Population genomics of a cephalopod species reflect oceanographic barriers and inbreeding patterns

Bernhard Bein[1], Françoise Lima[2], Henrique Lazzarotto[3], Luiz Rocha[3], Tatiana Leite[4], Sergio Lima[2], and Ricardo Pereira[1]

[1]LMU
[2]Universidade Federal do Rio Grande do Norte
[3]California Academy of Sciences
[4]Universidade Federal de Santa Catarina

March 8, 2022

Abstract

Temporal variation of effective population size and gene flow determine current patterns of genetic diversity within species, and hence the genetic variation upon which natural selection can act. Although such demographic processes are well understood in terrestrial organisms, they remain largely unknown in the ocean, where species diversity is still being described. Here, we present one of the first population genomic studies in a cephalopod, Octopus insularis, which is distributed in coastal and oceanic island habitats in the Atlantic Ocean, Mexican Gulf and the Caribbean Sea. Using genomic data, we identify the South Equatorial current as the main barrier to gene flow between southern and northern parts of the range, followed by discontinuities in the habitat associated with depth. We find that genetic diversity of insular populations significantly decreases after colonization from the continental shelf, also reflecting low habitat availability. Using demographic modelling, we find signatures of a stronger population expansion for coastal relative to insular populations, consistent with estimated increases in habitat availability since the Last Glacial Maximum. The direction of gene flow is coincident with unidirectional currents and bidirectional eddies between otherwise isolated populations, suggesting that dispersal through pelagic paralarvae is determinant for population connectivity. Together, our results show that oceanic currents and habitat breaks are determinant in the diversification of marine species, shaping standing genetic variability within populations. Moreover, our results show that insular populations are particularly vulnerable to current human exploitation and selective pressures, calling the revision of their protection status.

Title: Population genomics of a cephalopod species reflect oceanographic barriers and inbreeding patterns

Running title: Population genomics of *Octopus insularis*

Authors: Bernhard Bein[1]+, Françoise D. Lima[2]+[*], Henrique Lazzarotto[3], Luiz A. Rocha[3], Tatiana S. Leite[4], Sergio M. Q. Lima[2], Ricardo J. Pereira[1*]

1- Division of Evolutionary Biology, Faculty of Biology II, Ludwig-Maximilians-Universitat Munchen, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany.

2 - Laboratorio de Ictiologia Sistematica e Evolutiva, Departamento de Botanica E Zoologia, Universidade Federal do Rio Grande Do Norte, 59978–970 Lagoa Nova, Natal, RN, Brazil.

3 - Department of Ichthyology, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, CA, 90118, USA.

4 - Laboratorio de Metodos de Estudos Subaquaticos e Cefalopodes, Departamento de Ecologia e Zoologia, Universidade Federal de Santa Catarina, Florianopolis, SC, 88062–200, Brazil.

+- these authors contributed equally to this work

*** Corresponding authors:** Ricardo J. Pereira; **email:**ricardojn.pereira@gmail.com

Francoise D. Lima; **email:**francoisedl@yahoo.com.br

Abstract

Temporal variation of effective population size and gene flow determine current patterns of genetic diversity within species, and hence the genetic variation upon which natural selection can act. Although such demographic processes are well understood in terrestrial organisms, they remain largely unknown in the ocean, where species diversity is still being described. Here, we present one of the first population genomic studies in a cephalopod, *Octopus insularis* , which is distributed in coastal and oceanic island habitats in the Atlantic Ocean, Mexican Gulf and the Caribbean Sea. Using genomic data, we identify the South Equatorial current as the main barrier to gene flow between southern and northern parts of the range, followed by discontinuities in the habitat associated with depth. We find that genetic diversity of insular populations significantly decreases after colonization from the continental shelf, also reflecting low habitat availability. Using demographic modelling, we find signatures of a stronger population expansion for coastal relative to insular populations, consistent with estimated increases in habitat availability since the Last Glacial Maximum. The direction of gene flow is coincident with unidirectional currents and bidirectional eddies between otherwise isolated populations, suggesting that dispersal through pelagic paralarvae is determinant for population connectivity. Together, our results show that oceanic currents and habitat breaks are determinant in the diversification of marine species, shaping standing genetic variability within populations. Moreover, our results show that insular populations are particularly vulnerable to current human exploitation and selective pressures, calling the revision of their protection status.

Keywords : ddRADseq, gene flow, population structure, genetic diversity, conservation, *Octopus insularis*

Introduction

Genetic diversity segregating within populations is the main target of natural selection, and therefore is determinant for species adaptability and vulnerability (Frankham 1995; Lande 1995; DeWoody et al. 2021). Understanding how genetic diversity is conditioned on demographic responses to past environmental change has been a major task in evolutionary biology as it provides insights into how species might react to future environmental change (Hofmann & Sgro, 2011; Moritz & Agudo, 2013). Although demographic processes, such as temporal variation of effective population size and gene flow, are well understood in terrestrial systems, these processes remain obscured in marine systems.

Marine systems play a fundamental role on human consumption (Watson & Tidd, 2018; Pauly & Zeller, 2016, Lotze et al., 2006). Nevertheless, we are only starting to understand the number of species within marine environments (Mora et al., 2011), and the evolutionary processes underlying their diversification. Until recently, the presumed assumption was that marine systems are characterized by large populations and high dispersal (Riginos & Liggins, 2013; Sanford & Kelly, 2011; Palumbi, 1994, 1992), making them more resilient to environmental change, relative to terrestrial systems. However, demographic parameters in marine organisms and their dependence on oceanographic barriers and environmental change have remained largely untested.

