

SCIENTIFIC REPORTS

OPEN

Genetic homogeneity of the invasive lionfish across the Northwestern Atlantic and the Gulf of Mexico based on Single Nucleotide Polymorphisms

R. Pérez-Portela^{1,6}, A. Bumford¹, B. Coffman¹, S. Wedelich¹, M. Davenport¹, A. Fogg², M. K. Swenarton³, F. Coleman⁴, M. A. Johnston⁵, D. L. Crawford¹ & M. F. Oleksiak¹

Despite the devastating impact of the lionfish (*Pterois volitans*) invasion on NW Atlantic ecosystems, little genetic information about the invasion process is available. We applied Genotyping by Sequencing techniques to identify 1,220 single nucleotide polymorphic sites (SNPs) from 162 lionfish samples collected between 2013 and 2015 from two areas chronologically identified as the first and last invaded areas in US waters: the east coast of Florida and the Gulf of Mexico. We used population genomic analyses, including phylogenetic reconstruction, Bayesian clustering, genetic distances, Discriminant Analyses of Principal Components, and coalescence simulations for detection of outlier SNPs, to understand genetic trends relevant to the lionfish's long-term persistence. We found no significant differences in genetic structure or diversity between the two areas (F_{ST} p -values > 0.01 , and t -test p -values > 0.05). In fact, our genomic analyses showed genetic homogeneity, with enough gene flow between the east coast of Florida and Gulf of Mexico to erase previous signals of genetic divergence detected between these areas, secondary spreading, and bottlenecks in the Gulf of Mexico. These findings suggest rapid genetic changes over space and time during the invasion, resulting in one panmictic population with no signs of divergence between areas due to local adaptation.

The Indo-Pacific lionfish (*Pterois* spp.) invasion of the northwestern Atlantic (NW Atlantic) is remarkable for its speed and magnitude¹ and considered among the world's more critical conservation issues during the last decade². The lionfish invasion is the first recorded invasion of a marine fish species in United States Atlantic waters³. Lionfish are now the most abundant fish predators in many reefs within the Wider Caribbean (i.e. tropical NW Atlantic, Caribbean Sea and Gulf of Mexico)^{4–6}, including those at depths of 30–350 m, with densities that, in some cases, far exceed those in their native ranges⁷. The success of these invasive predators is likely related to a combination of niche availability in the introduced area, ability to colonize a wide variety of habitats and broad thermohaline tolerance, high fecundity, wide-ranging dispersal, venomous defences, and absence of natural predators in the newly colonized environments⁸. The rapid increase in lionfish abundance in the Wider Caribbean has been linked to direct and indirect impacts on invaded ecosystems^{6,7}, such as significant declines in native fish biomass due to lionfish's predatory success on native species^{7,9} and their purported ability to displace large reef fishes, including groupers^{7,9}. Lionfish are high-efficiency predators that feed primarily on post-settlement

¹Rosenstiel School of Marine & Atmospheric Science (RSMAS), University of Miami, 4600 Rickenbacker Cswy, FL, 33149, Miami, USA. ²Florida Fish and Wildlife Conservation Commission, Division of Marine Fisheries Management, University of Southern Mississippi, Gulf Coast Research Laboratory, 703 East Beach Drive, Ocean Springs, MS, 39564, USA. ³Fish and Wildlife Service 800S Guild Ave Lodi, Lodi, CA, 95240, USA. ⁴Florida State University, Coastal and Marine Laboratory, 3618 Coastal Highway 98, St. Teresa, FL, 32358, USA. ⁵Flower Garden Banks National Marine Sanctuary, 4700 Avenue U, Bldg. 216, Galveston, TX, 77551, USA. ⁶Present address: Department of Biology, Geology, Physics and Inorganic Chemistry, Rey Juan Carlos University, C/Tulipán s/n, Móstoles, 28932, Spain. A. Bumford, B. Coffman and S. Wedelich contributed equally to this work. Correspondence and requests for materials should be addressed to R.P.-P. (email: perezportela@gmail.com)

reef fishes, which they disorient by blowing water jets on them before capture¹⁰. Prey in the invaded range have not developed defence strategies against this predatory mechanism and so succumb easily¹⁰. In addition, female lionfish are capable of spawning year-round, every 2–9 days (season dependent), with an average annual output of more than 2 million eggs and several tens of generations. This tremendous output of buoyant gelatinous egg masses that drift with marine currents, provides opportunities for widespread dispersal over short time-periods¹¹.

The lionfish invasion has been chronologically well documented, and potential routes of invasion and secondary spreading tracked^{3,12,13}. Lionfish, likely introduced via the ornamental pet trade and aquarium releases^{14–16}, was first noticed in the Western Atlantic on the east coast of Florida off Dania Beach in 1985¹⁷. In 1992, six lionfish were reported to have escaped from the Miami Aquarium at Key Biscayne during Hurricane Andrew¹⁸. By 2001, lionfish populations were well established along the US Atlantic coast, from Florida to North Carolina and Bermuda, with sporadic observations reported as far as north Rhode Island, where cold winter temperatures constraint the development of permanent populations³. After 2001, lionfish rapidly spread into the Bahamas and throughout Caribbean Sea, with observations of well-established populations in 2004 and 2007, respectively^{3,19,20}. Following the colonization of the Caribbean, the first apparent arrival of lionfish into the Gulf of Mexico via larval transport was reported in 2009, where they quickly spread and increased in density^{13,16,21}. Unfortunately, lionfish dispersion is not limited to the North Hemisphere as in 2014 recreational divers collected one adult specimen approximately 5,500 km from the Caribbean in a subtropical reef off Brazil's southeast coast²².

During the last years, genetic approaches using mitochondrial DNA have been applied to investigate the lionfish invasion. Barcode analyses suggested that although two lionfish species, *Pterois volitans* (Linnaeus, 1758) and *P. miles* (Bennett, 1828), were introduced in the NW Atlantic^{15,23}, *P. volitans* is the most ubiquitous species, occurring throughout the US east coast, Caribbean Sea, and the Gulf of Mexico^{15,16,20,21,23,24}. Although some molecular data did not detect signs of mitochondrial introgression and/or hybridization between the two potential species¹⁶, the most recent morphological and molecular information revealed that *P. volitans* is a recent hybrid species between the Indian lineage of *P. miles* and a Pacific lineage encompassing *P. lunulata* and *P. russelli*²⁵.

Population genetic studies using the mitochondrial d-loop fragment of *P. volitans*, conducted across the Wider Caribbean, supported the chronological records of the invasion and confirmed the colonization routes followed by the species^{15,16,20,21,23,26,27}. While these studies discarded the hypothesis of multiple introductions from native sources during the invasion process, they revealed successive bottlenecks over the course of the invasion. Across the Wider Caribbean, only nine d-loop haplotypes were identified among the total 1,248 samples sequenced^{16,20,21,26,27}, a number that contrasts with the 37 different haplotypes sequenced in the native range at the Indo-Pacific in only 70 specimens²⁰. D-loop analyses showed that the first bottleneck occurred from the native range to the NW Atlantic, although the NW Atlantic area (east coast of the US), identified as the entrance point of lionfish, displayed the highest nucleotide and haplotype diversity and number of haplotypes within the Wider Caribbean, with the nine different haplotypes¹⁶. The second bottleneck occurred within the invasive range when of the nine haplotypes found at the NW Atlantic only four secondarily spread to the Caribbean Sea. Finally, three of the four haplotypes found in the Caribbean Sea invaded the Gulf of Mexico^{16,20,21,26,27}. Hence, the lionfish invasion initially contrasts with the perception that avoiding genetic bottlenecks by the influx of genetic diversity through repeated introductions from the native range increases the invasion success and posterior spread within the invasive area^{28,29}, as observed in a number marine invasions fuelled by multiple introductions from different native sources^{30,31}. The study by Johnson and co-authors (2016), which pooled together all d-loop sequences obtained from the Wider Caribbean between 2007 and 2013, also detected significant differences in genetic structure among the three mentioned areas: NW Atlantic, Caribbean Sea, and the Gulf of Mexico (See Fig. 1), but no significant differences within them²¹, despite the genetic divergence noticed among some populations of the Caribbean Sea in a previous study²⁷. Additionally, the sharp genetic discontinuity between the NW Atlantic and the Gulf of Mexico suggested no direct gene flow across the Strait of Florida²¹.

