while most reefs are self-recruiting, *A. palmata* stands are not inbred and harbor high microsatellite genetic diversity (Baums et al. 2005). Furthermore, *A. palmata* stands were structured into two long-separated

populations, one in the eastern and one in the western Caribbean (Baums et al. 2005). Here, we report that genome-wide SNPs (MAF ≥ 0.05) resolved further population structure in the endangered reef-building coral, *A. palmata* from Florida to the USVI compared to previous microsatellite-based analyses.

It was recently suggested that the East-West divide of *A. palmata* lies not in the Mona Passage (Baums et al. 2005b; Baums et al. 2006b) but rather to the east of Puerto Rico (Mège et al. 2014). The 307 SNPs analyzed here confirm earlier findings that Puerto Rico and the USVI regions are more similar to each other than Puerto Rico is to either the Bahamas or Florida without imposing any priors in a STRUCTURE analysis (MAF≥0.05). However, it is not always possible to determine, with confidence, the correct clustering solution that accurately reflects genetic population structure when there is an underlying isolation by distance pattern (Frantz et al. 2009). We show here that there is significant isolation by geographic distance from Florida to the USVI when using presumably neutrally evolving SNP and microsatellite loci. Interestingly, inclusion of microsatellite locus 166, flagged as being an outlier locus, obscured this isolation by distance pattern. Therefore, locus 166 is a strong candidate for a locus under selection (or it is linked to a locus under selection) and its functional significance might prove a fruitful subject for future studies.

An east-west Caribbean divide was also evident in the corals *Orbicella annularis* (Foster et al. 2012) and *Acropora cervicornis* (Vollmer & Palumbi 2007). An additional barrier to gene flow in *A. palmata* was reported by Porto-Hannes et al. (2014) between Venezuela and the Mesoamerican Barrier Reef System utilizing four of the microsatellites markers.

The total number of SNPs (n=307) retained for population genetic analysis was lower than expected. This was due to a 10-fold increase in the number of fragments retrieved from the genome digest using the enzymes MluCI (^AATT) and NlaIII (CATG^) compared to what was predicted from an in-silico restriction of an incomplete draft genome of *A. palmata* (Baums, unpublished). A larger set of SNP loci may reveal additional finer scale structure across the Caribbean in *A. palmata*. However, model based clustering methods of 905,561 SNPs failed to reveal population structure in *A. digitifera* collected from the Ryukyu Archipelago of Japan, although a principle component analysis clustered the 122 samples into 4 groups identified as Okinawa, Kerama, Yaeyama-North, and Yaeyama-South, respectively (Shinzato et al. 2015). Low coverage, 5X in this study, is also a concern. Yet in the coral

*Platygyra daedalea,* 5x coverage was sufficient to assign samples to two distinct clusters based on their geographic origin, the Persian Gulf or Sea of Oman and was consistent with their 20x coverage data set (Howells et al. 2016b).

**Genetic diversity indices in *A. palmata***

Several factors could account for negative $F_{IS}$ values including negative assortative mating, if a species is outcrossed and lacks selfed progeny or there is a selection pressure that favors the most heterozygous individuals. Of our samples, 49 out of 96 were ramets of larger genets. *A. palmata* colonies fragment frequently; the branches regrow into new colonies resulting in stands of genetically identical colonies (Baums et al. 2006a). [Note that samples included here all represented distinct genets]. Asexual reproduction could explain the excess of heterozygosity in *A. palmata* within the Florida region (see (Balloux et al. 2003; Carlon 1999; Delmotte et al. 2002). Excess hetereozygosity has been observed in other clonal organisms. For example, significant negative $F_{IS}$ values in a partially clonal but self-incompatible wild cherry tree was explained in part by asexual reproduction (Stoeckel et al. 2006).

The nucleotide diversity $\pi$, describes the degree of nucleotide polymorphism in a population and can be calculated based on variant sites only or on variant and non-variant sites combined. In *A. cervicornis*, nucleotide diversity based on variant sites only ranged from 0.239–0.44, with all means of reefs in Florida being higher than the Dominican Republic. (Drury et al. 2016b). However, in *A. palmata* we find that Florida is the least genetically diverse region when comparing variant sites only (0.203, Table 3), as would be expected in a marginal environment (Arnaud-Haond et al. 2006; Baums 2008; Baums et al. 2014b; Cahill & Levinton 2016; Eckert et al. 2008). In *Acropora austera* populations in the south-west Indian Ocean, nucleotide diversity ranged from 0.007 to 0.022, with lower estimates in the south than north (Macdonald et al. 2011). The nucleotide diversity estimate for *A. cervicornis*, including variant and non-variant SNP sites was 0.09 (Drury et al. 2016a). In the sea anemones, *Aiptasia* and *Nematostella* (Cnidaria) a genome-wide estimate of nucleotide diversity was 0.004 SNPs/bp surveyed (Bellis et al. 2016) and 0.0065 SNPs/bp (Putnam et al. 2007), respectively. In a survey of transcriptome derived SNPs in three gorgonian species synonymous nucleotide diversity ranged from 0.012 – 0.020 (Romiguier et al. 2014). Average pairwise nucleotide diversity in other metazoans include *Caenorhabditis elegans* (~0.001 SNPs/bp, (Swan et al. 2002)), *Drosophila*

*pseudoobscura* [~0.002 SNPs/bp (Kulathinal et al. 2009), and *homo sapiens* [$7.51 \times 10^{-4}$ SNPs/bp (Sachidanandam et al. 2001)].

Our estimates of nucleotide diversity (including variant and non-variant sites) was 0.0004 SNPs/bp for all populations, an order of magnitude lower than in other cnidarians. Based on a survey of 374 individual transcriptome derived SNPs from 76 non-model animal species, the level of nucleotide diversity found in *A. palmata* is well below that predicted for a long-lived species, with small propagule size and large adult size (Romiguier et al. 2014). This low nucleotide diversity could be due to either a relatively small long-term effective population size, a severe bottleneck associated with a selective sweep (Ellegren & Galtier 2016) or the small number of SNPs included in this study (Fischer et al. 2017).

Allelic richness of microsatellite data correlates better with genome-wide estimates of genetic diversity based on SNPs than heterozygosity (Fischer et al. 2017) and allelic richness is more sensitive to recent population bottlenecks than heterozygosity (Allendorf 1986). Average microsatellite-based allelic richness in 14 Indo-Pacific *Acropora* corals was 4.96 overall and 6.21 in the five geographically widespread species (calculated based on Table 6 in Richards & Oppen 2012) which compares favorably with an average allelic richness of 8.49 in *A. palmata* found here. Thus, allelic richness of microsatellite loci remains high in Caribbean *A. palmata* despite recent population declines and the documented loss of alleles in Florida (Williams et al. 2014).

Future studies should include several thousand SNPs assayed in samples from across the species range to provide conclusive data on the impact of recent population declines on overall genetic diversity in *A. palmata*.

**Genes under positive selection**

One of the SNP loci identified as being under positive selection was annotated as a *tyrosine-protein kinase transmembrane receptor ROR1-like*. ROR receptor protein is associated with the nervous system in the fruit fly *Drosophila* (Wilson et al. 1993), nematode *C. elegans* (Francis et al. 2005), and sea slug *Aplysia californica* (McKay et al. 2001). Functional analysis of *cam-1*, a gene that encodes for a ROR kinase in *C. elegans*, demonstrated roles in both the orientation of polarity in asymmetric cell division and axon outgrowth, and the ability to guide

migrating cells (Forrester et al. 1999). The role of *ROR1* receptors in Cnidaria is unknown although studies in *Hydra* suggest a function in regulating cell specification and tissue morphogenesis (Bertrand et al. 2014; Krishnapati & Ghaskadbi 2014; Lange et al. 2014).