Genetic and, more recently, genomic studies in marine species are shifting our understanding on diversification in the marine environment. Many morphologically conserved taxa that were taxonomized as single, widely distributed species have been recognized as multiple cryptic species (Amor et al., 2017; Duda, Kohn & Matheny, 2009), often composed by differentiated populations (Peijenburg et al., 2006), suggesting that strong barriers to gene flow can operate in marine taxa (Johannesson et al., 2018; Volk et al., 2020; Filatov et al., 2021). Genomic studies in economically important fish species revealed that effective population size varies between and within species (da Fonseca et al., 2021; Barry, Broquet & Gagnaire, 2022; Sodeland et

al., 2022), suggesting that genetic drift can be a strong driver of diversification in the ocean. Such studies also show that genetic barriers between populations of the same species are coincident with temperature and salinity gradients (Jorgensen et al., 2005; Guo, Li & Merila, 2016; Guo et al., 2015, Fisher et al., 2021), suggesting that local adaptation can further drive divergence in the presence of gene flow. Less is known about invertebrate oceanic species, particularly those with sedentary behavior, where dispersal is restricted to early developmental stages.

The cephalopod *Octopus insularis* Leite & Haimovici, 2008 was recognized as a different species from other more widespread octopuses' species, including *O. vulagris* , which is heavily targeted by fisheries worldwide. It inhabits tropical shallow reefs, from the interdidal to depths of 40 m (Leite et al., 2008; Leite et al., 2009a; Bouth et al., 2011), where it acts as an opportunistic predator (Leite et al, 2009b; Leite et al., 2016; da Silva et al., 2018). It is distributed along the eastern continental shelf of the American continent, from the Gulf of Mexico (Flores-Valle et al. 2018) to the south of Brazil and in several oceanic islands in the south Atlantic (Leite & Haimovici, 2008; Leite et al., 2009a). This cephalopod has a short life cycle (one generation per year; Lima et al., 2017), high fecundity (˜ 95,000 eggs; Lima et al., 2014), planktonic paralarvae (Lima et al., 2017), and adults display a sedentary life style (Lima et al., 2017). Studies on environmental niche modelling (Lima et al., 2020) found that coastal habitats are highly connected by suitable habitat, while most oceanic islands are disconnected from coastal habitats due to high depth. The exception is the oceanic archipelago of Rocas' Atoll and Fernando de Noronha, which are connected to the Northwestern coast of Brazil by shallow seamounts. By projecting the current ecological niche to the Last Glacial Maximum (some 24,000 kya), Lima et al. (2020) suggested that habitat suitability increased strongly along the coast, while no major changes were seen in oceanic islands, but it remains unclear whether such environmental changes shaped current patterns of genetic diversity. Recent genetic work using a fragment of the mitochondrial COI gene from populations collected throughout the known species range (Lima et al., under review) showed four haplotypic groups; two most divergent haplogroups separated by the South Equatorial Current (Fig. 1A), and two less divergent haplogroups separated by high depth between the coast and oceanic islands. These results raise the hypothesis that sea currents and habitat availability associated to seamounts may drive the diversification of *O. insularis* . Although valuable, molecular studies relying on mitochondrial markers alone are prone to reflect stochastic processes affecting a single marker, by selective constrains of the mitochondria, and by demographic processes specific to females (Galtier et al., 2009). Understanding how demographic history, in particular how temporal changes of effective population size and gene flow, resulted in extant patterns of genetic diversity within populations requires the sampling of independent nuclear markers and the use of coalescent-based methods (Sousa & Hey, 2013).

Here, using thousands of SNPs sampled randomly across the genome of *O. insularis* , and sampling populations of this species throughout its known range, including its coastal and disjunct oceanic habitats (Fig. 1A), we perform one of the first population genomic studies of a cephalopod species (but see Timm et al., 2020). First, we identify population structure and associated oceanographic barriers to gene flow along the species range. Second, we infer phylogeographic patterns of island colonization and associated changes in genetic diversity within populations. Lastly, we understand how temporal changes in effective population size and gene flow are conditioned by changes in habitat suitability and main oceanic currents. Our findings provide general insights on the drivers of diversification in marine species with low dispersal abilities, and provide recommendations for conservation and sustainable fisheries of this ecologically and commercially relevant species.

Material and Methods

Sampling of specimens

A total of 71 individuals of*Octopus insularis* were sampled from 11 localities encompassing most of its known species range (Table S1, Fig. 1A). Specimens were collected during snorkeling, scuba diving, or were purchased in fish markets, when the exact location of capture was known. Specimens were then stored in

96% ethanol at room temperature and deposited in the mollusks collection of the Federal University of Rio Grande do Norte.

Library preparation and sequencing

We extracted DNA from muscle tissue of stored specimens using DNeasy Tissue kits (Qiagen), following the manufacturer's protocol. We prepared double digest restriction site associated DNA sequencing (ddRADSeq) libraries following the protocol of Peterson et al. (2012), as adapted by Gaither et al. (2015). In short, we used the restriction endonucleases SphI and MluCl (New England Biolabs), following DNA purification with Dynabeads M-270 Streptavidin (Life Technologies), and ligation of the universal P2 adaptors and 24 different P1 adaptors containing individual barcodes of 5 bp, which differed from one another by at least 3 bp. Prior to pooling, we cleaned the DNA libraries using magnetic beads, and pooled individuals in three groups with unique Illumina indices. We then size-selected pooled DNA to recover fragments between 376 and 450 bp using a Pippin Prep (Sage Science). Libraries were quantified using a High Sensitivity DNA Kit on a 2100 Bioanalyzer (Agilent Technologies) and sequenced by QB3 Genomics at UC Berkeley on an Illumina HiSeq 2000, producing 100 bp single end reads. We obtained 260,002,245 raw reads, which after demultiplexing corresponded to an average of 3,298,408 reads per individual (sd 1,900,730 reads/ind; Table S2).