Despite the important insights revealed by the d-loop, using only one mitochondrial marker may prevent detection of fine-scale differentiation and connectivity within lionfish's invasive range¹⁶, and hence nuclear loci are required to provide a complete picture of the current genetic status of this invasion. In this study, we focus our attention on *P. volitans*, referred to as “lionfish” hereafter. We analyse two areas chronologically identified as the first and last invaded areas within the Wider Caribbean: the NW Atlantic (east coast of Florida) and the northern Gulf of Mexico, respectively. These two marine areas displayed highly significant differences in genetic diversity, structure and absence of connectivity across the Florida Strait for the d-loop marker in a previous study²¹.

We here apply Genotyping by Sequencing (GBS) techniques to identify over a thousand nuclear and independent single nucleotide polymorphic sites (SNPs) from samples collected over a short time. Our general aim is to understand the invasion progression by determining fine-scale population genomics between the NW Atlantic (east coast of Florida) and the northern Gulf of Mexico, and potential connectivity between them, knowing that the spatial genetic structure, temporal genetic trend, and levels of connectivity among areas is relevant to predict the potential long-term persistence of an invader²⁹ and to design management strategies for its control.

Results

From the Genotyping by Sequencing (GBS) library of 229 lionfish samples, a total of 404,254 sequence tags were retained in TASSEL³². Filtering for individuals with at least 75% of the called loci and loci that were present in at least 84% of individuals yielded a total of 1,654 SNPs and 162 individuals from the NW Atlantic and Gulf of Mexico (see Tables 1 and 2, and Fig. 1). From this dataset, 434 SNPs were removed: 61 SNPs that showed significant linkage disequilibrium ($r^2 > 0.2999$, FDR p -value < 0.01) and 373 SNPs with significantly greater observed than expected heterozygosity (Hardy Weinberg Equilibrium- HWE, p -value < 0.01) (see Supplementary Material FS1), leaving a final dataset of 1,220 SNPs in 162 specimens covered by 2,322,797 reads (mean: 11.75 reads per locus and sample).

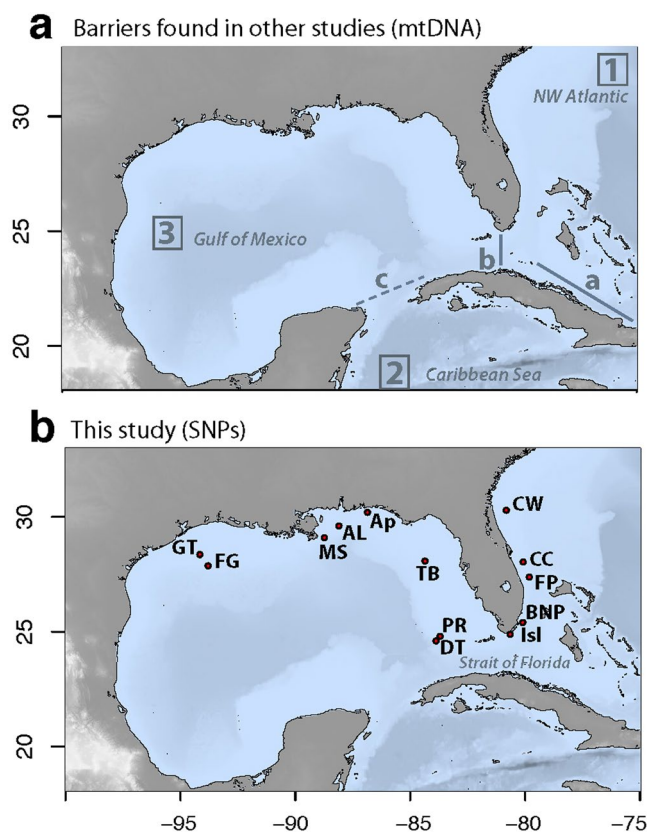


Figure 1. Sampling collection of lionfish. (a) Map showing major genetic discontinuities (a,b and c in blue) and genetically isolated areas (according to the temporal invasion progression as 1, 2 and 3) detected in previous studies from mitochondrial DNA of *P. volitans*²¹, and (b) Map of the sampling sites in this study. Maps were created for this study with the “marmap” package⁷³ in R⁶⁴.

Site	Area	Code	Ng	Date	Latitude	Longitude	Depth	Preservation
North East. FL	NW Atlantic	CW	10	2015	30.29	−80.82	no data	Ethanol
Cape Canaveral. FL	NW Atlantic	CC	6	2015	28.04	−80.09	27	Ethanol
Biscayne National Park. FL	NW Atlantic	BNP	38	2015	25.4	−80.1	4 to 36	Chaos
Ft. Pierce. FL	NW Atlantic	FP	5	2015	27.38	−79.82	24	Ethanol
Islamorada. FL	NW Atlantic	Isl	30	2015	24.88	−80.65	no data	Chaos
Dry Tortugas. FL	Florida Keys	DT	19	2013	24.59	−83.88	62	Ethanol
Pulley Ridge	Florida Keys	PR	26	2013	24.8	−83.70	62	Ethanol
Tampa. FL	Gulf of Mexico	TB	19	2014	28.08	−84.36	34	Ethanol
Apalachicola. FL	Gulf of Mexico	Ap	19	2014	30.2	86.86	35	Ethanol
Alabama Shelf. AL	Gulf of Mexico	AL	20	2014	29.61	−88.101	39	Ethanol
Mississippi Delta. LA	Gulf of Mexico	MS	10	2013	29.09	−88.734	44	Ethanol
Flower Garden Banks. TX	Gulf of Mexico	FG	10	2015	27.87	−93.80	24	Frozen + ethanol
Galveston. TX	Gulf of Mexico	GT	17	2014	28.36	−94.157	27	Ethanol

Table 1. General information of lionfish samples. Sampling site, major marine area, site code, number of individuals genotyped (Ng), year of collection (Date), geographical coordinates (Latitude and Longitude), collection depth (in meters), and preservation method.

Detection of SNP outliers. From the dataset of 1,220 SNPs, 23 outlier SNPs were identified as candidate markers under positive selection with Arlequin after FDR correction (24 SNPs were identified when the uncorrected p -value ≤ 0.01 was applied). No marker was identified to be under balancing selection from Arlequin (see Fig. 2). Lositan identified a total of 213 outlier SNPs, 73 of them candidates under positive selection (Fig. 2) and 140 candidates under balancing selection. Among all outlier SNPs, only 13 were found in common between both methods. We consider that only these 13 outlier SNPs had strong statistical support to be considered as

Code	Nf	mean n° alleles	Ho	He	F _{IS}	HWE
CW	7	1.85 ± 0.001	0.268 ± 0.006	0.289 ± 0.005	0.054 ± 0.013	0.001
CC	1	0.87 ± 0.017	0.121 ± 0.009	0.061 ± 0.005	—	—
BNP	36	1.99 ± 0.002	0.358 ± 0.006	0.337 ± 0.004	−0.027 ± 0.001	1.000
FP	4	1.76 ± 0.012	0.230 ± 0.008	0.280 ± 0.005	−0.080 ± 0.018	1.000
Isl	29	1.99 ± 0.002	0.359 ± 0.006	0.337 ± 0.004	−0.033 ± 0.001	1.000
DT	6	1.74 ± 0.013	0.186 ± 0.006	0.256 ± 0.005	0.216 ± 0.015	0.001
PR	20	1.98 ± 0.004	0.294 ± 0.005	0.310 ± 0.004	0.057 ± 0.01	0.001
TB	11	1.90 ± 0.009	0.225 ± 0.004	0.288 ± 0.005	0.193 ± 0.013	0.001
Ap	10	1.88 ± 0.009	0.237 ± 0.005	0.289 ± 0.004	0.151 ± 0.012	0.001
AL	20	1.98 ± 0.004	0.309 ± 0.006	0.320 ± 0.004	0.053 ± 0.011	0.001
MS	1	0.84 ± 0.016	0.099 ± 0.009	0.050 ± 0.004	—	—
FG	4	1.58 ± 0.015	0.191 ± 0.007	0.220 ± 0.006	0.081 ± 0.017	0.001
GT	13	1.93 ± 0.007	0.256 ± 0.005	0.298 ± 0.004	0.120 ± 0.012	0.001

Table 2. Main genetic descriptors for all 1,220 SNPs. Population code, number of individuals retained for analyses after filtering (Nf), mean number of alleles, observed and expected heterozygosity (Ho and He, respectively), Fixation index (F_{IS}) and *p*-values of Hardy Weinberg Equilibrium, Mean number of alleles, Ho, He and F_{IS} are presented with standard errors.

