SCIENTIFIC REPORTS

OPEN

Effective Dispersal of Caribbean Reef Fish is Smaller than Current Spacing Among Marine Protected Areas

Diana M. Beltrán[1], Nikolaos V. Schizas[1], Richard S. Appeldoorn[1] & Carlos Prada[2]

The oceans are deteriorating at a fast pace. Conservation measures, such as Marine Protected Areas, are being implemented to relieve areas from local stressors and allow populations to restore to natural levels. Successful networks of MPAs operate if the space among MPAs is smaller than the dispersal capacity of the species under protection. We studied connectivity patterns across populations in a series of MPAs in the common yellowhead Jawfish, *Opistognathus aurifrons*. Using the power of genome-wide variation, we estimated that the maximum effective dispersal is 8.3 km. We found that MPAs exchange migrants likely via intermediate unprotected habitats through stepping stone dispersal. At scales >50 km such connectivity is decreased, particularly across the Mona Passage. The MPA network studied would be unable to maintain connectivity of these small benthic fishes if habitat in between them is extirpated. Our study highlights the power of SNPs to derive effective dispersal distance and the ability of SNPs to make inferences from single individuals. Given that overall reef fish diversity is driven by species with life histories similar to that of the yellowhead jawfish, managers face a challenge to develop strategies that allow connectivity and avoid isolation of populations and their possible extinction.

Tropical ecosystems have degraded in the last decades as a result of human activities. This decline is marked on coral reefs, where species commonly seen in the 1970s are rarely observed today, especially apex predators[1]. One conservation initiative to restore marine populations is the designation of networks of Marine Protected Areas (MPAs), which in their most restrictive form are no-take. No–take MPAs are fishing-free spaces that, when properly managed can restore populations if no other stressors are present. These MPAs act in multiple ways to increase the spawning potential of overexploited species, but they also have associated direct benefits to ecosystem as a whole[2]. Thus, while MPAs were originally focused on the restoration of fished species, they are now used to protect biodiversity on reefs[3]. Networks of MPAs are conceptually based on the idea that most marine populations are ecologically connected by the dispersal of planktonic larvae[4]. The spatial scale of larval dispersal determine the degree of connectivity among marine populations, providing information on the ideal reserve size and the minimum spacing among reserves to achieve both self-recruitment and connectivity[5]. Determining the scale of this effective movement of larvae is critical to designing successful MPA networks.

In theory, most marine systems lack obvious barriers to dispersal and MPAs should be connected and work as networks over large spatial scales[4]. Reproduction in marine systems usually results in planktonic larvae that disperse from days to months in the ocean, potentially connecting populations over hundreds of kilometers. Early biophysical models of larval dispersal in the surface ocean have often found marine connectivity across populations[4, 6], and genetic inferences have provided evidence of gene flow over thousands of kilometers[7, 8]. However, recent studies have challenged previous assumptions of high connectivity among marine populations over large spatial scales[9–13].

Recent physical oceanographic models suggest that populations are locally maintained and connectivity among populations restricted[14–16]. Understanding whether populations are connected and the spatial scale at which this occurs is central to designing effective MPA networks. Estimating larval dispersal is challenging, as

[1]Department of Marine Sciences, University of Puerto Rico, Mayagüez, Call Box 9000, Mayagüez, 00681, Puerto Rico, USA. [2]Smithsonian Tropical Research Institute, Box 2072, Balboa, Panama. Correspondence and requests for materials should be addressed to C.P. (email: pradac@si.edu)

common mark-recapture studies do not work efficiently in populations that produce millions of planktonic larvae that quickly dilute in the sea and almost all of which die. Fortunately, genetic surveys provide a powerful indirect estimate of population connectivity[13]. If populations are connected, genetic variation across populations should be minor, but if genetic exchange is restricted and migrants are unable to reproduce in the new population[17], genetic differentiation among populations develops[18]. Now, easy to gather genomic data can capture subtle spatial genetic variation, revealing the effective movement of migrants and their per generation dispersal[19].

Connectivity inferred from genetic data often reflects the product of evolutionary processes across multiple time scales in the history of a species, including those at ecological scales, which are the most useful for designing MPA networks. Fortunately, we can measure connectivity at different time scales by applying different techniques. Gene genealogies within a coalescent framework integrate migration rates over time (e. g. whole history of a population) and provide long-term estimates of migration[20]. Conversely, inferences based on linkage disequilibrium[21] can capture more contemporary connectivity.

We used genome-wide Single Nucleotide Polymorphisms (SNPs) and microsatellites to test whether existing MPAs in the northeast Caribbean are genetically connected through migration among them, and large enough to sustain marine populations of the yellowhead Jawfish (*Opistognathus aurifrons*). In theory, populations of *O. aurifrons* should readily connect thanks to its three-week pelagic larval duration (about 400–500 km dispersal capacity)[15, 22, 23]. We studied genetic variation across a longitudinal gradient of seven MPAs that span the Mona Passage, a well-known biogeographic barrier between Puerto Rico and Dominican Republic[9, 24]. We first inferred genetic diversity and estimated effective population size in the common reef fish, *O. aurifrons*. We then quantified variation across geography and estimated the per generation effective dispersal distance using *in-situ* fish densities. Finally, we tested for differences between contemporary and historical connectivity across these MPAs.

Our data suggest an average effective movement of 10 km per generation. MPAs are connected at scales smaller than 100 km, likely as a result of migrants moving across generations through intermediate unprotected reef areas. For species with life histories similar to that of the yellowhead jawfish, our data indicate that current MPAs are too far apart to maintain connectivity over short time frames, which puts at risk the long-term conservation of the diverse, small benthic fish assemblage as represented by the yellowhead jawfish.

Results

Genetic Diversity.   To measure connectivity, we sampled populations in a network of MPAs along the southern (Caribbean) side of Puerto Rico and the Dominican Republic (Fig. S1). On average, the geographic distance between any two sampled MPAs is 90 km, with the largest being 486 km. A total of 260 individuals were collected from seven MPAs separated by the Mona Passage - two in the Dominican Republic: Parque Submarino La Caleta, and Parque Nacional Natural del Este, and five in Puerto Rico: Mona Island Natural Reserve, Desecheo Island Marine Reserve, Tres Palmas Marine Reserve, La Parguera Natural Reserve, and Canal Luis Peña Marine Reserve at Culebra Island. We genotyped all individuals at 12 newly designed microsatellites[25] (Tables S1 and S2) and a subset of 54 individuals from all seven populations at 10,047 SNPs (Tables S3 and S4).