Another SNP identified as being under positive selection was located in the gene annotated as unconventional *myosin-IXb isoform X7*, a Rho GTPase-activating protein (RhoGAP) that is essential for coordinating the activity of Rho GTPases. Invertebrates are thought to contain a single myosin class IX gene (the exception is *Drosphilia* which has none) whereas most vertebrates have two with fishes having four (Liao et al. 2010). In general, Rho GTPases control the assembly and organization of the actin cytoskeleton which includes many functions such as cell adhesion, contraction and spreading, migration, morphogenesis, and phagocytosis. Little is known about the function of myosin-IX in invertebrates. However, a recent study in which *Orbicella faveolata* were exposed to immune challenges identified Unconventional myosin-IXb as a transcript that was significantly correlated with melanin protein activity (Fuess et al. 2016). In humans, Myosin-IXb is highly expressed in tissues of the immune system such as the lymph nodes, thymus, and spleen and also in immune cells like dendritic cells, macrophages and CD4 + T cells (Wirth et al. 1996). Myosin-IXb knockout mice showed impaired recruitment of monocytes and macrophages when exposed to a chemoattractant demonstrating that Myosin-IXb has an important function in innate immune responses *in vivo* (Hanley et al. 2010). Because statistical screens for loci under selection carry a high rate of false positive results, further experimental evidence is necessary before these loci can be considered targets of selection.

**Restoration implications**

Restoration efforts should proceed under the assumption that *A. palmata* harbors a significant amount of population structure requiring close matches of collection and outplant sites. Hybridization of *A. palmata* from different regions may or may not result in heterosis depending on compatibility, but would be worth pursuing in an *ex situ* setting to enable close monitoring of offspring performance under elevated temperatures (van Oppen et al. 2015). With respect to the sharply declining Florida colonies, these findings underline the need to manage and restore Florida's *A. palmata* as an isolated, genotypically depleted population (Williams et al. 2014).

**Acknowledgements**

Thanks to PSU genome sequencing facility for expert library preparation and sequencing.

**Figure legends**

**Figure 1** A Cumulative distribution of sequencing coverage of Bowtie2 aligned reads to the *A. digitifera* using Bedtools. On average 2.5% percent of the *A. digitifera* genome had sequence coverage at a stack depth of 5.

**Figure 2** Bayesian cluster analysis with STRUCTURE. Reefs within regions 1-4 sorted by latitude: Florida, Bahamas, Puerto Rico, US Virgin Islands. Analysis of 307 SNPs (analysis included only one SNP per locus) after combining Read1 and Read2. Panels K=2 (A), and K = 3 (B), K=4 (C). The most probable K was 3 (B) for the minor allele frequency corrected SNPs based on the mean estimated log probability of the data at a given K (3 replicate runs per K, +/- 1 standard deviation).

**Figure 3** Bayesian cluster analysis with STRUCTURE. Panels (A-C). Analysis of 11 microsatellites with the most probable K being 4. Panels (D-E). Exclusion of locus 166 that was identified as an outlier resulted in an analysis of 10 microsatellites with the most probable K being 3.

**Figure 4** MANTEL matrix correlation test between genetic and geographic distances, and environmental parameters as calculated by a dissimilarity matrix expressed as the Euclidean distances between regions based on measured environmental data. *Acropora palmata* samples from four regions (Florida, Bahamas, Puerto Rico and USVI) were genotyped with 307 SNP (a, c) or 10 neutral microsatellite markers (d). Panel (a) $y = 0.0107x + 0.0104$, $R^2 = 0.6104$, $p = 0.09$. Panel (b) $y=0.002x + 0.4175$, $R^2= 0.1012$, $p = 0.21$. Panel (c) $y=0.000007x + 0.0098$. $R^2= 0.6483$, p-value=0.05. Panel (d) $y=0.000007x + 0.0027$. $R^2= 0.69$, $p = 0.04$.

**Supplemental Figure** MANTEL matrix correlation test between genetic and geographic distances. *Acropora palmata* samples from four regions (Florida, Bahamas, Puerto Rico and USVI) were genotyped with 11 microsatellite markers, including the 166 outlier locus. $y= 5E-06x + 0.0137$. $R^2 = 0.1147$, p-value=0.1.

458

459 **Tables**

460

461 **Table 1** *Acropora palmata* colonies included in SNP analysis. Samples were obtained from four
462 regions in the Caribbean and 3 – 6 reefs per region. Given are latitude and longitude in decimal
463 degrees.

464 **Table 2** RAD-tag sequencing summary table of *Acropora palmata* samples.

465 **Table 3** Summary statistics for Read 1 and Read 2 combined. % PL = percent polymorphic loci,
466 Obs Hom = observed homozygosity, Obs Het = observed heterozygosity, StdErr = standard
467 error, Exp = expected. $F_{IS}$ calculations with and without minor allele frequency restrictions.
468 Calculated by STACKS 1.30.

469 **Table 4** Pairwise $F_{ST}$ calculated from STACKS 1.3. Read 1 and 2 combined (duplicated stacks
470 between reads removed, MAF≥0.05) calculated on the AMOVA corrected (p-value <0.05) $F_{ST}$
471 measurements. Considered were loci present in all populations.

472 **Supplementary Table 1** DD-Rad sequencing**.** There were 12 pools with 8 unique barcodes in
473 each. The Database ID is a unique identifier for each coral specimen. Given is also the total
474 number of ramets for each genet that was included in the RAD sequencing. The indices are short
475 DNA sequences that uniquely identify products in the final libraries.

476 **Supplemental Table 2** GPS coordinates in decimal degrees for the World Ocean Atlas
477 2013 (WOA13 V2) environmental data averaged for a region.

478 **Supplementary Table 3** Outlier SNPs identified by programs LOSITAN and BAYESCAN. Stacks
479 locus_bp is the STACKS program locus ID with the SNP location basepair after the underscore.
480 Read category indicates whether the outlier SNP was found in read 1 or 2 or the paired-end
481 sequencing run. Digitifera scaffold identifies the scaffold where the Stacks locus aligned to,
482 followed by the basepair location in the next column. S start= sequence start. S end = Sequence
483 end.