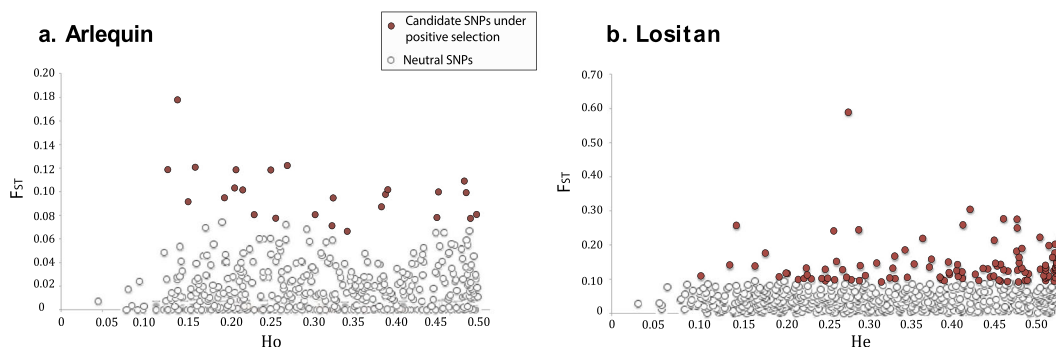


Figure 2. Detection of outlier SNPs. (a) Using coalescent simulations with Arlequin: F_{ST} and observed heterozygosity (Ho), and (b) Coalescent simulations using Lositan: F_{ST} and expected heterozygosity (He). Neutral SNPs are plotted as white circles and candidate SNPs under positive selection are represented as red circles. Candidate SNPs under balancing selection are not represented in this graph.

potentially under positive selection. The remaining 1,207 SNPs were assumed to be neutral, although their neutrality could not be directly proven.

Lack of divergence between the NW Atlantic and the northern Gulf of Mexico. The main genetic descriptors obtained for all 1,220 lionfish SNPs are listed in Table 2. Most populations showed lower values of observed (Ho) than expected heterozygosity (He), which translated into positive values of the fixation index F_{IS}, and significant deviation from HWE (Table 2). Genetic diversity values varied among populations: populations with only few individuals had the lowest diversity. For those populations with 10 or fewer individuals, the mean number of alleles per locus was lower than the potential maximum of 2 (see the curve of accumulative number of alleles related to sample size in Supplementary Material FS2). Our analyses did not detect significant differences in genetic diversity (assessed as Ho and He) between the NW Atlantic and the Gulf of Mexico (t-tests: *t* = 0.89 and 0.42, *p*-values = 0.39 and 0.68, respectively).

The Maximum Likelihood (ML) tree reconstructed from 1,220 SNPs of 162 specimens from 13 locations did not show geographical clustering of specimens related to the sampling site and/or geographical area where they were collected (Supplementary Material FS3) and displayed very low bootstrap value support on most nodes.

In the Bayesian clustering analysis performed in Structure for all 1,220 SNPs, the optimal numbers of homogeneous genetic clusters (K) for the whole data set³³ were three and five (K = 3 and K = 5) according to the ad hoc statistic ΔK (see values of Delta K- ΔK- in Supplementary Material FS4), but the Log likelihood for K (LK) did not significantly increased from 1 to 5 suggesting lack of spatial genetic clustering (see LK in Supplementary Material FS4). The individual-based cluster memberships from K = 2 to K = 8 (see Supplementary Material FS4) showed no spatial genetic heterogeneity among sampling sites and mixed membership of individuals to all clusters, supporting the hypothesis of panmixia across the Florida Strait (see Fig. 3 for K = 3 and K = 5, and Supplementary Material FS4 for all K values). Only slight differences in terms of a higher probability of one specific cluster (the blue cluster) could be observed in two sites located at south Florida, BNP and Isl (Fig. 3). Genetic admixture across the invasive range was also observed when 1,207 neutral SNPs and 13 outlier SNPs were separately analysed. In Fig. 3, we compare Structure results for K = 5 from the three different datasets.

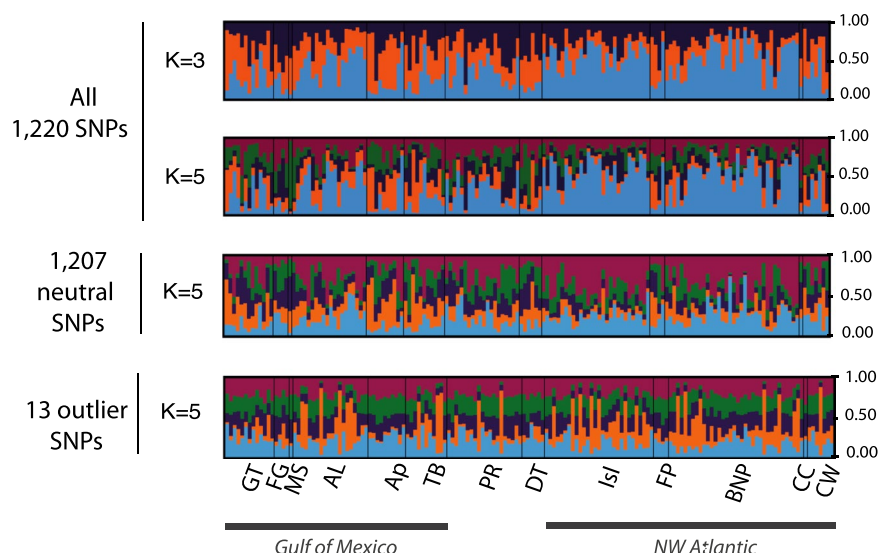


Figure 3. Structure barplot outputs. Posterior probabilities of individual assignment of the most probable number of clusters ($K = 3$ and 5 ; different clusters are represented by different colours) for all 1,220 SNPs. Barplots from 1,207 neutral SNPs and 13 candidate SNPs under selection (outliers) are also presented. For neutral and outlier SNPs the graph of $K = 5$ is shown for comparison with the whole dataset.

Source of variation	d.f	% Var.	Fixation index	<i>p</i> -value	% Var.	Fixation index	<i>p</i> -value
		All 1,220 SNPs			13 outlier SNPs		
<i>PR and DT within NW Atlantic</i>							
Among groups	1	0	−0.001	0.99	0	−0.012	0.58
Among populations within groups	11	0	−0.006	0.99	5.7	0.056	0.00
Within populations	149	0	−0.036	1	27.3	0.286	0.00
Within individuals	162	100	−0.044	0.85	68.2	0.318	0.00
<i>PR and DT within Gulf of Mexico</i>							
Among groups	1	0	−0.002	0.90	0	−0.012	0.58
Among populations within groups	11	0	−0.006	1	5.7	0.056	0.00
Within populations	149	0	−0.037	0.98	27.3	0.286	0.00
Within individuals	162	100	−0.449	0.95	68.2	0.318	0.00
<i>Without PR and DT</i>							
Among groups	1	0	−0.002	0.87	0	−0.016	0.52
Among populations within groups	11	0	−0.005	0.99	7.04	0.069	0.00
Within populations	125	0	−0.045	0.98	28.3	0.299	0.00
Within individuals	136	100	−0.053	0.99	68.2	0.338	0.00

Table 3. Results of analyses of molecular variance (AMOVAs) of lionfish from all 1,220 SNPs and 13 candidate SNPs under selection (outliers). Populations' grouping (Source of variation), degrees of freedom (d.f.), percentage of variation (% Var.), fixation indexes, and p -values.