Genetic variation segregates gradually across geography.   Genetic variation among populations across the MPA network is partitioned by geography (Fig. 1 and Table 1). Geography explains 40% (p < 0.05) of the variation for microsatellites and 42% (p < 0.05) for SNPs. Estimates from single individuals are also remarkably similar (40%; p < 0.05) and support the use of SNPs to sample fewer individuals per population, ameliorating the effects on wild fauna.

As a result of the geographic isolation, pairwise F_{ST}s were highest when comparing populations from West (La Caleta and PNNE, both in the Dominican Republic) and East of the Mona Passage (p < 0.01). The Mona population also showed genetic differentiation when compared to all other populations (p < 0.01), except when compared with Tres Palmas (Table 1).

At finer scales, F_{ST}s from SNPs suggest population differentiation at all levels (except Desecheo vs. Tres Palmas and Mona vs. Tres Palmas). The change in allele frequencies detected by pairwise F_{ST}s suggests high geographic structure even at scales smaller than 100 km. Our comparison of F_{ST}s between microsatellites and SNPs suggests that SNP data provide higher resolution (Table 1). For instance, significance of pairwise F_{ST}s from microsatellites disappeared in six of the pairwise population comparisons, while it was captured by SNPs.

Principal component analysis also shows the geographical partition of genetic variation and the subtle but noticeable genetic break within the Dominican Republic (Fig. 2). It also depicts the overlap in genetic variation at Mona Island from individuals sampled in the Dominican Republic and those sampled in western Puerto Rico. The first two principal components explain 72% and 18% of the genetic variation. Component one separates populations from west to east, while component two segregates populations within Puerto Rico (Fig. 2A). Results from both microsatellites and SNPs are similar. However, the SNP-derived PCA clearly delineates at least two tight groups: La Caleta of Dominican Republic and Puerto Rico (Fig. 2B). The PNNE population, while geographically close to Puerto Rico (Fig. 2B) is quite spread, with high intra-sample variability. The >10 k SNPs have by far the highest resolution, with reduced within-group variation and increased among-group distance.

Effective dispersal is less than ten kilometers.   Density of fish from field surveys ranged from 5,554–74,400 fish per km² [26, 27]. As expected these estimates are biased towards high densities as the surveys were conducted in habitats in which the species is common. In contrast, our estimates from genetic data are 21–282 times lower. The effective population size from 12 microsatellite loci as estimated by IMa2[28] varied from 1,000,290 to 34,318,231 (HPD in IMa2) with a mean of 5,119,320 breeding individuals across the whole Caribbean (26,000 km²), with a density of 263.4 fish per km². We considered a microsatellite mutation rate of 5×10^{-4} per locus per year a value commonly used in the literature[29–31] and one year per generation. Applying these density estimates to Rousset's equation, the effective dispersal distance of the yellowhead jawfish ranges from 0.1

2
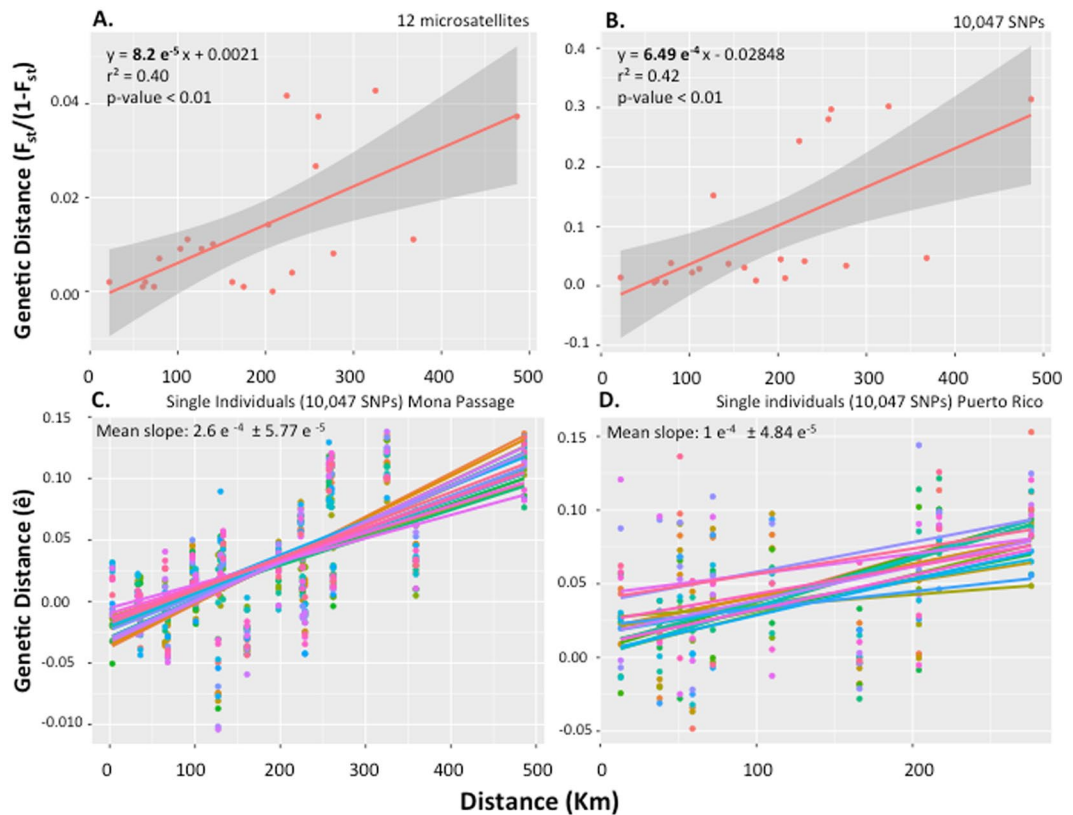
Figure 1. Isolation by Distance (IBD) in *Opistognathus aurifrons* from seven sampling sites along Puerto Rico MPAs (La Caleta, PNNE, Mona Island, Desecheo Island, Tres Palmas MR, La Parguera and Culebra). (**A**) IDB from 12 microsatellites; (**B**) IBD from 10,047 SNPs and (**C**) IBD from single individuals for seven sampling sites (La Caleta, PNNE, Mona Island, Desecheo Island, Tres Palmas MR, La Parguera and Culebra). (**D**) IBD from single individuals for five MPAs within Puerto Rico (Mona Island, Desecheo Island, Tres Palmas MR, La Parguera and Canal Luís Peña at Culebra) with 10,047 SNPs.