484

485

486

487 References

488

489 Allendorf FW. 1986. Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* 5:181-190.
490 10.1002/zoo.1430050212
491 Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, and Lander ES. 2000. An SNP map
492 of the human genome generated by reduced representation shotgun sequencing. *Nature*
493 407:513-516.
494 Antao T, Lopes A, Lopes RJ, Beja-Pereira A, and Luikart G. 2008. LOSITAN: A workbench to detect
495 molecular adaptation based on a Fst-outlier method. *Bmc Bioinformatics* 9:323.
496 Arnaud-Haond S, Teixeira S, Massa SI, Billot C, Saenger P, Coupland G, Duarte CM, and Serrao EA. 2006.
497 Genetic structure at range edge: low diversity and high inbreeding in Southeast Asian mangrove
498 (*Avicennia marina*) populations. *Molecular Ecology* 15:3515-3525. doi:10.1111/j.1365-
499 294X.2006.02997.x
500 Balloux F, Lehmann L, and de Meeus T. 2003. The population genetics of clonal and partially clonal
501 diploids. *Genetics* 164:1635-1644.
502 Baums IB. 2008. A restoration genetics guide for coral reef conservation. *Molecular Ecology* 17:2796-
503 2811. doi:10.1111/j.1365-294X.2008.03787.x
504 Baums IB. 2008 - 2011. Predicting the effects of ocean warming on larval dispersal by measuring
505 adaptive potential of corals. Caribbean: NSF Biological Oceanography.
506 Baums IB, Devlin-Durante MK, Brown L, and Pinzón JH. 2009. Nine novel, polymorphic microsatellite
507 markers for the study of threatened Caribbean acroporid corals. *Molecular Ecology Resources*
508 9:1155-1158.
509 Baums IB, Devlin-Durante MK, and LaJeunesse TC. 2014a. New insights into the dynamics between reef
510 corals and their associated dinoflagellate endosymbionts from population genetic studies.
511 *Molecular Ecology* 23:4203-4215. 10.1111/mec.12788
512 Baums IB, Durante MD, Laing AA, Feingold J, Smith T, Bruckner A, and Monteiro J. 2014b. Marginal coral
513 populations: the densest known aggregation of *Pocillopora* in the Galápagos Archipelago is of
514 asexual origin. *Frontiers in Marine Science* 1. 10.3389/fmars.2014.00059
515 Baums IB, Hughes CR, and Hellberg MH. 2005a. Mendelian microsatellite loci for the Caribbean coral
516 *Acropora palmata*. *Marine Ecology - Progress Series* 288:115-127.
517 Baums IB, Miller MW, and Hellberg ME. 2005b. Regionally isolated populations of an imperiled
518 Caribbean coral, *Acropora palmata*. *Molecular Ecology* 14:1377-1390.
519 Baums IB, Miller MW, and Hellberg ME. 2006a. Geographic variation in clonal structure in a reef building
520 Caribbean coral, *Acropora palmata*. *Ecological Monographs* 76:503-519. 10.1890/0012-
521 9615(2006)076[0503:GVICSI]2.0.CO;2
522 Baums IB, Paris CB, and Cherubin LM. 2006b. A bio-oceanographic filter to larval dispersal in a reef-
523 building coral. *Limnology and Oceanography* 51:1969-1981.
524 Beaumont MA, and Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population
525 structure. *Proceedings of the Royal Society of London Series B-Biological Sciences* 263:1619-
526 1626.
527 Becks L, Ellner SP, Jones LE, and Hairston NG, Jr. 2010. Reduction of adaptive genetic diversity radically
528 alters eco-evolutionary community dynamics. *Ecology Letters* 13:989-997. 10.1111/j.1461-
529 0248.2010.01490.x
530 Bedoya-Reina O, Ratan A, Burhans R, Kim H, Giardine B, Riemer C, Li Q, Olson T, Loughran T, vonHoldt B,
531 Perry G, Schuster S, and Miller W. 2013. Galaxy tools to study genome diversity. *GigaScience*
532 2:17.

533 Bellis ES, Howe DK, and Denver DR. 2016. Genome-wide polymorphism and signatures of selection in
534 the symbiotic sea anemone *Aiptasia*. *Bmc Genomics* 17:160.
535 Bertrand S, Iwema T, and Escriva H. 2014. FGF signaling emerged concomitantly with the origin of
536 Eumetazoans. *Molecular biology and evolution* 31:310-318.
537 Blankenberg D, Von Kuster G, Bouvier E, Baker D, Afgan E, Stoler N, Team G, Taylor J, and Nekrutenko A.
538 2014. Dissemination of scientific software with Galaxy ToolShed. *Genome Biology* 15:3.
539 10.1186/gb4161
540 Cahill AE, and Levinton JS. 2016. Genetic differentiation and reduced genetic diversity at the northern
541 range edge of two species with different dispersal modes. *Molecular Ecology* 25:515-526.
542 10.1111/mec.13497
543 Carlon DB. 1999. The evolution of mating systems in tropical reef corals. *Trends in Ecology & Evolution*
544 14:491-495.
545 Catchen J, Hohenlohe PA, Bassham S, Amores A, and Cresko WA. 2013. Stacks: an analysis tool set for
546 population genomics. *Molecular Ecology* 22:3124-3140. 10.1111/mec.12354
547 Catchen JM, Amores A, Hohenlohe P, Cresko W, and Postlethwait JH. 2011. Stacks: Building and
548 Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics* 1:171-
549 182. 10.1534/g3.111.000240
550 Delmotte F, Leterme N, Gauthier JP, Rispe C, and Simon JC. 2002. Genetic architecture of sexual and
551 asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite
552 markers. *Molecular Ecology* 11:711-723. 10.1046/j.1365-294X.2002.01478.x
553 Drury C, Dale K, Panlilio J, Miller S, Lirman D, Larson E, Bartels E, Crawford D, and Oleksiak M. 2016a.
554 Genomic variation among populations of threatened coral: Acropora cervicornis. *BMC genomics*
555 17:286.
556 Drury C, Dale KE, Panlilio JM, Miller SV, Lirman D, Larson EA, Bartels E, Crawford DL, and Oleksiak MF.
557 2016b. Genomic variation among populations of threatened coral: *Acropora cervicornis*. *Bmc*
558 *Genomics* 17:286. 10.1186/s12864-016-2583-8
559 Eckert CG, Samis KE, and Lougheed SC. 2008. Genetic variation across species' geographical ranges: the
560 central-marginal hypothesis and beyond. *Molecular Ecology* 17:1170-1188. doi:10.1111/j.1365-
561 294X.2007.03659.x
562 Ellegren H, and Galtier N. 2016. Determinants of genetic diversity. *Nat Rev Genet* 17:422-433.
563 10.1038/nrg.2016.58
564 Epstein N, Bak RPM, and Rinkevich B. 2001. Strategies for Gardening Denuded Coral Reef Areas: The
565 Applicability of Using Different Types of Coral Material for Reef Restoration. *Restoration Ecology*
566 9:432-442. 10.1046/j.1526-100X.2001.94012.x
567 Falush D, Stephens M, and Pritchard JK. 2003. Inference of population structure using multilocus
568 genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567-1587.
569 Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderegger R, and Widmer A.
570 2017. Estimating genomic diversity and population differentiation – an empirical comparison of
571 microsatellite and SNP variation in *Arabidopsis halleri*. *Bmc Genomics* 18:69. 10.1186/s12864-
572 016-3459-7
573 Fogarty ND. 2012. Caribbean acroporid coral hybrids are viable across life history stages. *Marine Ecology*
574 *Progress Series* 446:145-159.
575 Foll M, and Gaggiotti O. 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for Both
576 Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 180:977-993.
577 10.1534/genetics.108.092221
578 Forrester WC, Dell M, Perens E, and Garriga G. 1999. A *C. elegans* Ror receptor tyrosine kinase regulates
579 cell motility and asymmetric cell division. *Nature* 400:881-885.