Assembly of RAD-loci and filtering

Since there is no reference genome for *Octopus insularis* , we assembled de novo RAD-loci using ipyrad (v.0.9.50) (Eaton & Overcast, 2020). To assess the robustness of the de novo assembly we considered two sequence similarity thresholds (0.9 and 0.95) that are often used for studies within species (Amor, Johnson & James, 2020). We only considered reads with length >70 bp and otherwise kept the default settings of ipyrad. For each assembly we estimated number of all loci assembled across individuals, number of homologous loci after filtering, number of single nucleotide polymorphisms (SNPs), percentage of missing data, and number of parsimoniously informative SNPs.

Because the two assemblies were similar in their summary statistics (Table S3), we chose a sequence similarity threshold of 0.9 in order to avoid over splitting of loci. Two samples (SPS3 and FN5) that showed overall low read number (< 200,000) and low number of recovered loci (< 200) were discarded to reduce the amount of missing data (Table S2), leaving 69 individuals from 11 locations for our final assembly.

We exported SNPs in the variant call format (vcf), resulting in 572,012 SNPs and 299,304 homologous loci. We then generated three datasets for downstream analyses (Table S4) with the program vcftools v.0.1.12b (Danecek et al., 2011): 1) the original dataset without further filtering (herein "full"), 2) a more permissive dataset allowing for loci with a maximum of 40% missing data across all 69 individuals (–max-missingness 0.6; herein "69inds_40MD"), and 3) a more stringent dataset allowing for a maximum of 20% missing data across individuals after removing the 5 individuals with the lowest number of loci (–max-missingness 0.8 and –remove for individuals BA18, RN2A, STH2, OIC1, RN13; herein "64inds_20MD"). All of these data sets contain SNPs linked within the same ddRAD locus, as physical linkage is accounted for by the most methods (Table S4). To test if the percentage of missing data reflect the input of reads per individual, rather than a population-specific divergence, we fit a linear regression model to the comparison between the log of raw reads and the log of missing data per individual, using the "69inds_40MD" dataset, and assessed statistical significance.

Population structure

We inferred the most likely number of evolutionarily independent lineages within *O. insularis* using complementary methods that rely on different assumptions, using the permissive and stringent datasets ("69inds_40MD", "64inds_20MD").

First, we carried out a Principal Component Analysis developed for low coverage sequencing data (EMU

PCA, Meisner et al., 2021). This imputation method is suitable for genetic dataset with extensive missing data, which typically occurs in ddRAD datasets, and does not depend on previously defined groups and thus cannot over-estimate population structure. We converted both datasets to bed format using Plink v1.90b6.22 (Purcell et al., 2007) and ran EMU PCA assuming eleven eigenvalues (the number of sampling localities), filtering out alleles with a minor allele frequency < 0.001 (-f 0.001).

Second, we inferred the probability of assignment of each individual to a given number of K ancestral clusters, using TESS3R (Caye et al., 2016), also considering the geographic location of the sampled individuals. We converted the vcf-files to the lfmm input format, using the vcf2lfmm function of the R-package LEA (Frichot & Francois., 2015). We conducted 20 independent runs, assuming K between one and eleven and allowing for maximum 1,000 iterations.

Lastly, we estimated the co-ancestry matrix between every pair of individuals, using fineRADstructure (Malinsky et al., 2018). In contrast with the previous method, this method uses linkage information within the same ddRAD locus, without any prior assumption based on sampling location. In addition to the permissive and stringent datasets, we also converted vcf-files for the "full" datasets to finerad input files with RADpainter (RADpainter hapsfromVCF), and inferred the co-ancestry matrix with the default settings.

Phylogenetic analysis

We inferred the phylogenetic relationship between individuals, using the permissive and stringent datasets ("69inds_40MD", "64inds_20MD").

First, we estimated a maximum likelihood (ML) tree, concatenating all loci. We used fasta files to estimate the most likely evolution model in ModelFinder (Kalyaanamoorthy et al., 2017), and computed a ML tree in IQ-Tree v.2.1.2 (Minh et al., 2020). To evaluate the support values of the inferred topology, we used the ultrafast bootstrap approach (Hoang et al., 2018) with 1,000 replicates. We ran the analysis five times, chose the tree with the highest likelihood, and visualized it with ggtree v. 2.4.2. (Yu et al., 2017).

Second, we inferred a coalescent tree that incorporates incomplete lineage sorting, using the SVDquartets (Chifman & Kubatko, 2014) and gQMC (Avni, Cohen & Snir, 2015) algorithms implemented in tetrad, within ipyrad. This approach uses one randomly chosen SNP per locus for inferring the topology between any combination of four individuals and joins them into a super tree. In addition to the permissive and stringent datasets, we also ran tetrad on the "full" dataset, using 100 non-parametric bootstraps, and constructed a majority rule consensus tree.

Population summary statistics

To assess if coastal and island lineages of *O. insularis* differ in their levels of genetic diversity, we considered both the permissive and stringent datasets ("69inds_40MD", "64inds_20MD").

First, we measured the following individual-level statistics: 1) expected (H_e) and observed (H_o) heterozygosity, representing the average and observed number of polymorphic sites in each individual across ddRAD loci; 2) inbreeding coefficient (F_{IT}), representing the heterozygosity of an individual relative to that observed in the total sampling; and 3) nucleotide difference per SNP (π*), which represents the average number of nucleotide differences between the two haplotypes from one individual. We converted the datasets into fasta format using PGDSpider (Lischer & Excoffier, 2012), calculated H_e, H_o and F_{IT} in vcftools (Danecek et al., 2011), and π* in DnaSP6 (Rozas et al., 2017). We tested if there was a significant correlation between the percentage of missing data and F_{IT} or π* in each individual, using a linear regression model with R (R Core Team, 2017). To test if population-wide π* and F_{IT} were different between lineages, we used a pairwise Student's t-test with Bonferroni-Holm p-value correction in R (R Core Team, 2017).