to 3.4 km. To avoid the biased increase in slope due to the genetic break between the Dominican Republic and Puerto Rico, we also inferred isolation-by-distance (IBD) patterns only within Puerto Rico (Fig. 1D). If we use the smaller IBD slope obtained from samples only within Puerto Rico and the lowest density, the effective dispersal distance is 8.3 km.

Restricted contemporary gene flow contrasts with historical migration across the Mona Passage.   To further describe the genetically isolated groups, we used linkage disequilibrium across markers. STRUCTURE and Admixture analysis suggests the presence of two genetic pools across the sampled area, represented as orange and blue in Fig. 3 (Fig. S5). The power to assign individuals to two groups differs between microsatellites and SNPs. SNP analysis clearly delineates the two pools. One gene pool is confined to populations in Puerto Rico, including Mona Island, while the other consist of La Caleta, Dominican Republic. The PNNE is an admixed population from La Caleta and Puerto Rico. The pattern is subtler for microsatellites. The orange genotype is more common in the Dominican Republic and is replaced gradually by the blue genotype as one moves eastward to Puerto Rico (Fig. 3). Estimates of the proportion of individuals displaying the orange genotype is significantly different when comparing populations from the Dominican Republic and those from Puerto Rico ($p < 0.001$). Self-recruitment estimates (>0.6) from BayesAss using SNP data also supports limited connectivity among populations. The only significant genetic exchange occurs among populations within Puerto Rico. However, the lack of consistency across datasets does not allow to precisely estimating the magnitude and the geographic center of such migration (Fig. S3 and Table S5).

Sharp contemporary genetic structure contrasts with historical gene flow estimates across the Mona Passage (Fig. 4). We found extensive migration across the Mona Passage. Historical migration seemed higher in the eastern direction from the Dominican Republic to Mona and Puerto Rico. There is also a strong historical gene flow signal from elsewhere in Puerto Rico to Mona, making Mona a historical stepping-stone connecting Puerto Rico and the Dominican Republic.

Discussion

The benefits of creating networks of MPAs are evident when the spacing among protected areas is small enough to allow larvae of neighboring populations to achieve genetic connectivity[5]. It is thus critical to estimate the magnitude and direction of genetic exchange and evaluate their contemporary versus historical component. To

	Tres Palmas	La Parguera	PNNE	Mona Island	La Caleta	Desecheo	Culebra
Tres Palmas	—	0.002	0.002	0.001	**0.026**	0.002	0.005
Parguera	0.009	—	**0.014**	**0.011**	**0.041**	**0.007**	0.001
PNNE	**0.03**	0.043	—	**0.009**	0.009	**0.010**	**0.011**
Mona Island	0.006	**0.028**	0.022	—	**0.040**	0.001	**0.008**
La Caleta	0.219	0.232	0.132	0.196	—	**0.036**	**0.036**
Desecheo	0.014	**0.037**	0.036	**0.006**	0.229	—	0.004
Culebra	**0.013**	**0.009**	**0.045**	**0.033**	0.239	**0.04**	—

Table 1. Spatial genetic structure (F_{ST}) across the Mona Passage as derived from 12 microsatellites (above diagonal) and 10,047 SNPs (below diagonal). Alpha values were adjusted with a Bonferroni correction (0.05/21 = 0.023). Significantly different pairwise comparisons are shown in bold (p < 0.01).

understand contemporary and historical connectivity among populations, we studied genetic variation across a Caribbean network of MPAs in a common benthic reef fish with pelagic larvae. Analyzing variation at 12 micro-satellite loci and over 10k SNPs, we showed that: (i) The yellowhead jawfish has a large amount of genetic varia-tion with an effective population size on the order of millions of individuals. (ii) Effective migration distance per generation is less than 8.3 kilometers generating a pattern of genetic segregation by geography. (iii) Contemporary gene flow is restricted across the Mona Passage, segregating areas into two populations: those west of Mona, and all others. (iv) Contemporary genetic structure across the Mona Passage contrasts the historically high gene flow in both eastern and western directions.

In the broadest sense our genetic analyses suggest populations within the studied MPA network are connected in a stepping-stone fashion via intermediate unprotected reef habitats. Effective dispersal within Puerto Rico of 10 km allows only neighboring populations to exchange migrants, and such limited connectivity results in a posi-tive correlation between geographical and genetic distances as predicted from theory[32]. Across the Mona Passage, the effective dispersal distance was substantially lower: 4 km. Below we present a perspective on the current spacing and connectivity among these already designated MPAs, the critical role of the Mona Island Reserve to preserve connectivity across the Passage and the contrasting patterns of historical and contemporary connectivity across this network of MPAs.

Current MPAs are too far apart to allow connectivity in the absence of intermediate habitats.
Our data suggest a pattern of genetic isolation with geography, with some populations separated only by <100 km showing significant genetic differentiation. Our effective dispersal distance estimates also suggest the fish would move less than 10 km per generation, providing an upper bound for the largest spacing among, MPA networks if constant and full connectivity are necessary design criteria. Our dispersal estimates fall within previous estimates of fish dispersal such as for *Stegastes partitus* (9 km), *Hypoplectrus nigricans* (10 km) and *Hypoplectrus puella* (2–14 km)[33, 34] and the anemone fish *Amphiprion clarkii* (4–20 km)[35], but lower than in *Thalassoma bifasciatum* (27 km), *Haemulon flavolinueatum* (46 km) *Chaetodon capistratus* (52 km)[34].