580 Foster NL, Paris CB, Kool JT, Baums IB, Stevens JR, Sanchez JA, Bastidas C, Agudelo C, Bush P, Day O,
581 Ferrari R, Gonzalez P, Gore S, Guppy R, McCartney MA, McCoy C, Mendes J, Srinivasan A, Steiner
582 S, Vermeij MJA, Weil E, and Mumby PJ. 2012. Connectivity of Caribbean coral populations:
583 complementary insights from empirical and modelled gene flow. *Molecular Ecology* 21:1143-
584 1157. 10.1111/j.1365-294X.2012.05455.x
585 Francis MM, Evans SP, Jensen M, Madsen DM, Mancuso J, Norman KR, and Maricq AV. 2005. The Ror
586 receptor tyrosine kinase CAM-1 is required for ACR-16-mediated synaptic transmission at the C.
587 elegans neuromuscular junction. *Neuron* 46:581-594.
588 Frantz A, Cellina S, Krier A, Schley L, and Burke T. 2009. Using spatial Bayesian methods to determine the
589 genetic structure of a continuously distributed population: clusters or isolation by distance?
590 *Journal of Applied Ecology* 46:493-505.
591 Fuess LE, Weil E, and Mydlarz LD. 2016. Associations between transcriptional changes and protein
592 phenotypes provide insights into immune regulation in corals. *Developmental & Comparative*
593 *Immunology* 62:17-28.
594 Gagnaire PA, Normandeau E, Cote C, Hansen MM, and Bernatchez L. 2012. The Genetic Consequences
595 of Spatially Varying Selection in the Panmictic American Eel (*Anguilla rostrata*). *Genetics*
596 190:725-U703. 10.1534/genetics.111.134825
597 Griffin JN, Schrack EC, Lewis K-A, Baums IB, Soomdat N, and Silliman BR. 2015. Density-dependent
598 effects on initial growth of a branching coral under restoration *Restoration Ecology* 23:197-200.
599 10.1111/rec.12173
600 Griffin S, Spathias H, Moore DM, Baums IB, and Griffin BA. 2012. Scaling up *Acropora* nurseries in the
601 Caribbean and improving techniques. In: Yellowlees D, and Hughes TP, eds. *Proceedings of the*
602 *12th International Coral Reef Symposium*. Townsville, Australia: James Cook University.
603 Hanley PJ, Xu Y, Kronlage M, Grobe K, Schön P, Song J, Sorokin L, Schwab A, and Bähler M. 2010.
604 Motorized RhoGAP myosin IXb (Myo9b) controls cell shape and motility. *Proceedings of the*
605 *National Academy of Sciences* 107:12145-12150. 10.1073/pnas.0911986107
606 Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, and Glenn TC. 2016. RADcap:
607 sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced
608 missing data. *Molecular Ecology Resources* 16:1264-1278. 10.1111/1755-0998.12566
609 Howells EJ, Abrego D, Meyer E, Kirk NL, and Burt JA. 2016a. Host adaptation and unexpected symbiont
610 partners enable reef-building corals to tolerate extreme temperatures. *Global Change Biology*
611 22:2702-2714. 10.1111/gcb.13250
612 Howells EJ, Abrego D, Meyer E, Kirk NL, and Burt JA. 2016b. Host adaptation and unexpected symbiont
613 partners enable reef-building corals to tolerate extreme temperatures. *Global change biology*.
614 Hubisz M, Falush D, Stephens M, and Pritchard J. 2009. Inferring weak population structure with the
615 assistance of sample group information. *Molecular Ecology Resources* 9:1322-1332.
616 Jeffreys H. 1961. *The theory of probability*. Oxford: Oxford University Press.
617 Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, and Mayrose I. 2015a. CLUMPAK: a program for
618 identifying clustering modes and packaging population structure inferences across K. *Molecular*
619 *ecology resources*.
620 Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, and Mayrose I. 2015b. Clumpak: a program for
621 identifying clustering modes and packaging population structure inferences across K. *Molecular*
622 *Ecology Resources* 15:1179-1191.
623 Krishnapati L-S, and Ghaskadbi S. 2014. Identification and characterization of VEGF and FGF from Hydra.
624 *International Journal of Developmental Biology* 57:897-906.
625 Kulathinal RJ, Stevison LS, and Noor MA. 2009. The genomics of speciation in Drosophila: diversity,
626 divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet*
627 5:e1000550.

628 Lange E, Bertrand S, Holz O, Rebscher N, and Hassel M. 2014. Dynamic expression of a Hydra FGF at
629 boundaries and termini. *Development genes and evolution* 224:235-244.
630 Langmead B, and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357-
631 U354. Doi 10.1038/Nmeth.1923
632 Liao W, Elfrink K, and Bähler M. 2010. Head of Myosin IX Binds Calmodulin and Moves Processively
633 toward the Plus-end of Actin Filaments. *Journal of Biological Chemistry* 285:24933-24942.
634 10.1074/jbc.M110.101105
635 Lischer HEL, and Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting
636 population genetics and genomics programs. *Bioinformatics* 28:298-299.
637 10.1093/bioinformatics/btr642
638 Macdonald AH, Schleyer M, and Lamb J. 2011. Acropora austera connectivity in the south-western
639 Indian Ocean assessed using nuclear intron sequence data. *Marine Biology* 158:613-621.
640 Manel S, Schwartz MK, Luikart G, and Taberlet P. 2003. Landscape genetics: combining landscape
641 ecology and population genetics. *Trends in ecology & evolution* 18:189-197.
642 McKay SE, Hislop J, Scott D, Bulloch AG, Kaczmarek LK, Carew TJ, and Sossin WS. 2001. Aplysia ror forms
643 clusters on the surface of identified neuroendocrine cells. *Molecular and Cellular Neuroscience*
644 17:821-841.
645 Mège P, Schizas NV, Garcia Reyes J, and Hrbek T. 2014. Genetic seascape of the threatened Caribbean
646 elkhorn coral, *Acropora palmata*, on the Puerto Rico Shelf. *Marine Ecology*:online early.
647 10.1111/maec.12135
648 Morin PA, Martien KK, and Taylor BL. 2009. Assessing statistical power of SNPs for population structure
649 and conservation studies. *Molecular Ecology Resources* 9:66-73. 10.1111/j.1755-
650 0998.2008.02392.x
651 Nosil P, Funk DJ, and Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic
652 divergence. *Molecular Ecology* 18:375-402. DOI 10.1111/j.1365-294X.2008.03946.x
653 Peakall R, and Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for
654 teaching and research. *Molecular Ecology Notes* 6:288-295.
655 Peterson BK, Weber JN, Kay EH, Fisher HS, and Hoekstra HE. 2012a. Double Digest RADseq: An
656 Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model
657 Species. *PLoS ONE* 7. 10.1371/journal.pone.0037135
658 Peterson BK, Weber JN, Kay EH, Fisher HS, and Hoekstra HE. 2012b. Double digest RADseq: an
659 inexpensive method for de novo SNP discovery and genotyping in model and non-model
660 species. *PLoS One* 7:e37135.
661 Porto-Hannes I, Zubillaga AL, Shearer TL, Bastidas C, Salazar C, Coffroth MA, and Szmant AM. 2014.
662 Population structure of the corals *Orbicella faveolata* and *Acropora palmata* in the
663 Mesoamerican Barrier Reef System with comparisons over Caribbean basin-wide spatial scale.
664 *Marine Biology*:1-18. 10.1007/s00227-014-2560-1
665 Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E,
666 and Kapitonov VV. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire
667 and genomic organization. *science* 317:86-94.
668 Quinlan AR, and Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
669 *Bioinformatics* 26:841-842. 10.1093/bioinformatics/btq033
670 Richards ZT, and Oppen MJH. 2012. Rarity and genetic diversity in Indo–Pacific *Acropora* corals. *Ecology*
671 *and Evolution* 2:1867-1888. 10.1002/ece3.304
672 Rinkevich B. 2006. The coral gardening concept and the use of underwater nurseries: lessons learned
673 from silvics and silviculture. *Coral reef restoration handbook*:291-302.
674 Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre
675 N, Loire E, Lourenco JM, Nabholz B, Roux C, Tsagkogeorga G, Weber AAT, Weinert LA, Belkhir K,

Bierne N, Glemin S, and Galtier N. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261-263. 10.1038/nature13685

Ryman N, Palm S, Andre C, Carvalho GR, Dahlgren TG, Jorde PE, Laikre L, Larsson LC, Palme A, and Ruzzante DE. 2006. Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology* 15:2031-2045.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, and Willey DL. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933.