Second, to quantify how genetic diversity is partitioned between lineages, we calculated the pairwise fixation index (F_{ST}). We converted the vcf-files to the Arlequin-input format using PGDSpider (Lischer & Excoffier, 2012), and used Arlequin ver. 3.5.2.2 (Excoffier & Lischer, 2010) to compute pairwise $F_{ST,}$ considering the six lineages determined previously. We performed 5,000 boostraps to estimate significance from the null

hypothesis of no population structure. We assessed how much genetic variation is explained by four levels of population structure (between 2 major regions, within 2 major regions, between 6 minor regions and between individuals), by performing a locus-by-locus Analysis of Molecular Variance (AMOVA, Excoffier et al., 1992). We used 10,000 permutations and excluded loci missing in at least one of the lineages.

Lastly, to understand if the genetic diversity observed within each lineage is consistent with changes in effective population size, we computed Tajima's D using DnaSP (Rozas et al., 2017). We expect lineages that experienced a recent range expansion to have negative values of Tajima's D, significantly different from zero (alpha= 0.05).

Demographic analysis

To estimate the past demographic history of *O. insularis* , we applied demographic modelling with diffusion approximation methods as implemented in the program δaδi (Gutenkunst et al. 2009). First, to assess changes in effective population size within each lineage, we fit one-population models for the lineages containing more than 3 individuals (i.e. N-Coastal, S-Coastal, S-Oceanic). We compared four demographic models that differ in the number of population size changes (none, one, two) and in the mode of change (none, discrete, exponential) (details in Table S13). Second, to assess gene flow between lineages, we performed two-population models for the two pairs of adjacent lineages with a larger number of individuals (i.e. N-Coastal vs S-Coastal and S-Oceanic vs S-Coastal), which provide some power of estimating demographic parameters. We fitted 15 different demographic models (details in Table S14, following Portik et al., 2017) that incorporate different scenarios of presence and directionality of migration (m), time of a split between lineages (T), and change of effective population size (nu).

For both classes of demographic modeling, in order to retain a maximum number of segregating sites, we filtered the raw vcf output file ("full") to only include individuals from the relevant lineages, as in the more stringent dataset ("64ind_20MD dataset"). We excluded loci with more than 40 % missing data, and retained the first SNP per ddRAD locus, minimizing physical linkage. We converted the filtered vcf tiles to the dadi file format, projected down to the number of variants maximizing the number of segregating sites present (Table S15), and calculated the observed one- or two-dimensional folded site-frequency spectra (SFS). We optimized these models using the pipeline developed by Portik et al. (2017) with the following settings (one population modelling/ two population modelling): three/four rounds of optimization with 10, 20, 30/ 15, 10, 5, 5 replicates in each round; with maximum iterations of 5, 10, 50 / 15, 15, 25, 50 per replicate in each round; and a parameter fold of 3, 2, 1 / 3, 2, 1, 1 using the default Nelder-Mead method. To guarantee model convergence, we ran each optimization five times. We selected the model with the lowest Akaike information criterion (AIC) score (Akaike 1974), which penalizes the likelihood by the number of parameters, and estimated dAIC and wAIC (Burnham & Anderson, 2002). For the two-population models, we carried out a goodness-of-fit test of the most likely demgraphic model (Barrat et al., 2018) to test general model fit of our inferred parameters. The test is considered passed if the empirical log-likelihood value falls within the distribution of log-likelihood values fitted to simulated SFS.

Results

Assembly of RAD-loci and filtering

Assemblies with alternative similarity thresholds have equivalent summary statistics (Table S3; supplementary Fig. S1), and thus we chose a similarity threshold of 0.9 to avoid over splitting of loci. Assuming this threshold, the "full" dataset consisted of 299,304 homologous loci, with 572,012 SNPs, with 66,34% of missing data (Table S3); the more permissive "69inds_40MD" dataset consisted of 34,455 loci with 99,915 SNPs, and the more stringent "64inds_20MD" consisted of 25,702 loci with 72,816 SNPs. We found a significant correlation between the number of raw reads and percentage of missing data per individual (adjusted R^2: 0.9128, p-value: 2.2 x 10-16).

Population structure

In the EMU-PCA, for the "69inds_40MD" dataset, PC1 and 2 explain 19 and 14% of the genetic variance, respectively. Individuals are clustered into six groups that recapitulate their geographic location (Fig. 1B, Fig. S2A-B): 1) South-coastal (including AL and BA), 2) South-oceanic (the oceanic islands TM), 3) North-SPS (the oceanic islands SPS), 4) North-Caribbean (the coastal OIC), 5) North-Atlantic (the oceanic islands STH, ASC), and 6) North-coastal (including CE, RN, FN, and AR). Results for the more stringent dataset were qualitatively similar (Fig. S2A).

In the TESS3R analysis, using the more permissive dataset (Fig. 2A-B, Fig. S3), the first population split separates samples from the southern coast of Brazil from all the others, with Bahia (BA) and Alagoas (AL) bordering these clusters and containing individuals with admixed ancestry. At K=3, the northern oceanic islands (São Pedro / São Paolo SPS) form a cluster without showing shared ancestry. At K = 4, the southern populations are divided into a coastal (AL and BA) and an oceanic island cluster (TM), with the individuals from Bahia (BA) showing admixture between these two clusters. At K=5, the Caribbean individuals are assigned to a new cluster, showing admixture with the northern-coastal cluster. At K = 6, individuals from the most remote oceanic islands (ASC and STH) form their own unmixed cluster. The exception is the individual from Saint Helena, which contains the largest amount of missing data and therefore is assigned to every cluster with some probability. Assuming further number of ancestral clusters, we find that previous clusters split equally into two clusters, as thus we refrain from interpreting these biologically. Using the more stringent dataset, we obtained overall concordant results (Fig. S4).