Dispersal estimates contrast with the mean distance within the network of MPAs sampled (90 km) and all Management Areas (37 km) in the studied area, regardless if they are enforced or not. This is 8X (or 3X if we include all Management Areas) the ideal space among MPAs to ensure connectivity and a high level of protection. Our study suggests that the sustainability of MPAs separated by >20 km may exist only because of the existence of intermediate unprotected habitats. Continued decline in reef health, however, may result in the disappearance of those unprotected habitats, resulting in fragmented isolated populations prone to demographic stochasticity and extinctions.

Our results also highlight the power of SNP data to derive estimates from single individuals. Slopes calculated from population level data and single individuals were almost identical. Finding similar genetic patterns from fewer individuals diminishes harm to wild fauna in genetic surveys and thus preferable for conservation biology.

Mona Island as a link between the east and the west of the Caribbean.
We found a subtle genetic break west from the Mona Passage, dividing populations within the Dominican Republic, with Mona Island being part of the eastern cluster with unequivocal western genetic traces in some individuals. Historically, however, gene flow has occurred between the two populations, and Mona Island is a critical link that has allowed long-term migration in both east and west directions (Fig. 4). Our results fit the east-west pattern of differentiation, prob-ably generated by the effect of deep ocean water in creating undesirable conditions, which limit larval migration across the Passage. For instance, in summer currents flow southward along the Dominican Republic and north-ward around Puerto Rico, while in winter surface water (up to 50 m) flows northward and deeper water (below 50 m) flows southward[14, 16]. Such seasonal patterns of the currents are believed to limit larval dispersal across this narrow stretch during certain times of the year[14]. The population differentiation in the yellowhead jawfish high-lights the importance of Mona Passage as shown in genetic studies of *Acropora* corals[14, 24, 36] in a damselfish and a wrasse[37], goby fishes[9, 38], and also with morphometric and color pattern variations in populations of fishes[39, 40]. Our detailed sampling, however, revealed the biogeographic boundary in *O. aurifrons* occurs between La Caleta and PNNE within the Dominican Republic.

Limited gene flow at ecological scales despite historical long-term connectivity. Our historical estimates of genetic variation in the yellowhead jawfish suggest that the current genetic break in the Dominican Republic is a recent barrier and connectivity could be restored as has happened in the past. Variation between contemporary genetic structure and historical gene flow has also been found in the whitesnout anemone-fish *Amphiprion mccullochi* in southeastern Australia[41]. One caveat from this study is the use of the program Migrate-n[42] on microsatellite data for contemporary gene flow. Migrate-n uses the harmonic mean to integrate variation in gene flow over evolutionary times scales and provides an estimate of the long-term migration rate[43]. In contrast, we calculated long-term migration using Treemix 1.12[44] as a measure of historical gene flow and STRUCTURE 2.3.4 and Admixture 1.3[45, 46] to estimate recent genetic structure.

The genetic barrier around the Dominican Republic and Mona is a temporary barrier where over longer time scales the major westward current does not impede the flux of migrants going eastward. The high gene flow across evolutionary scales is not surprising because the yellowhead jawfish has a pelagic larval duration of more than two weeks[22], which could potentially allow the fish larvae to travel 400–500 km in a single generation[15]. While contemporary gene flow is of most use for conservation, our results stress the role of historical migration in shaping contemporary patterns. In the context of MPA theory, this suggests that present-day barriers to gene flow may hide the historical potential of dispersion of the species. Critically for MPA design, the PNNE and Mona Island represent mosaics of the genetic variation from the east and the west and have historically function as stepping-stones to connect those populations. Preserving Mona is a large step forward towards keeping a condensed amount of the Caribbean genetic diversity and the likelihood of connectivity between the east and the west, if oceanographic conditions allow and conditions in the existing MPAs do not degrade further.

Our results suggest MPAs along the Mona Passage only exchange migrants only through intermediate populations in unprotected habitats. Our findings stress the importance of creating intermediate MPAs to generate a network that ensures connectivity. Our contrasting pattern of contemporary genetic differentiation and historical gene flow highlights the importance of preserving populations, such as the Mona population, at the edges of current genetic breaks. Mona Island is a reservoir of genetic variation from both the eastern and western population and a critical link to allow pulses of gene flow over longer time scale.

Conclusions

Species diversity is driven by small benthic taxa such as the yellowhead jawfish (Fig. 5). Small benthic fish dominant on coral reef systems have an effective dispersal distance less than ten kilometers. MPAs are now widely used to preserve fish diversity across the globe, particularly across coral reef habitats. To use MPAs to protect fish biodiversity, and maintain the health of coral reefs, MPAs need to be closer.

Materials and Methods

Studied Species. To estimate effective dispersal among MPAs, we chose the benthic yellowhead jawfish (*Opistognathus aurifrons*), a common Caribbean species[47]. The yellowhead jawfish is a sedentary benthic fish that lives in decorated burrows on calcareous sand at depths between 3–50 m. Spawning occurs in the burrow and the male orally incubates the eggs[47, 48]. Density estimates for the species are known and available in Puerto Rico[26] and Panama[27]. Given our goal of capturing subtle genetic changes in a fish with pelagic larvae across a densely sampled geographical area, the benthic jawfish is ideal. Small egg-brooding fishes such as the yellowhead jawfish represent a large portion of the fish biodiversity on coral reefs. MPAs while originally developed to restore fished species, are now used to protect biodiversity on reefs[3]. All experimental procedures followed the regulations of the government of Puerto Rico. Our experimental procedure was approved by the University of Puerto Rico Animal Care and Use Committee, the Puerto Rico Department of Natural and Environmental Resources (permit number: 2011-IC-007) and *Secretaría de Estado y Medio Ambiente y Recursos Naturales* at Dominican Republic (permit number: 0611). In all instances we followed their guidelines and recommendations.