Schopmeyer SA, Lirman D, Bartels E, Byrne J, Gilliam DS, Hunt J, Johnson ME, Larson EA, Maxwell K, Nedimyer K, and Walter C. 2012. In Situ Coral Nurseries Serve as Genetic Repositories for Coral Reef Restoration after an Extreme Cold-Water Event. *Restoration Ecology* 20:696-703. 10.1111/j.1526-100X.2011.00836.x

Selkoe KA, and Toonen RJ. 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series* 436:291-305. 10.3354/meps09238

Shinzato C, Mungpakdee S, Arakaki N, and Satoh N. 2015. Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago. *Scientific Reports* 5:18211. 10.1038/srep18211

Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, Fujie M, Fujiwara M, Koyanagi R, Ikuta T, Fujiyama A, Miller DJ, and Satoh N. 2011. Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476:320-323. 10.1038/nature10249

Stoeckel S, Grange J, Fernandez-Manjarres JF, Bilger I, Frascaria-Lacoste N, and Mariette S. 2006. Heterozygote excess in a self-incompatible and partially clonal forest tree species - *Prunus avium* L. *Molecular Ecology* 15:2109-2118. 10.1111/j.1365-294X.2006.02926.x

Swan KA, Curtis DE, McKusick KB, Voinov AV, Mapa FA, and Cancilla MR. 2002. High-throughput gene mapping in Caenorhabditis elegans. *Genome research* 12:1100-1105.

Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR, and Bird CE. 2013. ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* 1:e203. 10.7717/peerj.203

van Oppen MJH, Oliver JK, Putnam HM, and Gates RD. 2015. Building coral reef resilience through assisted evolution. *Proceedings of the National Academy of Sciences* 112:2307-2313. 10.1073/pnas.1422301112

Vilas A, PÉRez-Figueroa A, and Caballero A. 2012. A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology* 25:1364-1376. 10.1111/j.1420-9101.2012.02526.x

Vollmer SV, and Palumbi SR. 2007. Restricted gene flow in the Caribbean staghorn coral *Acropora cervicomis*: Implications for the recovery of endangered reefs. *Journal of Heredity* 98:40-50.

Wang S, Meyer E, McKay JK, and Matz MV. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods* 9:808-+. 10.1038/nmeth.2023

Williams DE, Miller MW, and Baums IB. 2014. Cryptic changes in the genetic structure of a highly clonal coral population and the relationship with ecological performance. *Coral Reefs* 33:595-606. 10.1007/s00338-014-1157-y

Willing E-M, Dreyer C, and van Oosterhout C. 2012. Estimates of Genetic Differentiation Measured by F(ST) Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PLoS ONE* 7:e42649. 10.1371/journal.pone.0042649

Wilson C, Goberdhan D, and Steller H. 1993. Dror, a potential neurotrophic receptor gene, encodes a Drosophila homolog of the vertebrate Ror family of Trk-related receptor tyrosine kinases. *Proceedings of the National Academy of Sciences* 90:7109-7113.

723 Wirth J, Jensen K, Post P, Bement W, and Mooseker M. 1996. Human myosin-IXb, an unconventional
724 myosin with a chimerin-like rho/rac GTPase-activating protein domain in its tail. *Journal of cell*
725 *science* 109:653-661.

726

727

**Table 1** *Acropora palmata* colonies included in SNP analysis. Samples were obtained from four regions in the Caribbean and 3 – 6 reefs per region. Given are latitude and longitude in decimal degrees.

A)

| *Region* | *Reef* | *Count of Samples* | *Latitude* | *Longitude* |
|---|---|---|---|---|
| *Florida* | Sand Island | 6 | 25.018093 | -80.368472 |
| | French | 8 | 25.03393 | -80.34941 |
| | Little Grecian | 1 | 25.118433 | -80.31715 |
| | Horseshoe | 1 | 25.139467 | -80.29435 |
| | Elbow | 8 | 25.143628 | -80.257927 |
| *Bahamas* | Little Ragged Island | 1 | 22.15375 | -75.687208 |
| | Adelaine Cay | 8 | 22.173372 | -75.703016 |
| | Elkhorn Cay | 2 | 22.328253 | -75.783228 |
| | Johnson Cay | 3 | 22.33312 | -75.77892 |
| | Nairn Cay | 8 | 22.35199 | -75.79612 |
| | Middle Beach | 2 | 23.781199 | -76.10391 |
| *Puerto Rico* | San Cristobal | 8 | 17.56493 | -67.04515 |
| | Rincon | 6 | 18.21007 | -67.15849 |
| | Tres Palmas | 2 | 18.350133 | -67.266333 |
| | La Cordillera | 8 | 18.368522 | -65.571678 |
| *USVI* | Tague Bay | 8 | 17.763867 | 64.613397 |
| | Hawksnest Bay | 8 | 18.347183 | -64.780775 |
| | Johnsons Reef | 8 | 18.361733 | -64.7743 |
| *Grand Total* | | 96 | | |

B)

| ***Region*** | ***Reef*** | ***Count of Samples*** | ***Latitude*** | ***Longitude*** |
|---|---|---|---|---|
| *Florida* | Horseshoe | 1 | 25.1395 | -80.294 |
| | Little Grecian | 1 | 25.1184 | -80.317 |
| | Sand Island | 6 | 25.0179 | -80.369 |
| | Western Sambo | 6 | 24.4799 | -81.719 |
| | Rock Key | 4 | 24.456 | -81.86 |
| | Dry Tortugas | 1 | 24.6209 | -82.868 |
| | Marker 3 | 1 | 25.3733 | -80.16 |
| | Boomerang Reef | 1 | 25.3525 | -80.179 |
| | Carysfort | 4 | 25.2219 | -80.211 |
| *Bahamas* | Great Iguana | 19 | 26.7075 | -77.154 |
| | Middle Beach | 2 | 23.7812 | -76.104 |

| *Region* | *Reef* | *Count of Samples* | *Latitude* | *Longitude* |
|---|---|---|---|---|
| | Charlies Beach | 1 | 23.7808 | -76.104 |
| | Black Bouy | 1 | 23.8022 | -76.146 |
| | Bock Cay | 1 | 23.8075 | -76.16 |
| | Little Darby | 2 | 23.8474 | -76.209 |
| | Rocky Dundas | 1 | 24.2788 | -76.539 |
| | Halls Pond | 2 | 24.3539 | -76.57 |
| | LSI | 3 | 23.7691 | -76.096 |
| | Little Ragged Island | 1 | 22.1538 | -75.687 |
| | Adelaine Cay | 1 | 22.1734 | -75.703 |
| | Johnson Cay | 1 | 22.3331 | -75.779 |
| | Nairn Cay | 4 | 22.352 | -75.796 |
| *Puerto Rico* | San Cristobal | 14 | 17.5649 | -67.045 |
| | Rincon | 24 | 18.2101 | -67.159 |
| | Aurora | 3 | 17.9425 | -66.871 |
| | Paraguera | 1 | 17.997 | -67.052 |
| *USVI* | Hawksnest Bay | 6 | 18.3472 | -64.781 |
| | Johnsons Reef | 12 | 18.3617 | -64.774 |
| | Haulover Bay | 13 | 18.3489 | -64.677 |
| | Buck Island | 14 | 18.2774 | -64.894 |
| | Flat Key | 4 | 18.317 | -64.989 |
| | Hans Lollik | 4 | 18.4019 | -64.906 |
| | Sapphire | 6 | 18.3333 | -64.85 |
| | Botany | 3 | 18.3572 | -65.036 |
| *Belize* | unknown | 3 | | |
| | Bugle Caye | 1 | | |
| | Curlew | 5 | -68.896 | -88.083 |
| | Gladden | 1 | 16.4401 | -88.192 |
| | Glovers Atoll | 3 | | |
| | GSTF1 | 5 | 16.5499 | -88.05 |
| | GSTF12 | 7 | 16.5499 | -88.05 |
| | LarksCaye | 1 | | |
| | Laughing Bird Caye | 4 | 16.4367 | -88.199 |
| | Loggerhead | 2 | | |
| | Sandbores | 3 | 16.7791 | -88.118 |