For the fineRADstructure analysis (Fig. 2C) with the "full" dataset, the highest degree of co-ancestry was observed in individuals from São Pedro / São Paulo (N-SPS), followed by individuals from Ascension Island that cluster together with Saint Helena (N-Atlantic), and individuals from the Caribbean (N-Carribean). These three groups are nested within a northern group that also includes individuals from N-Coastal localities (Ceara, Rio Grande do Norte) and nearby islands (Fernando de Noronha, and Rocas Atoll). All individuals from the archipelago of Trindade and Martim Vaz (S-Oceanic) have a relatively high co-ancestry and are nested in a southern group encompassing individuals from the S-Coastal localities (Alagoas and Bahia), reflecting hierarchical population structure also in the southern group. Results from runs with the more permissive ("69inds_40MD") and more stringent ("64inds_20MD") datasets reflected the same relative levels of co-ancestry and hierarchical population structure (Fig. S5, S6).

Phylogenetic analysis

The ML trees estimated with the most likely model (TVM+F+R3) were largely consistent between datasets. Individuals belonging to the island lineages (N-Atlantic, N-SPS, and S-Oceanic) and to the northernmost coastal lineage (N-Caribbean) form well supported (ultrafast bootstrap > 90) monophyletic clades (Fig. 3A). In contrast, individuals from the geographically widespread coastal lineages (N-Costal, and S-Coastal) form paraphyletic clades, with the respective oceanic clades nested within. The N-Caribbean lineage is the sister of the N-Coastal lineage. The clades leading to the two northern oceanic islands (N-Atlantic and N-SPS) show remarkably long branches, reflecting high genetic divergence. While in the most restrictive dataset these two clusters are sister taxa (Fig. S7, S9), this relationship is not supported in the most permissive dataset (Fig. 3A, S8).

The topology of the coalescent trees was largely congruent with that from the ML trees. Differences include the monophyly of the S-Coastal lineage with high bootstrap support (>90), the paraphyly of the N-Caribbean lineage and the monophyly of the N-Coastal lineage with bootstrap support of >50 (Fig. 3B). Comparing topologies based on the three datasets, we observe that the N-SPS and N-Atlantic lineages form a monophyletic clade in the more permissive dataset (Fig. 3B, Fig. S10), while this monophyly is not recovered in the "full" and more stringent datasets (Fig. S11, S12). Similar to our ML trees, the branch length of lineages from the northern oceanic islands (N-Atlantic and N-SPS) are substantially longer than those of coastal lineages (Fig. S10).

Population summary statistics

Using both datasets, we find that π^* and observed (H_o) heterozygosity are lower in the oceanic islands relative to the coastal lineages (Tables S7, S8). Conversely, the inbreeding coefficient F_{IT} is lower in the coastal lineages and highest in all oceanic islands, with N-SPS showing 4 times higher inbreeding relative to the coast (Fig. 4, Fig. S13, Tables S7, S8). For both π^* and F_{IT}, all comparisons between coastal and oceanic lineages yielded significant p-values (Tables S5, S6), except the comparisons between S-Oceanic and N-Caribbean (F_{IT}: 0.111, π^*: 0.182). We did not find a significant correlation between the percentage of missing data and individual measures of diversity (p-values: 0.324 for F_{IT}, 0.145 for π^*), confirming that the observed patterns are not driven by missing data.

For both datasets, F_{ST} was significant between all six lineages (Tables S9, S10), except for pairwise comparisons where lineages containing three or less individuals (N-SPS, N-Caribbean, N-Atlantic), reflecting the limited sampling. F_{ST} values ranged from 0.056 observed between N-Coastal and N-Atlantic, to 0.703 observed between the two oceanic islands (N-SPS and N-Atlantic) of the northern region. The AMOVA of the more permissive dataset (Table S11) showed that variation was highest within individuals (67.88 %), followed by variation among the lineages within groups (14.62 %), variation explained by the broader regions (North-South, 8.93 %), and finally variation within lineages (8.56%). The AMOVA of the more stringent dataset was qualitatively similar (Table S12).

For both datasets, Tajima's D (Tables S7, S8) was negative for the three coastal lineages, however it was only significantly different from demographic stability in the N-Coastal lineage. The three lineages of oceanic islands had Tajima's D values closer to zero.

Demographic history

For the single population modelling, in all cases the neutral model showed the highest AIC values (Table S16), showing that any model accounting for size changes in effective population size (Ne) explains the observed SFS significantly better. For the S-Coastal and S-Oceanic lineages, the simpler "two_epoch" model showed the lowest AIC score (Fig. S14) and similar residuals to other expansion models (Fig. S15). Assuming this model, parameter estimates indicate an increase in Ne of 1.34 (S-Oceanic) and 4.17 (S-Coastal) times, relative to the Ne of the ancestral population "Na" (Table S16, Fig. S16). For the N-Coastal lineage, AIC scores favored the "three_epoch" model incorporating an additional change of Ne. Here, the population first increased to 3.16 times and subsequently to 11.89 times the Na (Table S16, Fig. S16).

For two population modelling, in both cases we obtained the highest AIC values for the neutral model of no population split, rejecting this scenario (Table S17, Fig. S17). In the case of N-Coastal vs S-Coastal, the model integrating a period without gene flow, followed by secondary contact and asymmetric migration (sec_contact_asym_mig_size) shows the lowest AIC value (Fig. S17). Assuming this model, we find that after the initial population split, the S-Coastal population shrinks to 0.0124 times the Na, while the N-Coastal population increases to 2.33 times of the Na. After secondary contact, both effective population sizes increase to 1.3 and 8.46 times of the Na, respectively (Fig. 4A, Table S17). After secondary contact, migration rates are highly asymmetric, being 2.6401 from N-Coastal to S-Coastal and 0.0628 in the opposite direction. For S-Coastal vs S-Oceanic, the model incorporating a population split followed by continuous asymmetric migration (asym_mig, Table S14, S17, Fig S17) showed the lowest AIC values. Assuming this model, S-Coastal increases to an effective population size of 14.1 times the Na after the split, while S-Oceanic increases to 1.19 times the Na. Migration rates are moderately asymmetric at 1.01 from S-Oceanic to S-Coastal and 1.52 in the other direction (Fig. 4B). Residuals for the of the modelled JSFS are higher in the Coastal vs S-Coastal comparison, relative to the S-Coastal vs S-Oceanic comparison (Fig. 4A). Yet, both models passed the goodness-of-fit test (Fig. S18), showing that these are idealized models are fair representations of the evolutionary history acting in these populations.