Molecular Markers and Genotyping. To generate genotypic data, we extracted genomic DNA across all 260 individuals using the QIAGEN DNeasy Kit following the manufacturer's protocol. We genotyped all individuals at 12 newly designed microsatellites[49] (Table S1) and a subset of 54 individuals at 10,047 SNPs. We included representatives from all seven populations. Both microsatellite and SNP datasets are available through at Dryad (doi:10.5061/dryad.8gg2p). All specimens were vouchered at the Marine Sciences Fish Collection of the University of Puerto Rico and are available upon request.

To genotype individuals at microsatellite loci, we used our described two-step protocol[49]. All PCR amplicons were visualized using an automated sequencer (ABI 3130xl). An internal size standard (ROX-400) was used to estimate microsatellite sizes. We analyzed chromatograms with size standards and scored fragments in Geneious R8 8.1.4[50]. For spurious scoring peaks, we repeated the PCRs and ran them individually on the sequencer as single samples. Individuals were repeated at least five times before deeming them as failed.

To evaluate deviations from Hardy–Weinberg equilibrium in microsatellite data, we used MICROCHECKER 2.2.3[51]. All loci were under Hardy-Weinberg equilibrium, except loci 1588 and 7983. Genetic differentiation and gene flow were estimated including and excluding these two loci, and results were similar. Analyses for all loci are presented.

In addition, we obtained SNP data across all seven populations from a subset of 54 individuals using Genotyping-by-Sequencing (GBS)[52]. The 54 individuals are also a subset of 95 individuals from a larger Caribbean wide study. GBS was performed at the Cornell Institute of Genomic Diversity. GBS is a well-standardized approach to collect massive amounts of loci. In brief, DNA samples are barcoded and adapter pairs ligated. Restriction enzymes are then used to fractionate the genomic DNA. An optimization procedure determined that the *PstI* (recognition site: CTGCAG) was the best for the yellowhead jawfish. Once fragments are generated, adapters are ligated to the end of the fragments. Samples are then pooled and size selected to exclude unreacted
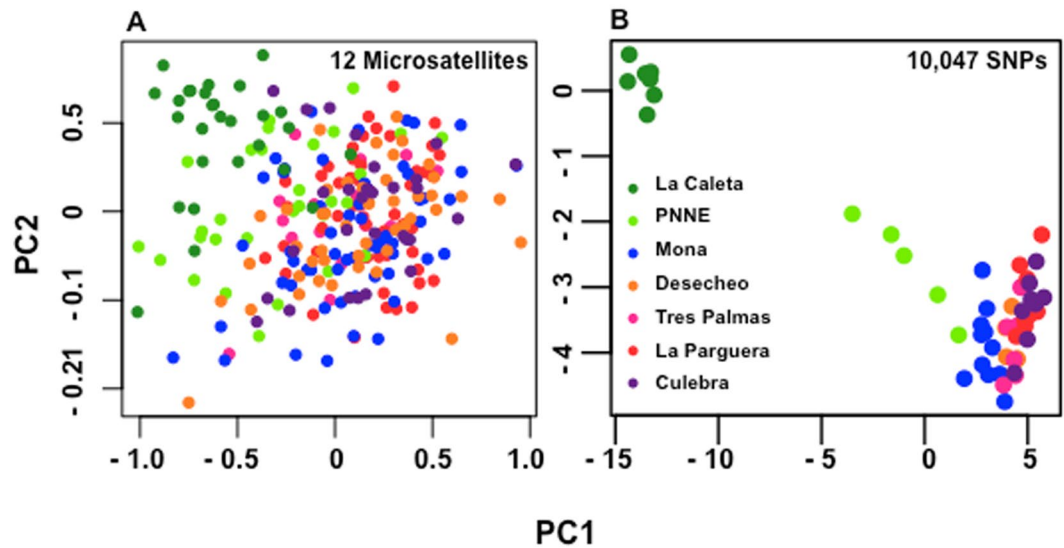
Figure 2. Multidimensional summary of genetic variation. (**A**) Principal component analysis (PCA) of the individuals genotypes with 12 microsatellites. (**B**) Principal component analysis (PCA) of the individuals genotypes with 10,047 SNPs.
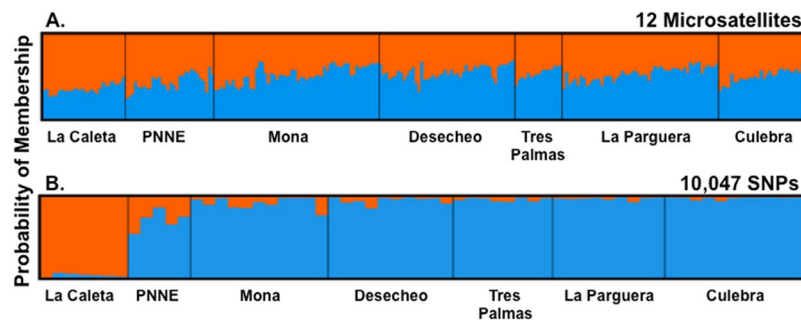

Figure 3. Graphical summary of clustering Analysis for *Opistognathus aurifrons* genotypes from seven Marine Protected Areas along Mona Passage using Structure. Probability of Membership of each sampling site (n = 7) from individuals per site scored with 12 microsatellites (**A**) and 10,047 SNPs (**B**). Each vertical line represents an individual and the estimated proportion of the individual's genome from each inferred cluster.

adapters. Libraries were then amplified using an 18-cycle PCR with long primers complementary to the barcode and common adapters[52]. Samples were run on a 100-bp single-end Illumina HiSeq 2000 lane.

Raw reads were processed using UNEAK, an addition to TASSEL 3.0[53]. UNEAK retains all reads with a barcode, cut site, and no missing data in the first 64 bp after the barcode. Reads are then grouped into tags by 100% identity, tags are aligned pairwise, and any tag pairs differing by one bp are identified as potential SNPs. We identified 1,528,449 total tags after merging, with 639,274 total tag networks identified. To remove sequencing errors, and individuals and SNPs with a high amount of missing data, any alleles represented by a frequency of less than 3% were filtered out.