| *Region* | *Reef* | *Count of Samples* | *Latitude* | *Longitude* |
|---|---|---|---|---|
| | Carrie Bow | 13 | 16.8021 | -88.082 |
| *Curacao* | Blue Bay | 7 | 12.1352 | -68.99 |
| | Boka Patrick | 8 | 12.2873 | -69.043 |
| | Directors Bay | 2 | 12.0664 | -68.8603 |
| | East Point | 4 | 12.0407 | -68.783 |
| | PuntuPicu | 9 | 12.0831 | -68.896 |
| | Red Bay | 2 | 12.1355 | -68.99 |
| | Sea Aquarium | 9 | 12.0838 | -68.896 |
| | Water Factory | 3 | 12.1085 | -68.9528 |

736

737

738

739 **Table 2** RAD-tag sequencing summary table of *Acropora palmata* samples.

740

| *Region* | | *Pool* | *Coral colonies* | *Lane* | *Total Reads* | *Retained Reads after processing* | *Average number of retained sequence reads per sample* | *Standard Deviation* |
|---|---|---|---|---|---|---|---|---|
| *West* | Bahamas | B1 | 8 | 2 | 50,900,230 | 41,199,646 | 5,149,956 | 1,915,875 |
| | | B2 | 8 | 2 | 56,097,984 | 45,237,633 | 5,654,704 | 1,853,265 |
| | | B3 | 8 | 2 | 58,379,852 | 47,706,860 | 5,963,358 | 2,734,261 |
| | Florida | F1 | 8 | 1 | 50,925,548 | 39,750,070 | 4,968,759 | 1,681,820 |
| | | F2 | 8 | 1 | 48,752,776 | 42,036,153 | 5,254,519 | 4,422,737 |
| | | F3 | 8 | 1 | 49,942,322 | 38,611,895 | 4,826,487 | 2,518,097 |
| *East* | Puerto Rico | P1 | 8 | 1 | 43,979,338 | 36,237,997 | 4,529,750 | 4,166,551 |
| | | P2 | 8 | 1 | 55,267,402 | 47,235,081 | 5,904,385 | 4,096,287 |
| | | P3 | 8 | 1 | 47,324,190 | 34,835,445 | 4,354,431 | 3,117,707 |
| | USVI | U1 | 8 | 2 | 40,616,766 | 33,170,324 | 4,146,291 | 2,187,597 |
| | | U2 | 8 | 2 | 43,215,386 | 34,291,498 | 4,286,437 | 1,187,166 |
| | | U3 | 8 | 2 | 45,849,098 | 38,439,719 | 4,804,965 | 1,555,938 |
| | | Grand Total | 96 | | 591,250,892 | 478,752,321 | | |

741
742
743

744 **Table 3** Summary statistics for Read 1 and Read 2 combined. % PL = percent polymorphic loci,
745 Obs Hom = observed homozygosity, Obs Het = observed heterozygosity, StdErr = standard
746 error, Exp = expected. $F_{IS}$ calculations with and without minor allele frequency restrictions.
747 Calculated by STACKS 1.30.

748

| | | **Bahamas** | **Florida** | **Puerto Rico** | **USVI** |
|---|---|---|---|---|---|
| All positions: variant and fixed | Total Sites | 200425 | 200425 | 200425 | 200425 |
| | Variant Sites | 390 | 390 | 390 | 390 |
| | Private Alleles | 2 | 1 | 0 | 2 |
| | % PL | 0.1732 | 0.1497 | 0.1694 | 0.1668 |
| | Fis | 0.00005 | 0 | 0 | 0.00005 |
| | Nucleotide diversity (π) | 0.0004 | 0.0004 | 0.0004 | 0.0004 |
| Variant positions only | Obs Hom | 0.7728 | 0.7874 | 0.7791 | 0.7815 |
| | Std Err | 0.0164 | 0.0164 | 0.0154 | 0.0154 |
| | Obs Het | 0.2273 | 0.2126 | 0.2210 | 0.2186 |
| | Std Err | 0.0164 | 0.0164 | 0.0154 | 0.0154 |
| | Exp Hom | 0.7832 | 0.8050 | 0.7919 | 0.7916 |
| | Exp Het | 0.2169 | 0.1951 | 0.2081 | 0.2085 |
| | Fis | 0.02235 | **-0.0086** | 0.0035 | 0.02065 |
| | Nucleotide diversity (π) | 0.2254 | 0.2034 | 0.2174 | 0.21705 |

749
750

**Table 4** Pairwise $F_{ST}$ calculated from STACKS 1.3. Read 1 and 2 combined (duplicated stacks between reads removed, MAF≥0.05) calculated on the AMOVA corrected (p-value <0.05) $F_{ST}$ measurements. Considered were loci present in all populations.

A)

| | *Bahamas* | *Florida* | *Puerto Rico* | *USVI* |
|---|---|---|---|---|
| *Bahamas* | | | | |
| *Florida* | 0.018 | | | |
| *Puerto Rico* | 0.013 | 0.022 | | |
| *USVI* | 0.018 | 0.022 | 0.009 | |

B)

| | Belize | Florida | Bahamas | Puerto Rico | USVI | Curacao |
|---|---|---|---|---|---|---|
| Belize | | | | | | |
| Florida | 0.0040 | | | | | |
| Bahamas | 0.0115 | 0.0097 | | | | |
| Puerto Rico | 0.0206 | 0.0153 | 0.0063 | | | |
| USVI | 0.0206 | 0.0174 | 0.0098 | 0.0037 | | |
| Curacao | 0.0240 | 0.0138 | 0.0181 | 0.0173 | 0.0208 | |