Discussion

Oceanic currents and depth cause cryptic diversification in *O. insularis*

Whereas diversification processes in terrestrial habitats have often been linked to geographic barriers inhibiting gene flow, oceanic barriers driving diversification in marine organisms are less well understood, particularly in cephalopods. Our genomic methods in *Octopus insularis* recovered 299,304 ddRAD loci, containing 572,012 linked SNPs, offering the first insights into the genome-wide patterns of genetic differentiation in this species, and in the oceanographic barriers associated with it.

At a deeper phylogenetic scale, we find that populations of *O. insularis* are structured into two widely distributed northern and southern groups (Fig. 1B, 2B-C, Fig. 3). This deeper division between the northern and southern groups coincides with the South Equatorial Current (SEC) (Fig. 1A). This finding is in agreement with the distribution of the two major haplogroups in mitochondrial DNA (Lima et al., under revision). Given that *O insularis* produced up to ˜ 95,000 eggs (Lima et al., 2014) and that planktonic paralarvae disperse with oceanic currents (Lima et al., 2017), it is perhaps not surprising to find such a strong role of the SEC in the genetic divergence of this species. This current-mediated North-South division has been reported for other co-distributed species showing a pelagic propagule or larval dispersal, such as corals (Peluso et al., 2018) and mangrove trees (Francisco et al., 2018). Accordingly, a recent review of marine barriers to gene flow, Martins et al. (under review) found the SEC to compose the largest value of phylogeographic concordance among Brazilian coastal organisms, suggesting that this current imposes a major biogeographic barrier across Atlantic species. Studies in other species of octopuses (*Octopus vulgaris* , Melis et al. 2018; *Macroctopus maorum* , Higgins et al. 2013) have shown that population structure coincides with oceanic currents in the Mediterranean Sea and the southern Pacific, suggesting that oceanic currents might pose a strong oceanographic barrier for sedentary species with pelagic paralarvae.

At a shallower phylogenetic scale, we find six evolutionarily independent lineages (Fig. 2). The broader northern group is substructured into four lineages: a more widespread coastal lineage encompassing four localities near the coast of Brazil (N-Coastal), a Caribbean lineage (N-Caribbean), a first oceanic lineage encompassing the archipelago of São Pedro and São Paulo (N-SPS), and a second oceanic lineage encompassing the two islands off the coast of Africa (N-Atlantic), Saint Helena and Ascension. In turn, the broader southern group is substructured into two lineages: a more widespread coastal lineage (S-Coastal), and an oceanic lineage representing the archipelago Trindade and Martim Vaz (S-Oceanic). Importantly, these six lineages explain almost double of the genetic variation explained by the two broader groups (AMOVA, Table S11), underscoring their evolutionary significance. This finer population structure is coincident with previous ecological models of habitat suitability for this species that are largely driven by ocean depth (Lima et al., 2020). Oceanic lineages are coincident with volcanic islands that provide suitable habitat highly isolated from the continuous habitat along the coastal shelf of the American continent, with the exception of the S-Coastal lineage, which is connected to the southern coast of Brazil through a chain of seamounts. Genetic differentiation between populations from oceanic islands and the Brazilian coast have also been reported for dolphins (Oliveira et al., 2019), rockpool blennies (Neves et al., 2016), posobranch gastropods (Barroso et al., 2016) and corals (Peluso et al., 2018), confirming that such breaks in habitat constitute important drivers of divergence for coastal taxa with very different dispersal rates.

Island colonization is associated with reduction of genetic diversity

A well-supported observation in terrestrial island biogeography is the lower genetic diversity of recently colonized island populations compared to their mainland source populations (White & Searle, 2007; Boessenkool et al., 2007). Decreased diversity in insular populations leads to increasing inbreeding, decreasing their adaptive capability (Spielman, Brook & Frankham, 2004), and further exposing these to a higher risk of extinction (Frankham, 1997). Yet, it is less clear how genetic diversity changes during island colonization in marine systems, such as in *O. insularis.*

Our phylogenetic analyses shed some light on the evolutionary relationships between the six lineages (Fig.

3). We show that within the southern group, the S-Oceanic clade is nested within the broadly distributed S-Coastal clade, consistent with a colonization of this oceanic archipelago from the southern coast of Brazil. This direction of colonization is concordant with phylogenetic studies in co-distributed fish species (Macieira et al., 2015; Simon et al., 2021; Pinheiro et al., 2017), which suggest that the colonization of the oceanic archipelago Trindade and Martim Vaz occurred during the LGM (Pinheiro et al., 2017), when seamounts now submerged likely formed an ecological corridor for coastal taxa (Mazzei et al., 2021; Simon et al., 2021). Within the northern group, while the coalescent tree is consistent with a single colonization of the islands from the coast (i.e. N-SPS and N-Atlantic are sister clades; Fig 3B), the ML tree is consistent with two independent colonization events (i.e. N-SPS and N-Atlantic are not sister clades; Fig. 3A). Yet, the ML analysis using the more stringent dataset is again more consistent with a single colonization event (Fig. S7, S9), corroborating our findings with the coalescent methods, which are more appropriate to study recent radiations due to the expected large amounts of incomplete lineage sorting (Giarla & Esselstyn, 2015). Given our finding that these two northern oceanic lineages show the highest genetic differentiation observed in the species ($F_{ST:}$ 0.703) and show the longest phylogenetic branches (Fig. 3), more samples would be needed in these islands in order to establish if colonization occurred once or twice within the northern group. Nevertheless, our results conclusively demonstrate that the island populations represent at least two independent colorization events from the coast – one from the South-Coast and one or two from the North-Coast – allowing us to test evolutionary consequences of island colonization for patterns of genetic diversity.