To further remove data artifacts, we filtered out SNPs that deviated from Hardy–Weinberg equilibrium ($p > 0.01$) in at least five populations. We initially had over 58,756 SNPs; after removing individuals with <10% of the markers and SNPs with <5% frequency we retained 18,616 SNPs. We also excluded SNPs suspected under selection as identified by BayesScan 2.01[54] (12 loci) and consider only SNP per tag marker. To overcome any bias associated with having missing SNPs, we first generated a file with no missing information. This file has 90 individuals genotyped at 525 SNPs. In the second dataset we have individuals that have at least 90% of the SNPs and SNPs present in at least 70% of the samples. This last dataset has 54 individuals and 10,047 SNPs.

Genetic Diversity, Isolation by Distance and Effective Dispersal. We quantified genetic diversity across markers in GENODIVE 2.0[55] and estimated heterozygosity and number of alleles per population across all 12 microsatellites and 10,047 SNPs. We also estimated genetic variation across geography by calculating pairwise F_{ST}s among locations. We then estimated pairwise linear distance between locations in Google Earth 6.2 using the shortest nautical distance among populations. We then plotted genetic and geographical distances and estimated isolation-by-distance (IBD), assuming a linear regression in R[56].
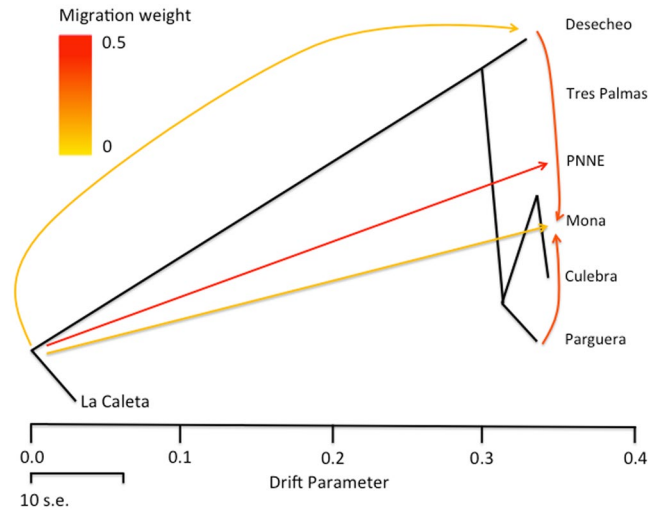
Figure 4. Phylogram of the inferred *Opistognathus aurifrons* population relationships. Scale bar shows 10 times the average standard error of the estimated entries in the sample covariance matrix. Arrows depict migration patterns. Red arrows represent stronger migration.
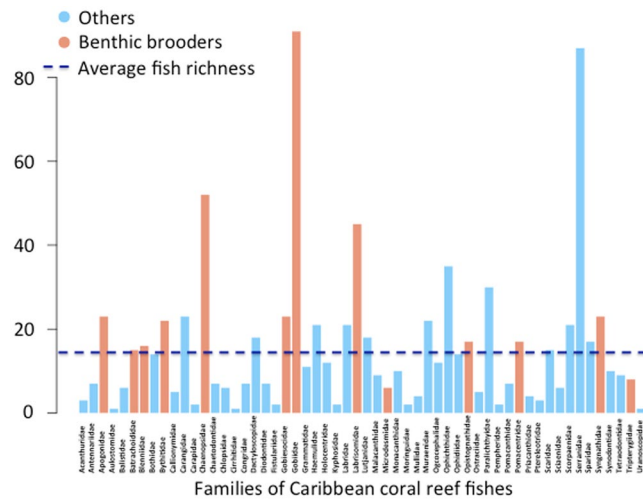
Figure 5. Number of species per Caribbean reef fish families.

An added feature of SNP data is that one can sample thousands of semi-independent markers and estimate more precisely the average genome-wide genetic distance between individuals. This improved power allows estimating IBD patterns using single individuals per population. To test whether single individuals would give similar results than patterns derived from populations, we randomly selected an individual from each population and iterated the process twenty times. We used the ê distance[57] to measure differences between pairs of individuals.

To estimate effective dispersal, we used Rosset's (1997) approach:

$$\sigma = \sqrt{\frac{1}{4D_e m}}$$

(1)

where D_e is the effective density and m is the slope derived from our analysis of IBD. To estimate D_e we used both a direct density estimate of fish per kilometer and one derived from genetic data by estimating effective population size.

Field densities were compiled from studies in Panama[27], and two locations in Puerto Rico: Rincón and La Parguera[26]. We estimated differences in habitat areas by measuring habitat patch sizes from GIS maps[58]. These estimates are likely biased upwardly and represent the upper bound of the highest fish density.

To measure effective dispersal from genetic data, we used effective population size as implemented in IMa2 ($\Theta = 4 N_e \mu$; where Θ is genetic variation in the population; N_e is the effective population size and μ is the mutation rate). We ran IMa2 (version 8/27/12)[28] on 12 microsatellites using as populations La Caleta and Puerto Rico (all populations except La Caleta and PNNE). We run IMa2 for 3,000,000 steps after an initial burnin of 1,000,000

steps. We repeated the process three times and check for convergence on similar results. All ESS values for all runs were above 50. We estimated effective population size as the mean across runs. We used a mutation rate of 5×10^{-4} [29–31] and one year generation time to estimate population sizes in demographic units. Once we had an estimate of the effective population size, we obtained fish density per km^2 by dividing by the total coral reef area across the Caribbean (26,000 km^2)[59]. Our density from genetic data is biased towards the lowest possible density, as long-term effective sizes are lower compared to census size (e.g., for instance in humans N_e is only 10,000 and census is >7 billion;[60]. In addition, we divided N_e by the reef area across the Caribbean, even though the yellowhead jawfish is a habitat specialist and occurs only on colonized pavement with sand channels on coral reefs, which in Puerto Rico make <30% of the reef area[26]. By having the direct census size as the upper bound of the highest density and the genetic inference with the largest possible distributional area, as the lowest possible density, we provide the widest range of effective dispersal estimates.