The Analyses of Molecular Variance (AMOVA) for all 1,220 SNPs, 1,207 neutral SNPs, and 13 outlier SNPs also did not detect genetic differentiation associated with the NW Atlantic and Gulf of Mexico (“Among groups”), regardless of whether the two populations from the Florida Keys (Dry Tortugas- DT and Pulley Ridge- PR) were pooled or removed from the analyses, and most genetic variation was retained within individuals (Table 3). No differences were observed from AMOVA results between the datasets including all 1,220 SNPs and only 1,207 neutral SNPs (see Table 3 and Supplementary Material TS1). Nevertheless, from the 13 outlier SNPs, we detected significant differences at two additional variance components: between populations within the NW Atlantic and Gulf of Mexico, and among individuals within populations, but still most genetic variation was retained within individuals (Table 3). F_{ST} statistics gave us more details of the pairwise differences between populations within the NW Atlantic and the Gulf of Mexico. The F_{ST} statistics were in general very low, and only two pairwise comparisons were significant (between AL and BNP, and Isl) after FDR correction of the p -values when all 1,220 SNPs (Table 4) or only 1,207 neutral SNPs (data not shown) were included in the analyses. F_{ST} distances from the 13 outlier SNPs revealed significant genetic differentiation between other sites: Ap and FG seemed to be the most

	CW	BNP	FP	Isl	DT	PR	TB	Ap	AL	FG	GT
CW	—	0.05904	0.12398	0.07693	0.08734	0.08357	0.09997	0.12432	0.02388	0.26866	0.11102
BNP	−0.00585	—	0.13273	0.01039	0.05440	0.01263	0.04417	0.08623*	0.01792	0.08177	0.05091
FP	−0.0091	−0.01163	—	0.15015	0.11214	0.13219	0.12923	0.13048	0.12197	0.28386	0.10115
Isl	−0.00569	−0.00205	−0.00986	—	0.01259	0.00150	0.00448	0.08536*	0.00696	0.13068*	0.03878
DT	−0.01861	−0.03608	−0.01768	−0.04071	—	−0.04126	0.05904	0.04868	0.01374	0.30744	−0.03108
PR	−0.00091	−0.00059	−0.00483	−0.00267	−0.02645	—	0.04344	0.03418	0.00650	0.16721	−0.00889
TB	−0.00174	−0.01366	0.00119	−0.01649	−0.00057	−0.00173	—	0.11251	0.01298	0.26521*	0.01745
Ap	−0.00228	−0.01181	−0.0075	−0.01483	−0.00275	−0.00333	0.01082	—	0.09401*	0.26914*	0.02245
AL	−0.00374	0.00136*	−0.00754	0.00211*	−0.02251	0.00291	−0.00487	−0.0038	—	0.20537*	0.01672
FG	−0.00706	−0.03463	−0.00706	−0.03926	0.01363	−0.023	0.01023	0.01088	−0.02301	—	0.24847*
GT	0.00236	−0.00695	−0.00131	−0.00917	−0.01311	−0.00199	−0.00225	−0.00287	−0.00034	−0.00234	—

Table 4. Lionfish F_{ST} values between sampling sites. F_{ST} values from all 1,220 SNPs below the diagonal, and 13 candidate SNPs under selection (outliers) above the diagonal. *Significant after false discovery rate (FDR) correction set at $p \leq 0.01$.

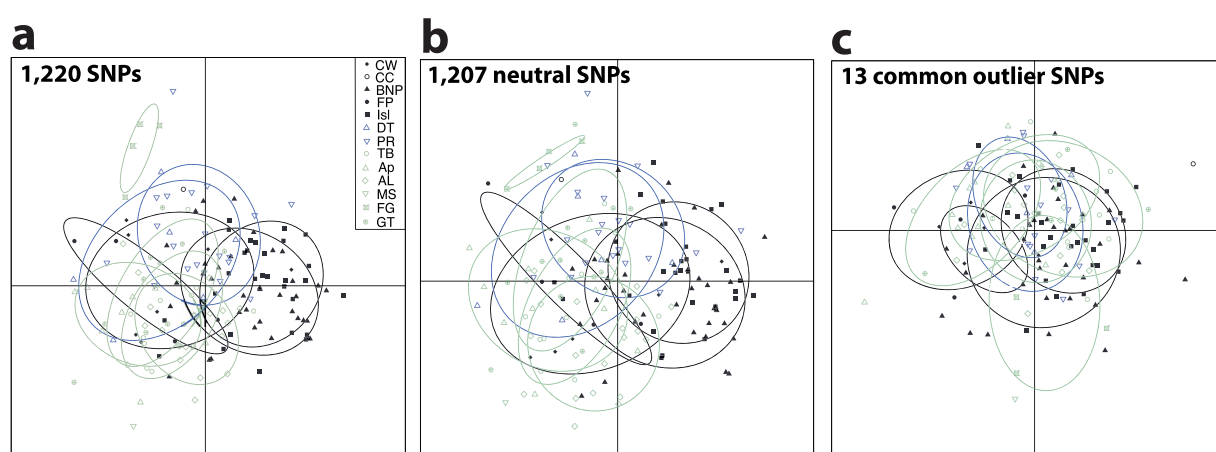


Figure 4. DAPC results. The DAPC graphs represent results from three different analyses: (a) from all 1,220 SNPs, (b) 1,207 neutral SNPs, and (c) 13 candidate SNPs under selection (outliers). In the DAPC graph points represent different individuals, point patterns different sampling sites, and colours different marine areas (grey = NW Atlantic, blue = Dry Tortugas and Pulley Ridge, and light green = northern Gulf of Mexico).

genetically divergent sites (see Table 4), but this genetic divergence was not fully supported by Bayesian clustering analyses (see previous results from Structure) or discriminant analyses of principal components (see explanation below).

The discriminant analyses of principal components (DAPC) also showed a general pattern of low genetic differentiation, as that observed from previous analyses. According to the Bayesian Information Criterion (BIC) that compares different DAPC clustering solutions, two clusters were the optimal number to describe our data. The DAPC plot, including all sampling sites and all 1,220 SNPs, showed no clear separation of populations or clusters between the NW Atlantic and Gulf of Mexico (Fig. 4a). Only FG seemed lightly isolated from all the other sampling sites. This pattern was maintained when only 1,207 neutral SNPs were included in the DAPC analysis (Fig. 4b). When 13 outlier SNPs were separately analysed, the divergence between FG and all the other populations decreased (Fig. 4c).

Discussion

Our genomic data of 1,220 SNPs from 162 *P. volitans* specimens across the NW Atlantic and the northern Gulf of Mexico represents the first study using nuclear loci to explore the genetic structure of this invasive predator and is among the few studies applying genome-wide scanning, based on Next Generation Sequencing technologies, to investigate population structure of a marine invader (see a review in ref.³⁴ and examples in refs.^{35,36}). Although 18 nuclear microsatellite loci were isolated for *P. volitans* and *P. miles* a few years ago³⁷, to our knowledge, those markers have not yet been used for population analyses.