760

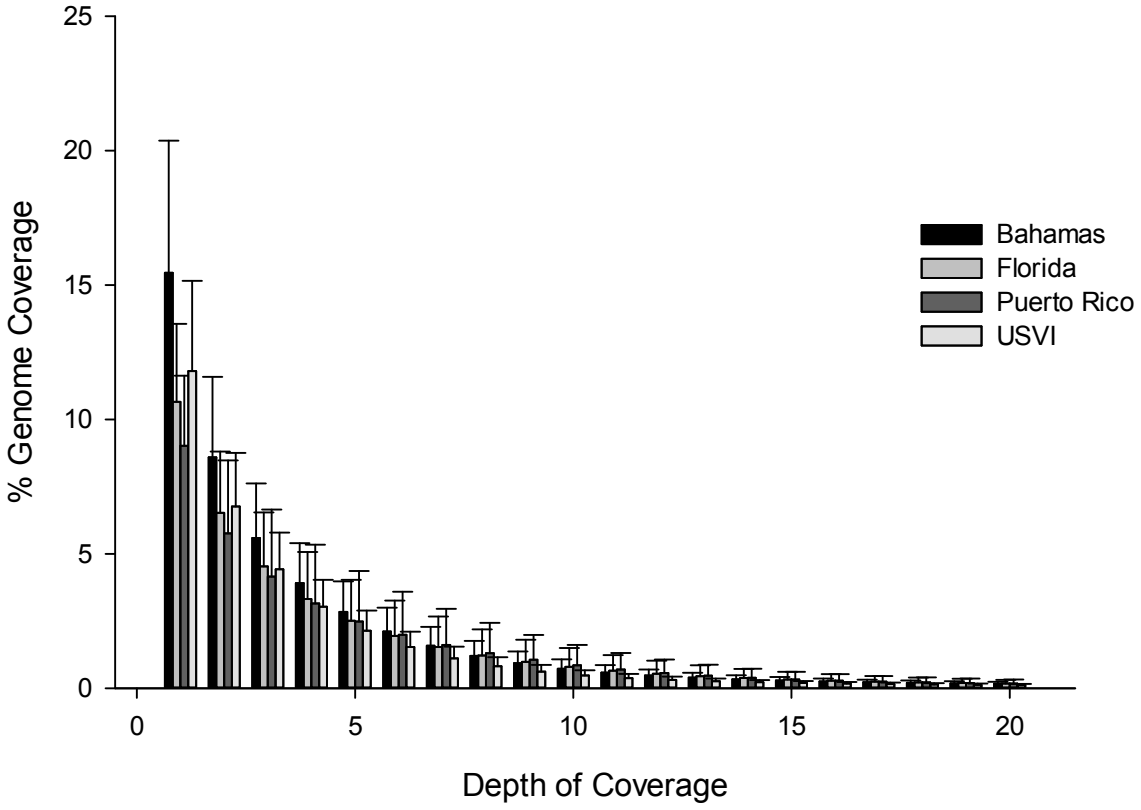


**Figure 1** A Cumulative distribution of sequencing coverage of aligned reads to the *A. digitifera* genome using Bedtools. On average, 2.5% percent of the *A. digitifera* genome had sequence coverage at a stack depth of 5.

761

762

763

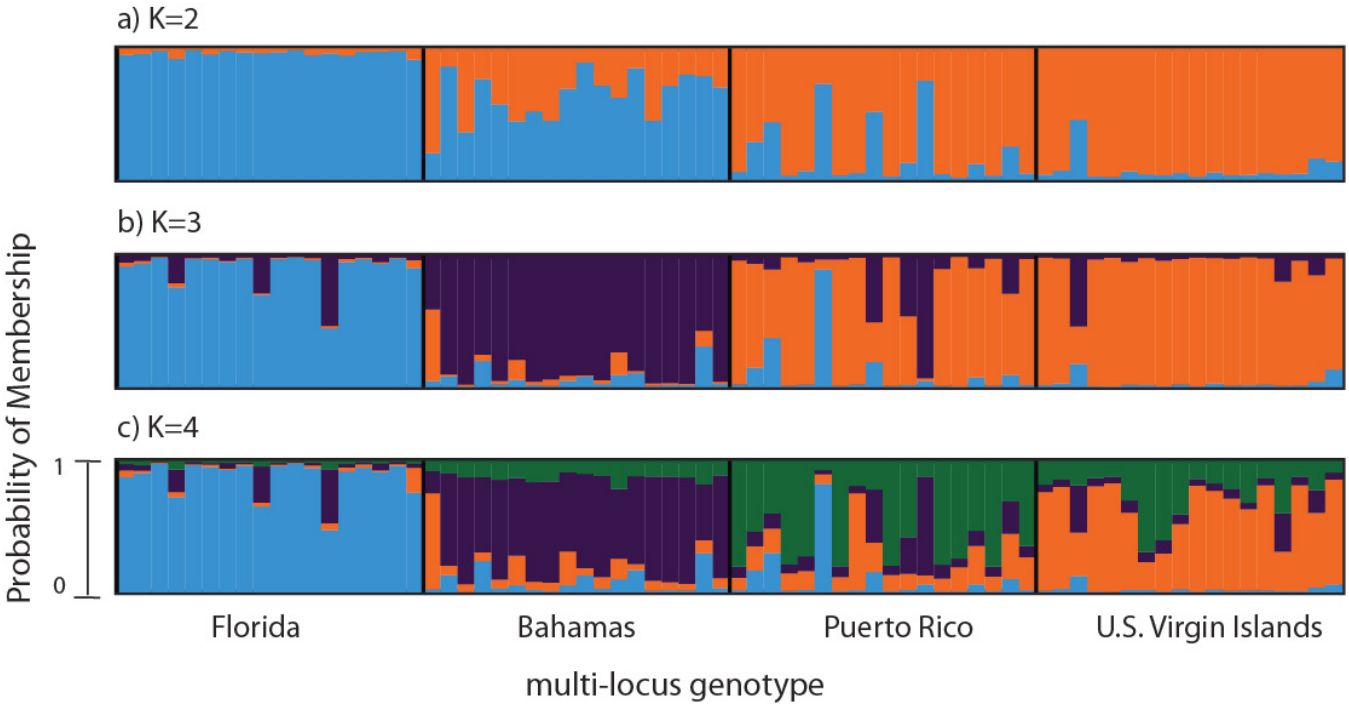


**Figure 2** Bayesian cluster analysis of 307 *Acropora palmata* SNP loci from four regions (Pritchard et al. 2000). Reefs within the regions (Florida, Bahamas, Puerto Rico, US Virgin Islands) are sorted by latitude: Only one SNP per locus were included, after combining Read1 and Read2. Panels show the combined results of 3 replicate run per K. K=2 (a), and K = 3 (b), K=4 (c). The most probable K was 3 (B) for the minor allele frequency corrected SNPs based on the mean estimated log probability of the data at a given K.