Regarding genetic diversity segregating within the six lineages of *O. insularis* , we observe that values of nucleotide diversity (π^*) are significantly lower in the three oceanic lineages (N-SPS, N-Atlantic, and S-Oceanic) compared to their respective source of colonization (N-Coastal or S-Coastal; Fig. 4B, Fig. S13, Table S5). Conversely, individual inbreeding coefficients (F_{IT}) are significantly higher in the oceanic lineages (Fig. 4A, Table S6), reflecting the same pattern observed in the co-ancestry matrix (Fig. 2A). These results suggest that island colonization is associated with strong decreases in genetic diversity, both due to the demographic founder effect reducing effective population size, and to the smaller availability of suitable habitat in the islands relative to the coast (Lima et al., 2020); e.g. the available habitat up to 50 m depth in N-SPS is of 1.1 km^2 (Ávila et al., 2018). In accordance with our results, similar decreases of genetic variability in oceanic islands have been reported for several reef fish species (Pinheiro et al., 2017). This suggest that, although marine species of low dispersal as *O. insularis* have been able to successfully colonize novel isolated habitats, such as the volcanic islands of the Atlantic, island colonization has led to a significant decrease of genetic variability in the insular range of the species, mirroring what was established in terrestrial systems (White & Searle, 2007; Boessenkool et al., 2007).

Demographic history is conditioned on oceanic currents and environmental change

Historical changes in the demographic parameters such as effective population size and gene flow are determinant for current patterns genetic diversity within species (Excoffier, Foll & Petit, 2009; Hewitt et al., 2000, Ellegren & Galtier, 2016). Although demographic changes driven by glacial cycles are well known to condition genetic diversity in terrestrial systems (Canestrelli, Sacco & Nascetti, 2012; Cheddadi et al., 2006), it is less clear how such global changes have affected marine systems that are highly dependent on shallow habitats, such as *O. insularis* .

Our demographic models per population detect contrasting demographic histories for coastal and oceanic lineages. For all the three lineages tested, we conclusively reject a scenario of a stable population size in favor scenarios showing one (for S-Oceanic and S-Coastal lineages) or two (for N-Coastal) instantaneous changes of effective population size. Consistent with relative values of Tajima's D (Table S7, S8), we estimate that demographic expansions are larger for the two coastal lineages (of 11.96 and 4.17-fold for the N-Coastal and S-Coastal lineages, respectively) than for the oceanic lineage (1.38 fold for the S-Oceanic; Fig. S16). The magnitude of these expansions is consistent with studies of environmental niche modelling that suggested a strong increase of habitat suitability since the LGM along the American coast for this species (Lima et al., 2020), associated with the increase of sea level and the consequent exposure of the continental shelf and increase of sea temperature (Ludt & Rocha, 2015). In contrast, in the oceanic islands covered in this study,

increases in sea level likely exposed less coastal habitat (Ávila et al., 2018), explaining the more modest expansion of population size.

Our demographic models for two populations reflect similar changes in effective population size, but reveal patterns of genetic connectivity between lineages. For the populations along the coast (N-Coastal vs S-Coastal comparison), the best model reflects a population split without gene flow, followed by a population expansion with gene flow (Fig 4A). Gene flow is strongly asymmetric, with migration from the N-Coastal to the S-Coastal lineage being 42-fold larger than in the opposite direction. This observation is coincident with the Brazil Current (BC), which moves southwards from the range of the N-Coastal lineage (Fig. 1A). This current has been associated to unidirectional gene flow in rockpool blennies (Neves et al., 2016), suggesting that it may be an important driver of diversification of marine organisms with pelagic dispersal associated to reef habitats.

For the S-Oceanic/S-Coastal comparison (Fig. 4B), the best model consists of a population split with instantaneous size change followed by constant migration. Although migration rates are asymmetric, migration from the coast to the island is only 1.5-fold larger than in the opposite direction. This mildly asymmetric gene flow is consistent with cyclonic eddies that lay north and south of the sea mountain chain connecting the coast with Trindade and Martim Vaz (Arruda et al., 2013, Mill et al., 2015), likely facilitating gene flow in both directions (Pinheiro et al., 2015). Together, these findings suggest that oceanic current not only work as major oceanographic barriers discussed above, but also mediate genetic connectivity between genetically isolated populations.

Implications for conservation biology

Our findings on the evolutionary processes shaping the diversification of *O. insularis* have direct implications for the conservation of this species. As the taxonomic status of the octopus species mainly targeted by American fisheries has been resolved only recently, little is known about the catch compositions of octopus fisheries in this continent. Annual catches of *O. americanus* Monfort, 1802 are around 15,000 t in Mexico and 2,000 t in Brazil (Jereb et al. 2014). Although *O. insularis* has a distribution similar to *O. americanus* , the reported catch of *O. insularis* species are much lower (around 500 t; Haimovici et al., 2014). Considering that *O. insularis* was described only in 2008, and since then new studies have identified a greater distribution range than originally thought, the conservation status of its stocks remains unclear, mainly due to the common misidentification between both species. With the reduction of pelagic fish stocks due to overexploitation, there is a tendency to target fishing towards cephalopods (Rosa et al. 2020), which was indeed observed by Lopes et al. (2020) on *O. insularis* fisheries over the last 10 years in Northeast Brazil. Thus, identifying different stocks within this species is crucial to propose management strategies that avoid overexploitation of this important fishery resource.