Our fine-scale population genomic analyses of lionfish demonstrate lack of a current genetic break between the first and the last invaded areas in US waters: the NW Atlantic and the Gulf of Mexico, respectively. The Bayesian clustering analysis and DAPC showed different clustering solutions due to a lack of clear genetic differentiation across the whole analysed area. From all 1,220 SNPs and 1,207 neutral SNPs, we only noticed significant genetic differences between three sites based on F_{ST} distances; the 13 outlier SNPs, FG and Ap showed significant

differences with five additional sites, but that divergence was not mirrored by other analyses (e.g. Structure and DAPC). Only FG seemed to be genetically divergent for most analyses and databases. However, this location only includes four individuals, a sample size that cannot be considered representative of the genetic diversity and structure of this location, as demonstrated by the low number of accumulated alleles within these four individuals (see Table 2). Hence, independently of the database used (all 1,220 SNPs, 1,207 neutral SNPs, or 13 outlier SNPs), results show a picture of general genetic homogeneity with enough gene flow between the NW Atlantic and the Gulf of Mexico to erase signals of secondary spreading detected from d-loop analyses in previous years and studies²³. Whether current gene flow occurs directly across the Florida Strait or indirectly throughout the Caribbean Sea cannot be assessed by the presented data, and further analyses including samples from the Caribbean Sea are necessary to investigate connectivity routes.

Our findings, therefore, contrast with the discontinuity between the NW Atlantic and the Gulf of Mexico and the lower genetic diversity of the Gulf of Mexico observed in previous studies based on d-loop data^{16,20,21,27}. Different mitochondrial and nuclear DNA patterns (mito-nuclear discordance), such as those noticed here for lionfish, are becoming more commonly reported as the number of nuclear multilocus datasets increases (see examples in refs^{38,39}) and can be explained by several non-exclusive causes⁴⁰. Demographic asymmetry due to sex-biased dispersal can cause mito-nuclear discordance in motile animals, including marine fish species^{40,41}. Although recent findings suggest that lionfish adults move more than initially thought⁴², they are not migratory, and buoyant eggs are the dispersal stage¹¹, so different migratory behaviour between males and females cannot explain the pattern we found, and other hypotheses should be considered, including temporal genetic shifts and selection.

A plausible cause of discordance between lionfish studies using different markers is temporal changes in the genetic structure over the invasion process. Population genetics theory anticipates fast genetic changes in introduced populations characterized by bottlenecks, founder effects, strong genetic drift, and new selective pressures in the introduced environments^{43–45}. Despite the importance of temporal genetic trends for the invasion dynamics, this point is overlooked in most studies of marine invaders, and it is assumed that genetic diversity remains stable over time³⁴. The few studies investigating temporal trends of genetic structure in introduced marine species showed variable outcomes^{46–51}. Whereas some invasive ascidians suffering massive seasonal die-off events maintained stable levels of genetic diversity and homogeneous structure over time due to the re-establishment of populations from the survivors or recolonization from nearby sites^{48,51}, other introduced species exhibited changes in genetic architecture over short time periods. For instance, in the introduced colonial ascidian, *Perophora japonica*, a genetically isolated population from Plymouth (South England) displayed a linear reduction in mitochondrial genetic diversity and large haplotype frequency changes over a 9 year-monitoring period, due to either genetic drift and/or selection⁴⁶. Rapid allele frequencies changes over time, high heterozygous deficiency, and inbreeding were also detected in isolated populations of the colonial ascidian *Botryllus* along the coast of Israel⁴⁹, but genetic isolation was not always associated with genetic diversity loss in this species. Invasive *Botryllus* populations along the Californian coast, isolated from other genetic sources and highly influenced by genetic drift and selection, maintained stable levels of genetic diversity thanks to high mutation rates generating a complex pattern of allele gains and losses⁴⁷. Nevertheless, marine invasive populations are, in many cases, characterised by high levels of genetic diversity due to multiple introductions from genetically distinct sources²⁸. An outstanding example of multiple cryptic introductions and genetic admixture within the invaded range, which might be related to the high invasion success, is that of the European green crab, currently one of the most important aquatic invaders established across all temperate shores around the world^{30,36}.

In lionfish, d-loop data revealed strong bottlenecks and scientists discarded the idea of multiple introductions into the NW Atlantic from the native range^{16,20}, which would result in a small initial effective population size. Moreover, changes in genetic structure are expected to occur faster in mitochondrial than nuclear DNA because mitochondrial DNA, a haploid, maternally inherited molecule, has an effective population size of one-quarter that of nuclear DNA and therefore is more sensitive to diversity changes associated with genetic drift. For this reason, the comparison of the NW Atlantic d-loop data (collected between 2007 and 2009) and the Gulf of Mexico (collected between 2011 and 2013)^{16,20,21,27} should be taken with caution since it assumes that NW Atlantic populations remained static over a six-year period with no changes in haplotype frequencies due simply to genetic drift. This assumption of temporal stability might obscure the most recent pattern of genetic diversity in lionfish. In this sense, the analyses presented here, based on 1,220 nuclear SNPs from samples collected during a brief period of 20 months seems to be a more reliable way to determine lionfish's current genetic structure.

Besides neutral temporal trends, differential selection can also promote discrepancy patterns between mitochondrial and nuclear DNA. Mitochondrial selection under divergent environmental conditions plays an important role for the distribution of mitochondrial variants in other marine fish species^{52,53}. Different environmental pressures between the NW Atlantic and the Gulf of Mexico could also have favoured some lionfish mitochondrial haplotypes over others, thus shaping the spatial distribution of haplotypes during the invasion process. However, the sampling scheme used in our study and the lack of new mitochondrial sequences do not allow mitochondrial selection and/or adaptation hypotheses to be tested. None of the 1,220 tags containing the SNPs were identified as mitochondrial fragments, and the different analyses performed did not reveal evidence of local adaptation and/or nuclear selection across the lionfish's invasive range. In some marine species with long-dispersal potential, outlier SNPs unravelled significantly finer genetic structure than neutral markers, suggesting the existence of local adaptation^{36,54–57}. For example, in the European hake, Atlantic and Mediterranean populations showed sharper divergence in analyses using outlier SNPs than neutral SNPs⁵⁴, a pattern of higher resolution that was also found in other fish species at small geographical scales of a few hundred kilometres when analysing outlier SNPs^{55,56}. In some marine invaders, local adaptation also seemed to play an important role in shaping populations' genetic structure, showing either latitudinal clines in outlier allele frequencies³⁶ or significant correlation with environmental variables such as salinity and water temperature⁵⁷. Nevertheless, in lionfish, differential selection between

the NW Atlantic and the Gulf of Mexico, and mito-nuclear interactions remains as an open question because although we did not find strong evidence of local adaptation, 1,220 SNPs still represent a small proportion of the species' genome, and selection on non-explored genomic areas could be possible.

The SNP data here presented yield valuable information of the genetic trend in the lionfish invasion, which could potentially be affected by genetic changes over time and across space, although other hypotheses such as different selective pressures between mitochondrial and nuclear DNA cannot be completely discarded. Additionally, we perceive some limitations in our study that should be taken in consideration for further investigations. For instance, we noticed that population analyses based on SNPs should include: sizes over 10 individuals to retain the potential maximum genetic diversity within populations, representative populations from the native range to shed light on the current impact of bottlenecks in genetic diversity across the whole invasive range, and populations from the Caribbean Sea to clarify the most important connectivity routes within the invaded area.

As demonstrated by previous publications based on mitochondrial DNA, which detected strong bottlenecks during the first introduction steps and invasion progression^{16,20,21,27,28}, the lionfish invasion is an example of how reductions in genetic diversity do not necessarily compromise population establishment and spreading^{16,20,21,27,28} and points out the importance of primary (pre-border) and secondary (post-border) introductions, e.g., secondary introductions to the Caribbean Sea and later to the Gulf of Mexico^{16,20,21,27}. The potential of lionfish to overcome these initial steps of the invasion with low genetic diversity at the mitochondrial DNA^{16,20,21,27}, and to homogenize nuclear genetic structure across the invaded area (as shown in this study with SNPs), should be considered when developing theoretical models on the expected geographical spreading of this invasion⁵⁸ and implementing appropriate strategies for its management and control.