764
765
766
767

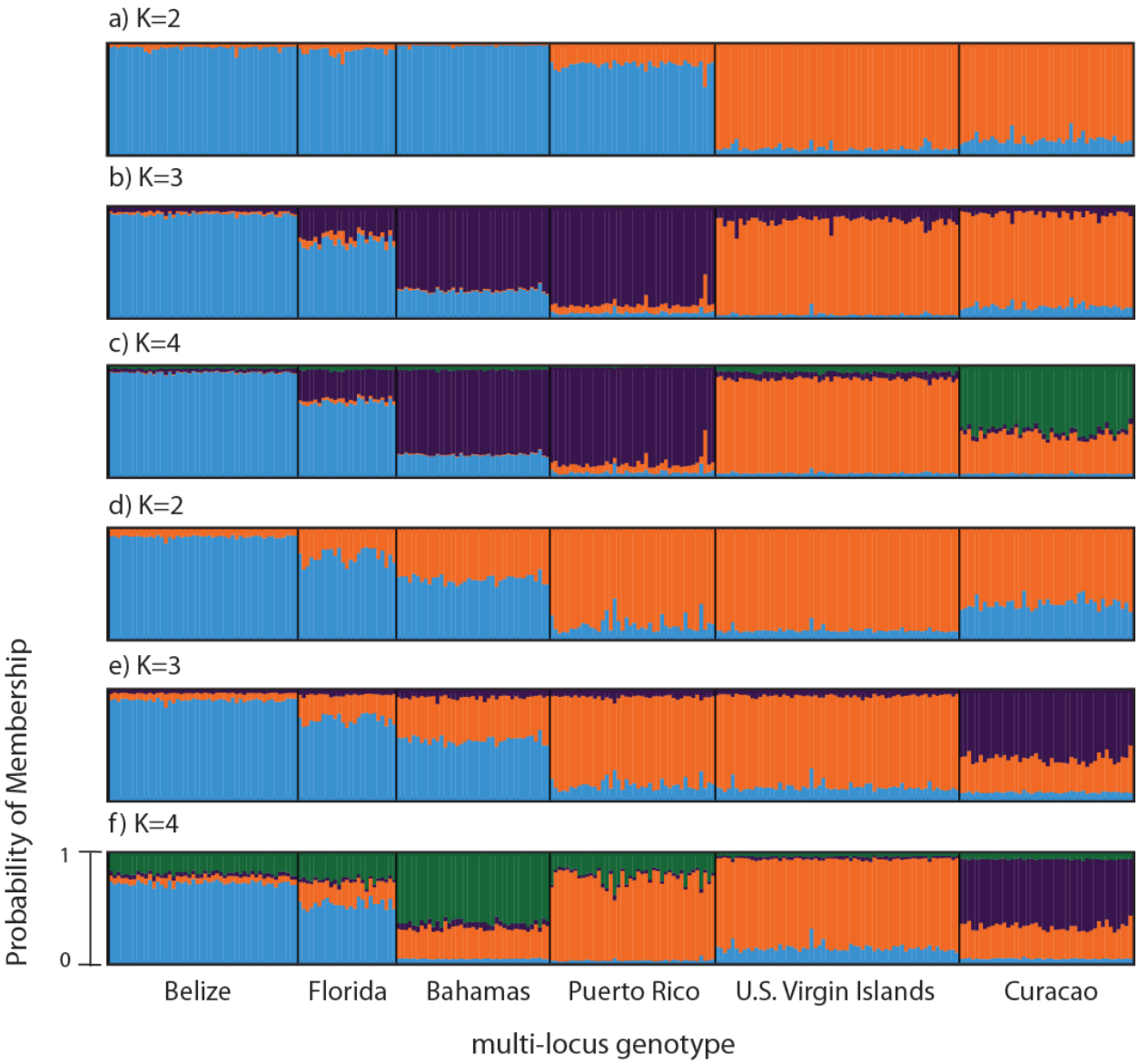


**Figure 3** Bayesian cluster analysis of 11 (a-c) and 10 (d-f) *Acropora palmata* microsatellite loci (Pritchard et al. 2000). Analysis of 11 microsatellites with the most probable K being 4. Panels (D-E). Exclusion of locus 166 that was identified as an outlier resulted in an analysis of 10 microsatellites with the most probable K being 3.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3043v1 | CC BY 4.0 Open Access | rec: 22 Jun 2017, publ: 22 Jun 2017

769

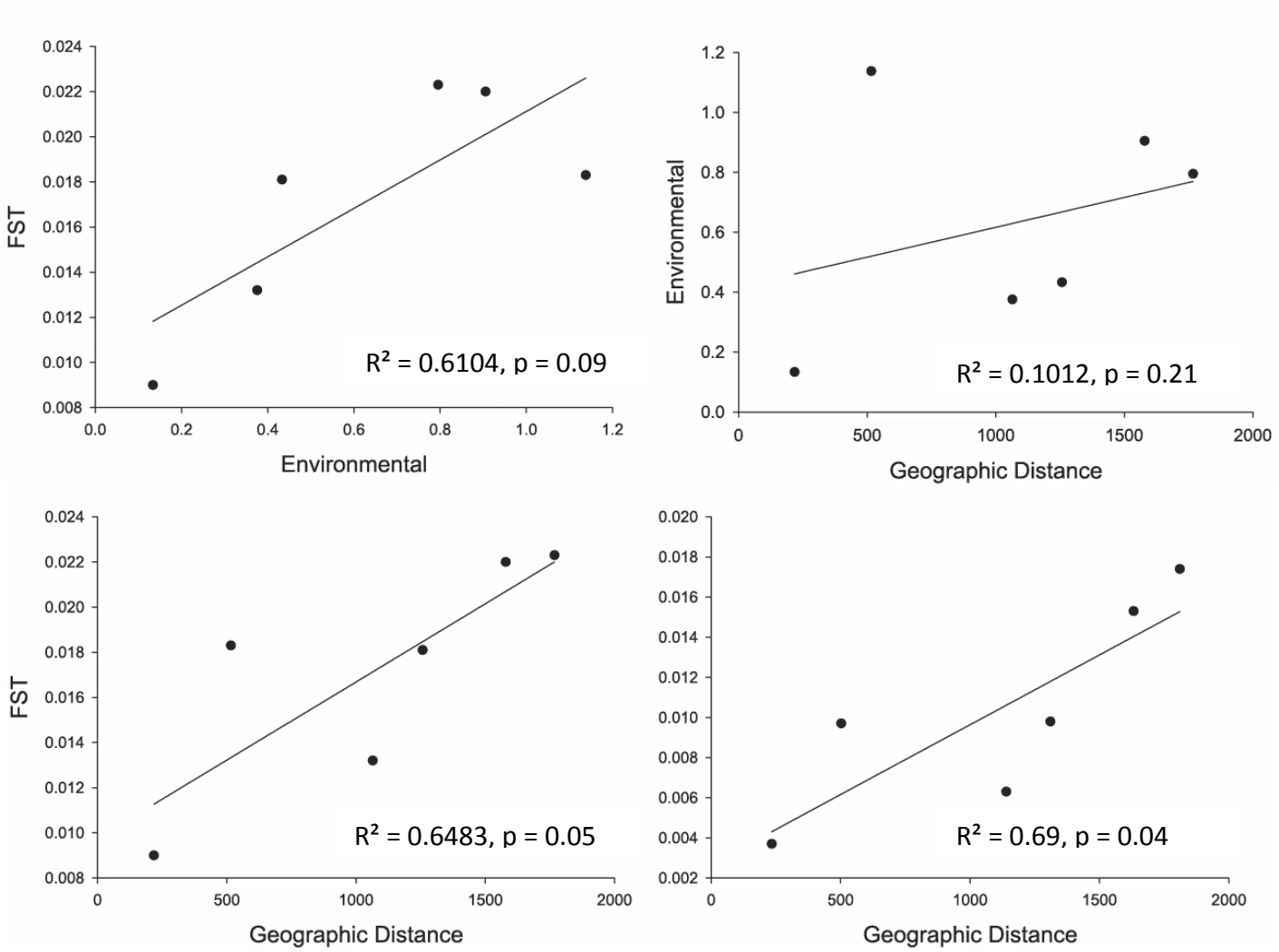


**Figure 4** MANTEL matrix correlation test between genetic and geographic distances, and environmental parameters as calculated by a dissimilarity matrix expressed as the Euclidean distances between regions based on measured environmental data. *Acropora palmata* samples from four regions (Florida, Bahamas, Puerto Rico and USVI) were genotyped with 307 SNP (a, c) or 10 neutral microsatellite markers (d). Panel (a) y = 0.0107x + 0.0104. Panel (b) y=0.002x + 0.4175. Panel (c) y=0.000007x + 0.0098. Panel (d) y=0.000007x + 0.0027.

770
771