Genetic Differentiation and Contemporary Connectivity. To test for genetic differentiation, we used linkage disequilibrium among markers via Bayesian clustering[45]. We ran STRUCTURE 2.3.2 on microsatellites and SNPs assuming an admixture model with a burn-in of 900,000 steps followed by 10 million iterations and 3 replicates per run. We assumed a number of inferred populations from 1 to 7 and then used the Evanno method[61] to infer the optimal number of populations (K's) in our dataset. In addition, we run Admixture 1.3[46] with cross-validation, and 1000 bootstrapping replicates for K values 1, 2, 3, 4 and 5 (Fig. S4). To further understand the distribution of genetic variation across geography, we calculated principal components in GENODIVE 2.0[55] and principal coordinates in Genalex 6.5[62, 63]. PCA analysis of both SNPs and microsatellites are shown in the main text (Fig. 2) and principal coordinates in the supplement (Fig. S5).

To test for contemporary gene flow, we used BayesAss 3[64]. To analyze our data in BayesAss, we used all seven populations sampled. For each analysis we ran BayesAss for a burn-in of 9,000,000 steps and inferred parameters using the next 90,000,000–300,000,000 iterations sampled at 10,000 intervals. BayesAss is constrained on using <400 loci so we generated 18 random subsets of our data and run each individually. The program was run using different random seed values. Mixing and convergence were inspected using TRACER 1.6[65]. Given the lack of convergence across runs, we also run the program three times with smaller datasets consisting of only populations from Puerto Rico.

Historical Connectivity. To quantify historical connectivity across the Mona Passage, we used Treemix 1.12[44] which can handle SNP data without ascertainment bias. To analyze the data in Treemix, the same population partitions were used as in BayesAss. To infer migration parameters, we used La Caleta, DR (western-most population) as the outgroup, as a larger study across the Caribbean suggested it to be less related to all other populations[25]. To increase our confidence in the Treemix estimates we ran the analysis with 500 bootstrap replicates and drew trees in R.

Both contemporary and historical levels of migration inferred from BayesAss[64] and Treemix[66] do come with some limitations and assumptions. Among others, that the data are presumably derived from "neutrally" evolving markers and there is independence (not in significant LD) among the markers. We have excluded loci suspected to be under selection as inferred from Bayescan 2.01[54] (12 outlier loci) and only consider one SNP per tag marker.

References

1. Steneck, R. S. *et al.* Thinking and managing outside the box: coalescing connectivity networks to build region-wide resilience in coral reef ecosystems. *Coral Reefs* **28**, 367–378 (2009).
2. Sale, P. F. *et al.* Critical science gaps impede use of no-take fishery reserves. *Trends in Ecology & Evolution* **20**, 74–80 (2005).
3. Day, J. *et al.* Guidelines for applying the IUCN Protected Area Management Categories to Marine Protected Areas. 36 (2012).
4. Roberts, C. M. Connectivity and Management of Caribbean coral reefs. *Science* **278**, 1454–1457 (1997).
5. Sale, P. F. *et al.* Critical science gaps impide use of no-take fishery reserves. *Trends in Ecology & Evolution* **20**, 74–80 (2005).
6. Heck, K. L. & McCoy, E. D. Long distance dispersal and the reef building corals on the Eastern Pacific. *Marine Biology* **48**, 349–356 (1978).
7. Lessios, H. A. & Robertson, D. R. Crossing the impassable: genetic connections in 20 reef fishes across the Eastern Pacific Barrier. *Proceedings of the Royal Society*, *London series* B, **273**, 2201–2208 (2006).
8. McFadden, C. S., Grosberg, R. K., Cameron, B. B., Karlton, D. P. & Secord, D. Genetic relationships within and between solitary and clonal forms of the sea anemone Anthopleura elegantissima revisited: Evidence for the existence of two species. *Marine Biology* **128**, 127–139 (1997).
9. Taylor, M. S. & Hellberg, M. E. Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science* **299**, 107–109 (2003).
10. Underwood, J. N., Smith, L. D., Van Oppen, M. J. H. & Gilmour, J. P. Multiple scales of genetic connectivity in a brooding coral on isoled reefs following catastrophic bleaching. *Molecular Ecology* **16**, 771–784 (2007).
11. Warner, R. R. Evolutionary ecology: how to reconcile pelagic dispersal with local adaptation. *Coral Reefs* **16**, 115–128 (1997).
12. Prada, C. & Hellberg, M. E. Long pre-reproductive selection and divergence by depth in a Caribbean candelabrum coral. *Proceedings of the National Academy of Sciences USA* **119**, 53–60 (2013).
13. Jones, G. P., Milicich, M. J., Emslie, M. J. & Lunow, C. Self-recruitment in a coral reef fish population. *Nature* **402**, 802–804 (1999).
14. Baums, I. B. & Paris, C. B. A bio-oceanographic filter to larval dispersal in a reef-building coral. *Limnology and Ocenography* **51**, 1969–1981 (2006).
15. Cowen, R. K., Paris, C. B. & Srinivasan, A. Scaling of connectivity in marine populations. *Science* **311**, 522–527 (2006).
16. Galindo, H. M. *et al.* Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal. *Molecular Ecology* **19**, 3692–3707 (2010).
17. Prada, C. & Hellberg, M. E. Strong natural selection on juveniles maintains a narrow adult hybrid zone in a broadcast spawner. *American Naturalist* **184**, 702–713 (2014).
18. Slatkin, M. Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792 (1987).
19. Rousset, F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219–1228 (1997).
20. Pinho, C. & Hey, J. Divergence with Gene Flow: Models and Data. *Annual Review of Ecology, Evolution, and Systematics* **41**, 215–230 (2010).