Finally, the lionfish invasion to the Wider Caribbean can be used as a lesson to anticipate the genetic trend and potential impacts of *P. miles* invasion in the Mediterranean Sea. As *P. volitans* across the Wider Caribbean, *P. miles* has quickly colonized wide areas of the eastern Mediterranean⁵⁹, which adds an additionally threaten in a small sea that is at the same time a hotspot of marine biodiversity and one of the world's most impacted seas.

Methods

Sampling collection. *P. volitans* samples were collected over a 20 month period, between June 2013 and February 2015, from thirteen locations along Florida's eastern coast (NW Atlantic and Florida Keys) and the northern Gulf of Mexico, at depths between 4 and 62 meters. Sampling sites and number of individuals genotyped are detailed in Table 1 and Fig. 1. Collections were often opportunistic by SCUBA divers, so collection depths could not always be recorded. Fin or gill clips were obtained from the collected specimens and preserved in absolute ethanol, frozen at -20°C or stored in 320 μl of chaotropic buffer (4.5 M guanadimium thiocyanate, 2% N-lauroylsarcosine, 50 mM EDTA, 25 mM Tris-HCL pH 7.5, 0.2% antifoam, 0.1 M β -mercaptoethanol) (see Table 1).

Ethics Statement. No endangered or protected species were involved in this study. Lionfish were sampled opportunistically by the authors from lionfish derbies or state and federal collections (as stated below); only dead lionfish were obtained. Lionfish were collected by a number of organizations in areas open to fishing with a spear or permitted by methods utilized. These fish were collected as a result of other activities such as tournaments, commercial harvest, and general fisheries surveys, and were sampled opportunistically for this study. No permits were required to collect lionfish beyond a state saltwater fishing license, which was in possession of divers at each collection. In the case of lionfish collected from offshore Florida, no fishing license is required. The University of Miami Institutional Animal Care and Use Committee (IACUC) did not require a protocol for this study since only dead specimens were donated to the University. State and Federal government organizations, although exempt from IACUC requirements, follow best practices to minimize pain and suffering of specimens. These are approved Institutional Animal Care and Use Committee protocols via the American Veterinary Medical Association Guidelines for the Euthanasia of Animals and the American Society of Ichthyologists and Herpetologists Guidelines for Use of Fish in Research.

Library construction and SNP isolation. Genomic DNA was extracted from tissue clips using silica columns. DNA quality was assessed via agarose gel electrophoresis, and DNA concentrations were quantified using Biotium AccuBlue™ Broad Range dsDNA Quantitative Solution according to the manufacturer's instructions. After quantification, 100 ng of DNA from each sample was dried down in 96-well plates in a SpeedVac concentrator. Samples were then rehydrated overnight with 5 μl of ultrapure milliQ water before further processing.

Genotyping by Sequencing libraries were constructed using the restriction enzyme *ApeKI*. A total of 50 ng of genomic DNA per sample was digested at 75°C for 2 hours. Unique barcoded adapters were used for library construction as described in⁶⁰. A total of 229 DNA samples were pooled together and fragments approximately 300 bp in length were selected with magnetic beads. Primers complementary to the adapters were then used for library amplification⁶⁰. Before sequencing, library quality was checked in an Agilent 2100 Bioanalyzer. The GBS library including the 229 individuals was sequenced in two lanes of an Illumina Hi Seq. 2500 using 75 bp single end reads at Elim Biopharmaceuticals, Inc. Hayward, CA.

The UNEAK GBS analysis pipeline in TASSEL³² for species without a reference genome was used to call SNPs using Bowtie⁶¹. The software identifies SNPs found on single non-overlapping "tags" (64 bp sequences) initiated at the restriction sites. Only SNPs that had a minimum of five reads across all samples were retained to reduce the impact of sequencing errors. Loci with significant linkage disequilibrium (D' p -value False Discovery Rate correction-FDR- adjusted to 0.01) identified in TASSEL and those with significantly greater observed than expected heterozygosity (p -value < 0.01) were removed from the database before performing further analyses. SNPs were then filtered to select individuals with at least 75% of the called loci and loci that were present in at

least 84% of individuals. Maximum heterozygosity during filtering was set at 0.5 to avoid excess of heterozygotes due to sequencing errors. All sequences containing selected SNPs were blasted (e-value $< 10^{-5}$) against the mitochondrial DNA of *Salmo salar* and the Genbank database to identify mitochondrial fragments.

Data availability. The HapMap file including the whole dataset of SNPs here analysed, a coverage file and the genepop file including allele frequencies have been deposited in PANGAEA (<https://doi.org/10.1594/PANGAEA.886118>).

Detection of outlier SNPs. Two different software programs, Arlequin⁶² and Lositan⁶³, were used to identify non-neutral SNPs, as candidate markers under selection, based on an F_{ST} -outlier detection method and coalescence simulations. Arlequin uses simulations based on observed heterozygosity (H_o) to create a null distribution of F_{ST} values and associated p -values for each locus. We performed a total of 20,000 simulations, with 100 demes, under a finite island model. This model was chosen due to the general lack of genetic structure (see Results). FDR correction of the p -values was applied to detect significant outliers; we also considered a more conservative approach with significance at $p < 0.01$ since strong corrections can increase type II error thereby assuming neutrality in SNPs that are not neutral⁵⁵ (although both approaches showed similar results). Lositan, on the other hand, creates a distribution based on the relation between F_{ST} values and expected heterozygosity (H_e). We performed a first run using all loci to estimate mean F_{ST} values with 20,000 simulations, 99% confidence interval, infinite alleles mutation model and false discovery rate of 0.1%. After the first run, loci in the confidence interval were removed, and “neutral” F_{ST} values were recalculated. A third run was finally performed using all loci, and the neutral F_{ST} values previously calculated were implemented to detect outliers. Finally, outliers recovered from both software programs, Arlequin and Lositan, were considered as candidate SNPs under selection.

Genetic structure analyses. General descriptors of genetic diversity as mean number of alleles, observed heterozygosity (H_o), expected heterozygosity (H_e), fixation index F_{IS} , and the Hardy Weinberg Equilibrium were calculated for all markers per population using Arlequin 3.5.1.2⁶² and the “adeget” package in R⁶⁴.

A Maximum Likelihood (ML) tree, including all genotypes obtained, was reconstructed in RAxML with a GTR+ G model and 100 rapid bootstrap replicates⁶⁵ to explore potential clustering of individuals related to different geographical areas and/or sampling sites. The ML tree was then visualized and edited in Figtree 1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>).

A Bayesian clustering analysis, performed with the software Structure 2.3.4⁶⁶, was used to investigate the optimal number of major homogeneous genetic clusters (K) found within our datasets under the null hypothesis of genetic homogeneity. Because Bayesian analysis can be computationally very intense and long, an initial fast run was performed with a K from 1 to 13 with five independent replicates, 20,000 Markov chain Monte Carlo (MCMC) per replicate, and a 2,000 burn-in period to get a general idea about the maximum number of clusters expected. Then, a definitive run was performed with K from 1 to 8 with five independent replicates, 100,000 MCMC per replicate, and a 10,000 burn-in period. We used an “admixture model” and correlated gene frequencies as implemented in Structure. The five independent runs were averaged using the clumpak server⁶⁷ (<http://clumpak.tau.ac.il>). The K value was determined by comparing the rate of change in the likelihood of K , using the ad hoc statistic ΔK in Structure Harvester 0.6.94⁶⁸.

Analyses of Molecular Variance (AMOVA), based on allele frequencies, were performed to specifically explore the potential genetic break between the two genetically different regions previously identified from mitochondrial DNA, the NW Atlantic and the Gulf of Mexico. The locations of PR and DT, at the Florida Strait, are rich mesophotic reefs and part of the Florida Keys reef complex but far inside the Gulf of Mexico. Since the genetic break between the NW Atlantic and the Gulf of Mexico shifts at different points of the Florida Strait depending on the species^{69,70}, we could not *a priori* assign these two sites (PR and DT) to one or the other area. Therefore, we performed three different AMOVA analyses: the first analysis included PR and DT within the NW Atlantic pool, the second included them within the Gulf of Mexico pool, and the third one excluded these two sites from the analysis. After testing differences between major marine areas, pairwise F_{ST} distances based on allele frequencies between all sampling sites were calculated with the same software. The significance of AMOVA and F_{ST} values was assessed after 50,000 non-parametric permutations of individuals among populations and/or populations between geographical areas and under the null hypothesis of genetic homogeneity. FDR correction of these p -values was applied for F_{ST} multiple testing⁷¹.

Significant difference in genetic diversity (H_o and H_e) between the two marine regions, the NW Atlantic and Florida Keys (CW, CC, FP, BNP, Isl, PR and DT, see Results section), and the Gulf of Mexico (TB, Ap, AL, MS, FG, GT) was evaluated with a t -test.

Additionally, discriminant analyses of principal components (DAPC)⁷² were computed for the complete dataset. DAPC does not assume any underlying population genetic model and is not as affected by Hardy Weinberg disequilibrium as other methods based on genetic distances (e.g. F_{ST} and AMOVA) and Bayesian clustering analyses. We used collection sites as populations with the “adeget” package in R⁶⁴. DAPC extracts multivariate information from genetic datasets by first performing a principal component analysis (PCA) on predefined groups (collection sites in this case) and then using the PCA factors as variables for a discriminant analysis (DA), which seeks to maximize the inter-site component of variation. Thus, DAPC allows the visual identification of genetic clusters and can outperform more computer-intensive approaches, such as Structure, in detecting genetic structure⁷². Since the number of principal components (PCs) retained may have large impact on the DAPC output, the optimal number of PCs to be retained was first explored by the cross-validation method implemented by this package.

To understand whether selection and/or local adaptation within the lionfish's invasive range is an important driver of the genetic structure, the searching strategy explained before for the Bayesian clustering analysis, AMOVA, F_{ST} and DAPC was comparatively applied to three different SNP datasets: for all isolated SNPs, for neutral SNPs, and for candidate SNPs under selection (outliers).

References

- Green, S. J., Akins, J. L., Maljković, A. & Côté, I. M. Invasive lionfish drive Atlantic coral reef fish declines. *PLoS one* **7**, e32596 (2012).
- Sutherland, W. J. *et al.* Horizon scan of global conservation issues for 2011. *Trend Ecol. Evol.* **26**, 10–16 (2011).
- Whitfield, P. E. *et al.* Biological invasion of the Indo-Pacific lionfish *Pterois volitans* along the Atlantic coast of North America. *Mar. Ecol. Prog. Ser.* **235**, 289–297 (2002).
- Albins, M. A. & Hixon, M. A. Invasive Indo-Pacific lionfish *Pterois volitans* reduce recruitment of Atlantic coral-reef fishes. *Mar. Ecol. Prog. Ser.* **367**, 233–238 (2008).
- Albins, M. A. & Hixon, M. A. Worst case scenario: potential long-term effects of invasive predatory lionfish (*Pterois volitans*) on Atlantic and Caribbean coral-reef communities. *Environ. Biol. Fishes* **96**, 1151–1157 (2013).
- Côté, I. M., Green, S. J. & Hixon, M. A. Predatory fish invaders: Insights from Indo-Pacific lionfish in the western Atlantic and Caribbean. *Biol. Conserv.* **164**, 50–61 (2013).
- Green, S. J. & Côté, I. M. Record densities of Indo-Pacific lionfish on Bahamian coral reefs. *Coral Reefs* **28**, 107–107 (2009).
- Morris, J. A. Jr & Freshwater, D. W. Phenotypic variation of lionfish supraocular tentacles. *Environ. Biol. Fishes* **83**, 237–241 (2008).
- Ballew, N. G., Bacheler, N. M., Kellison, G. T. & Schueller, A. M. Invasive lionfish reduce native fish abundance on a regional scale. *Sci. Rep.* **6**, 32169 (2016).
- Albins, M. A. & Lyons, P. J. Invasive red lionfish *Pterois volitans* blow directed jets of water at prey fish. *Mar. Ecol. Prog. Ser.* **448**, 1–5 (2012).
- Morris, J. A. & Whitfield, P. E. Biology, ecology, control and management of the invasive Indo-Pacific lionfish: an updated integrated assessment. NOAA, Technical Report (2009).
- Schofield, P. J. Geographic extent and chronology of the invasion of non-native lionfish (*Pterois volitans* [Linnaeus 1758] and *P. miles* [Bennett 1828]) in the Western North Atlantic and Caribbean Sea. *Aquat. Invasions* **4**, 473–479 (2009).
- Schofield, P. J. Update on geographic spread of invasive lionfishes (*Pterois volitans* [Linnaeus 1758] and *P. miles* [Bennett 1828]) in the Western North Atlantic Ocean, Caribbean Sea and Gulf of Mexico. *Aquat. Invasions* **5**, 117–122 (2010).
- Semmens, B. X., Buhle, E. R., Salomon, A. K. & Pattengill-Semmens, C. V. A hotspot of non-native marine fishes: evidence for the aquarium trade as an invasion pathway. *Mar. Ecol. Prog. Ser.* **266**, 239–244 (2004).
- Hamner, R. M., Freshwater, D. W. & Whitfield, P. E. Mitochondrial cytochrome b analysis reveals two invasive lionfish species with strong founder effects in the western Atlantic. *J. Fish Biol.* **71**, 214–222 (2007).
- Betancur-R, R. *et al.* Reconstructing the lionfish invasion: insights into Greater Caribbean biogeography. *J. Biogeography* **38**, 1281–1293 (2011).
- Morris, J. A. & Akins, J. L. Feeding ecology of invasive lionfish (*Pterois volitans*) in the Bahamian archipelago. *Environ. Biol. Fishes* **86**, 389–398 (2009).
- Courtenay, W. R. Marine fish introductions in southeastern Florida. *AFS Introduced Fish Section Newsletter* **14**, 2–3 (1995).
- Whitfield, P. E. *et al.* Abundance estimates of the Indo-Pacific lionfish *Pterois volitans/miles* complex in the Western North Atlantic. *Biol. Invasions* **9**, 53–64 (2007).
- Freshwater, D. W. *et al.* Mitochondrial control region sequence analyses indicate dispersal from the US East Coast as the source of the invasive Indo-Pacific lionfish *Pterois volitans* in the Bahamas. *Mar. Biol.* **156**, 1213–1221 (2009).
- Johnson, J., Bird, C. E., Johnston, M. A., Fogg, A. Q. & Hogan, J. D. Regional genetic structure and genetic founder effects in the invasive lionfish: comparing the Gulf of Mexico, Caribbean and North Atlantic. *Mar. Biol.* **163**, 216 (2016).
- Ferreira, C. E. L. *et al.* First record of invasive lionfish (*Pterois volitans*) for the Brazilian Coast. *PLoS one* **10**, e0123002 (2015).
- Freshwater, D. W., Hamner, R. M., Parham, S. & Wilbur, A. Molecular evidence that the lionfishes *Pterois miles* and *Pterois volitans* are distinct species. *J. N. C. Acad. Sci.* **125**, 39–46 (2009).
- Guzmán-Méndez, I. A. *et al.* First genetically confirmed record of the invasive devil firefish *Pterois miles* (Bennett, 1828) in the Mexican Caribbean. *Biol. Invasions Rec.* **6**, 99–103 (2017).
- Wilcox, C. L., Motomura, H., Matsunuma, M. & Bowen, B. W. Phylogeography of lionfishes (*Pterois*) indicate taxonomic over splitting and hybrid origin of the invasive *Pterois volitans*. *J. Heredity*, <https://doi.org/10.1093/jhered/esx056> (2017).
- Toledo-Hernández, C. *et al.* Population ecology and genetics of the invasive lionfish in Puerto Rico. *Aquat. Invasions* **9**, 227–237 (2014).
- Butterfield, J. S. *et al.* Wide-ranging phylogeographic structure of invasive red lionfish in the Western Atlantic and Greater Caribbean. *Mar. Biol.* **162**, 773–781 (2015).
- Roman, J. & Darling, J. A. Paradox lost: genetic diversity and the success of aquatic invasions. *Trends Ecol. Evol.* **22**, 454–464 (2007).
- Holland, B. S. Genetics of marine bioinvasions. *Hydrobiologia* **420**, 63–71 (2000).
- Darling, J. A., Bagley, M. J., Roman, J. O. E., Tepolt, C. K. & Geller, J. B. Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Mol. Ecol.* **17**, 4992–5007 (2008).
- Lejeune, C. *et al.* High genetic diversity and absence of founder effects in a worldwide aquatic invader. *Sci. Rep.* **4**, 5808 (2014).
- Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
- Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
- Rius, M., Bourne, S., Hornsby, H. G. & Chapman, M. A. Applications of next-generation sequencing to the study of biological invasions. *Curr. Zool.* **61**, 488–504 (2015).
- Tepolt, C. K. & Palumbi, S. R. Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Mol. Ecol.* **24**, 4145–4158 (2015).
- Jeffery, N. W. *et al.* RAD sequencing reveals genomewide divergence between independent invasions of the European green crab (*Carcinus maenas*) in the Northwest Atlantic. *Ecol. Evol.* **7**, 2513–2524 (2017).
- Schultz, T. F., Fitzpatrick, C. K., Wilson Freshwater, D. & Morris, J. A. Characterization of 18 polymorphic microsatellite loci from invasive lionfish (*Pterois volitans* and *P. miles*). *Conserv. Genet. Resour.* **5**, 599–601 (2013).
- Larmuseau, M. H. D., Raeymaekers, J. A. M., Hellemans, B., Van Houdt, J. K. J. & Volckaert, F. A. M. Mito-nuclear discordance in the degree of population differentiation in a marine goby. *Heredity* **105**, 532–542 (2010).
- Pérez-Portela, R., Rius, M. & Villamor, A. Lineage splitting, secondary contacts and genetic admixture of a widely distributed marine invertebrate. *J. Biogeography* **44**, 446–460 (2017).
- Toews, D. P. L. & Brelsford, A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* **21**, 3907–3930 (2012).
- Portnoy, D. S. *et al.* Selection and sex-biased dispersal in a coastal shark: the influence of philopatry on adaptive variation. *Mol. Ecol.* **24**, 5877–5885 (2015).