21. Rosenberg, N. A. *et al.* Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* **159**, 699–713 (2001).
22. Young, F. A. The yellowhead jawfish: breeding the marine mouthbrooder in captivity. *Freshwater and Marine Aquarium* **4**, 50–51 (1982).
23. Ho, A. L. F. C., Pruett, C. L. & Lin, J. Population genetic structure, coloration, and morphometrics of yellowhead jawfish *Opistognathus aurifrons* (Perciformes: Opistognathidae) in the Caribbean region. *Marine Ecology Progress Series* **444**, 275–287 (2012).
24. Baums, I. B., Miller, M. W. & Hellberg, M. E. Regionally isolated population of an imperiled Caribbean coral. *Acropora palmata. Molecular Ecology* **14**, 1377–1390 (2005).
25. Beltrán, D. *The Scale of Connectivity in Benthic Reef Fishes: The Population Genomics of Opistognathus aurifrons* DOCTOR OF PHILOSOPHY (Biological Oceanography) thesis, University of Puerto Rico, Mayagüez (2015).
26. LeGore, R. S., Hardin, M. P. & Garĉia-Sais, J. R. Puerto Rico Marine Ornamental Fishery Evaluation Phase II: Wild Population Assessments 1–63 (2006).
27. Hess, H. C. Male mouthbrooding in jawfishes (Opistognathidae): constraints on polygyny. *Bulletin of Marine Science* **52**, 806–818 (1993).
28. Hey, J. Isolation with Migration Models for More Than Two Populations. *Molecular Biology and Evolution* **27**, 905–920 (2010).
29. Ellegren, H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551–558 (2000).
30. Sun, J. X. *et al.* A direct characterization of human mutation based on microsatellites. *Nature Genetics* **44**, 1161–1165 (2012).
31. Yue, G. H., David, L. & Orban, L. Mutation rate and pattern of microsatellites in common carp (*Cyprinus carpio L.*). *Genetica* **129**, 329–331 (2007).
32. Slatkin, M. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264–279 (1993).
33. Puebla, O., Bermingham, E. & Guichard, F. Estimating dispersal from genetic isolation by distance in a coral reef fish (*Hypoplectrus puella*). *Ecology* **90**, 3087–3098 (2009).
34. Puebla, O., Bermingham, E. & Mcmillan, W. O. On the spatial scale of dispersal in coral reef fishes. *Molecular Ecology* **21**, 5675–5688 (2012).
35. Pinsky, M. L., Montes, H. R. & Palumbi, S. R. Using Isolation by Distance and effective Density to Estimate Dispersal Scales in Anemonefish. *Evolution* **64**, 2688–2700 (2010).
36. Vollmer, S. V. & Palumbi, S. R. Restricted gene flow in the Caribbean staghorn coral Acropora cervicornis: Implications for the recovery of endangered reefs. *Journal of Heredity* **98**, 40–50 (2007).
37. Shulman, M. & Bermingham, E. Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* **49**, 897–910 (1995).
38. Taylor, M. S. & Hellberg, M. E. Comparative phylogeography in a genus of coral reef fishes: biogeographic and genetic concordance in the Caribbean. *Molecular Ecology* **15**, 695–707 (2006).
39. Colin, P. L. *Neon Gobies.*, (T.F.H. Publications, Inc, Neptune City, New Jersey 1975).
40. Dennis, G. D., Smith-Vaniz, W. F., Colin, P. L., Hensley, D. A. & McGehee, M. A. Shore Fishes from Islands of the Mona Passage, Greater Antilles with Comments on Their Zoogeography. *Caribbean Journal of Science* **41**, 716–743 (2005).
41. Van der Meer, M. H., Hobbs, J.-P. A., Jones, G. P. & van Herwerden, L. Genetic Connectivity among and Self-Replenishment within Island Populations of a Restricted Range Subtropical Reef Fish. *PLOS ONE* **7**, 11 (2012).
42. Beerli, P. Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341–345 (2006).
43. Beerli, P. & Palczewski Unified Framework to Evaluate Panmixia and Migration Direction Among Multiple Sampling Locations. *Genetics* **185**, 313–326 (2010).
44. Pickrell, J. K. & Pritchard, J. K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *Plos Genetics* **8**, e1002967 (2012).
45. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945–959 (2000).
46. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–1664 (2009).
47. Colin, P. L. Interspecific relationships of the yellowhead jawfish, *Opistognathus aurifrons* (Pisces, Opistognathidae). *Copeia* **1971**, 469–473 (1971).
48. Colin, P. L. Burrowing behaviour of the yellowhead jawfish. *Opistognathus aurifrons. Copeia*, 84–90 (1973).
49. Beltrán, D., Schizas, N. V., Prada, C. & Appeldoorn, R. S. Isolation and characterization of 12 microsatellite loci to study connectivity in the yellow jawfish *Opistognathus aurifrons. Conservation Genetic Resources* **7**, 543–545 (2014).
50. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
51. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535–538 (2004).
52. Elshire, R. J. *et al.* A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One* **6**, 1–10 (2011).
53. Bradbury, P. J. *et al.* TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
54. Fischer, M. C., Foll, M., Excoffier, L. & Heckel, G. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology* **20**, 1450–1462 (2011).
55. Meirmans, P. G. & Van Tienderen, P. H. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology* **4**, 792–794 (2004).
56. Team, R. D. C. R: A language and environment for statistical computing R Foundation for Statistical Computing. Vienna, Austria (2013).
57. Watts, P. C. *et al.* Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion merculiale*: Odonata: Zygoptera) populations: analysis of 'neighbourhood size' using a more precise estimator. *Molecular Ecology* **16**, 737–751 (2007).
58. NOAA/NOS/NOAA/Biogeography Program. Benthic Habitats of Puerto Rico and the U.S. Virgin Islands; Photomosaic of St. John, USVI (2001).
59. Burke, L. & Maidens, J. *Reefs at Risk in the Caribbean*, Online at: http://www.wri.org/publication/ (2004).
60. Hill, W. G. Genetics Research. *Estimation of effective population size from data on linkage disequilibrium* **38**, 209–216 (1981).
61. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611–2620 (2005).
62. Peakall, R. & Smouse, P. E. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288–295 (2006).
63. Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537–2539 (2012).
64. Wilson, G. A. & Rannala, B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177–1191 (2003).
65. Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. Tracer v1.6. *Available from* http://beast.bio.ed.ac.uk/Tracer (2014).
66. Pickrell, J. K. & Pritchard, J. K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *Plos Genetics* **8**, 1–17 (2012).

Author Contributions

D.B., R.A., N.V.S. and C.P. conceived the study; D.B. and C.P. collected the material; D.B. carried out all lab analysis and D.B. and C.P. performed all data analysis. D.B. and C.P. wrote the paper with input from N.S. and R.A.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-04849-5

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.