42. Dahl, K. A., Patterson Iii, W. F. & Snyder, R. A. Experimental assessment of lionfish removals to mitigate reef fish community shifts on northern Gulf of Mexico artificial reefs. *Mar. Ecol. Prog. Ser.* **558**, 207–221 (2016).
43. Novak, S. J. The role of evolution in the invasion process. *PNAS* **104**, 3671–3672 (2007).
44. Keller, S. R. & Taylor, D. R. History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecol. Lett.* **11**, 852–866 (2008).
45. Sakai, A. K. *et al.* The population biology of invasive species. *Annu. Rev. Ecol. Evol. Syst.* **32**, 305–332 (2001).
46. Pérez-Portela, R., Turon, X. & Bishop, J. D. D. Bottlenecks and loss of genetic diversity: spatio-temporal patterns of genetic structure in an ascidian recently introduced in Europe. *Mar. Ecol. Prog. Ser.* **105**, 93–105 (2012).
47. Reem, E., Douek, J., Katzir, G. & Rinkevich, B. Long-term population genetic structure of an invasive urochordate: the ascidian *Botryllus schlosseri*. *Biol. Invasions* **15**, 225–241 (2013).
48. Pineda, M. C., Turon, X., Pérez-Portela, R. & López-Legentil, S. Stable populations in unstable habitats: temporal genetic structure of the introduced ascidian *Styela plicata* in North Carolina. *Mar. Biol.* **163**, 1–14 (2016).
49. Paz, G., Douek, J., Mo, C., Goren, M. & Rinkevich, B. Genetic structure of *Botryllus schlosseri* (Tunicata) populations from the Mediterranean coast of Israel. *Mar. Ecol. Prog. Ser.* **250**, 153–162 (2003).
50. Pineda, M. C., Lorente, B., Lopez-Legentil, S., Palacín, C. & Turon, X. Stochasticity in space. persistence in time: genetic heterogeneity in harbour populations of the introduced ascidian *Styela plicata*. *PeerJ* **4**, e2158 (2016).
51. Calazans, S. H., Walters, L. J., Fernandes, F. C., Ferreira, C. E. L. & Hoffman, E. A. Genetic structure provides insights into the geographic origins and temporal change in the invasive charru mussel (*Sururu*) in the southeastern United States. *PloS one* **12**, e0180619 (2017).
52. Silva, G., Lima, F. P., Martel, P. & Castilho, R. Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. *Proc. Biol. Sci.* **281**, 20141093 (2014).
53. Consuegra, S., John, E., Verspoor, E. & De Leaniz, C. G. Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genet. Sel. Evol.* **47**, 58 (2015).
54. Milano, I. *et al.* Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Mol. Ecol.* **23**, 118–135 (2014).
55. Carreras, C. *et al.* Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Sci. Rep.* **7**, 43417 (2017).
56. Xu, S. *et al.* Genomic evidence for local adaptation in the ovoviparous marine fish *Sebastes marmoratus* with a background of population homogeneity. *Sci. Rep.* **7**, 1562 (2017).
57. Lin, Y. *et al.* Genetic signatures of natural selection in a model invasive ascidian. *Sci. Rep.* **7**, 44080 (2017).
58. Johnston, M. W. & Purkis, S. J. Lionfish in the eastern Pacific: a cellular automaton approach to assessing invasion risk. *Biol. Invasions* **16**, 2681–2695 (2014).
59. Kletou, D., Hall-Spencer, J. M. & Kleitou, P. A lionfish (*Pterois miles*) invasion has begun in the Mediterranean Sea. *Mar. Biodiver. Rec.* **9**, 46 (2016).
60. Elshire, R. J. *et al.* A Robust. Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PloS one* **6**, e19379 (2011).
61. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
62. Schneider, S., Excoffier, L. & Laval, G. Arlequin (version 3.5. 1.2): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2010).
63. Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A. & Luikart, G. LOSITAN: a workbench to detect molecular adaptation based on a FST-outlier method. *BMC bioinformatics* **9**, 323 (2008).
64. Team. R. C. R: A language and environment for statistical computing (2013).
65. Silvestro, D. & Michalak, I. RaxmlGUI: a graphical front-end for RAXML. *Org. Divers. Evol.* **12**, 335–337 (2012).
66. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
67. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
68. Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).
69. Fedrizzi, N. *et al.* Population genetic structure of the dwarf seahorse (*Hippocampus zosterae*) in Florida. *PloS one* **10**, e0132308 (2015).
70. Broughton, R. E., Stewart, L. B. & Gold, J. R. Microsatellite variation suggests substantial gene flow between king mackerel (*Scomberomorus cavalla*) in the western Atlantic Ocean and Gulf of Mexico. *Fish. Res.* **54**, 305–316 (2002).
71. Narum, S. R. Beyond Bonferroni: less conservative analyses for conservation genetics. *Conserv. Genet.* **7**, 783–787 (2006).
72. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94 (2010).
73. Pante, E. & Simon-Bouhet, B. Marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS One* **8**, e73051 (2013).

Acknowledgements

The National Oceanic and Atmospheric Administration Center supported this research under the award NA11NOS4780045 to the University of Miami. We thank the University of Miami, and in particular Will Drennan and Gary Hitchcock in the Marine Science Program for supporting the undergraduate research which developed some of these data. We also thank Carlos Carreras, Sergio Taboada and Ana Riesgo for their assistant with some analyses.

Author Contributions

M.F.O. and D.L.C. designed this study, constructed GBS libraries and performed the first data filtering with A.B, B.C. and S.W. R.P.-P. computed the final data filtering, performed population genomic analyses, and wrote the manuscript. M.D., A.F., M.K.S., M.A.J., and F.C. provided lionfish samples. All authors contributed to improve the final version of this manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-23339-w>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018