Our findings on population structure imply that management plans for *O. insularis* must consider at least six evolutionarily independent units with confined distributions (Fig. 2B). Moreover, our finding of significantly higher levels of inbreeding and lower genetic diversity in the island lineages relative to the coastal ones (Fig. 4) imply that three island lineages deserve a higher protection status relative to the three coastal ones. Currently, the insular lineages receive some protection status (Marine Protection Atlas, accessed on 24.01.2022): N-SPS is part of a Brazilian Marine Protected Areas (MPA of São Pedro/ São Paulo), S-Oceanic is a protected Brazilian military area, and N-Atlantic is partially (Marine Protection Zone of Saint Helena) or fully protected (MPA of Ascension). Apart from Rocas Atoll, which is a fully protected area, the other archipelagos are only partial no-take areas (Giglio et al., 2018), being exploited mainly by artisanal fishing in part of the territory and potentially exposing the entire stock of the local lineage of *O. insularis* . In other species, decreased genetic diversity of natural populations has been associated with breakdown of multiple life history traits (Mills et al., 2012; Reed & Frankham, 2003; DeWoody et al., 2021), a hypothesis that needs to be evaluated in this system by future studies. Yet, given that the adaptive potential of natural populations directly depends on the amount of genetic diversity segregating within populations, our finding of low genetic connectivity and high inbreeding in N-SPS and N-Atlantic imply that these populations deserve a higher

protection status. We thus advocate that the entirety of these protected areas be classified as areas of no-take as new conservation measures of *O. insularis*.Restricting fishery activities to the more genetically diverse coastal lineages, educating local stakeholders on the morphological differences between sympatric octopus species, and establishing set quotas for landings of *O. insularis* will favor the sustainable management of this economically and ecologically important species.

Acknowledgements

This work was supported by the Brazilian National Research Council (CNPq 481492/2013- 9) and Coordination for the Improvement of Higher Education Personnel (CAPES, Projeto Ciências do Mar II-23038.004807/2014-01). They provided a grant for FDL and logistic and financial support to TSL, FDL and SM. We are thankful to the Chico Mendes Institute for Biodiversity Conservation (MMA/ICMBIO) and to the Brazilian Navy for logistics support during the field work.

Data Accessibility

Raw reads sequenced for this study were deposited in the dryad archive (doi: XXX). All the bioinformatic scripts are available in a public GitHub repository https://github.com/casparbein/OctopusInsularis.

Author Contributions

FDL, SML, and RJP conceived the project. FDL and TSL performed the specimen sampling and HL and LAR produced the genomic data. BB and RJP analysed the data and interpreted the results. BB and RJP wrote the first version of the manuscript, and FDL, TSL and SML contributed to the final version.

Figure legends

Fig. 1: Diversification of *Octopus insularis* throughout its species range. A: Sampling sites in this study. Black arrows depict the major oceanic currents (Sissini et al., 2017): the South Equatorial Current (SEC) runs westwards from the African to the Brazilian coast, splitting into the Brazil Current (BC) running southwards, and the North Brazil Current (NBC) running northwards and continuing as the Caribbean Current (CC). Gray scale represents depth. B: EMU PCA considering 11 eigenvalues. Abbreviations of sampling localities: AL - Alagoas, AR – Atol das Rocas, ASC – Ascension Island, BA – Bahía, CE -Ceará, FN – Fernando do Noronha, OIC – Panama, RN – Rio Grande do Norte, SPS – São Pedro/ São Paulo archipelago, STH – Saint Helena Island, TM – Trindade and Martim Vaz archipelago.

Fig. 2: Population structure of *Octopus insularis* . A: Spatial interpolation of ancestral clusters assuming K=2 and 6, using the "69inds_40MD" dataset in TESS3R. B: Co-ancestry matrix of all samples ("full" dataset) inferred in fineRADstructure. Colors on the axis correspond to the individual assignment in the two hierarchical levels shown in A.

Fig. 3 : Phylogenetic relationship of the 69 individuals of*O. insularis* . A: mid-rooted Maximum Likelihood tree inferred by IQtree from concatenated SNP. B: unrooted Coalescent tree inferred by tetrad from resampling unlinked SNPs. Individuals are colored according to their lineage identity. Black nodes demarcate UF bootstrap/bootstrap support of >90, white nodes demarcate bootstrap support of >50.

Fig. 4: Genetic diversity and inbreeding coefficients within inferred clusters for the 69inds_40MD dataset. A: Individual F_{IT}, B: π^* values. Individuals are grouped by clusters inferred from tess3R at K = 6 and fineRADstrucuture. Boxplots are drawn for each cluster, with the solid black line marking the median, the top and bottom end of the box the 25% and 75% quartile boundaries and the whiskers the 1.5 interquartile range.

Fig. 5: Demographic modelling of two pairs of adjacent lineages**.** The top depicts the best model according to AIC values, with the top block representing the ancestral effective population Size (Na), subsequent

blocks effective population size (nu1, nu2, nu1a, nu2a, nu1b, nu2b) scaled to Na, arrows between blocks migration (2*Na*migrants/generation, m12,m21) and height of the blocks represent time since a particular demographic event (2*Na*generations, T, T1, T2). Bottom shows empirical and modelled SFS as well as per-site residuals and a histogram with the distribution of all residuals. A; Best model for N-Coastal vs S-Coastal lineages, B: Best model for S-Oceanic vs S-Coastal lineages.

References

Akaike, H. (1974). A new look at the statistical model identification.*IEEE transactions on automatic control* , *19* (6), 716-723.

Amor, M. D., Norman, M. D., Roura, A., Leite, T. S., Gleadall, I. G., Reid, A., ... & Strugnell, J. M. (2017). Morphological assessment of the Octopus vulgaris species complex evaluated in light of molecular-based phylogenetic inferences. *Zoologica Scripta* , *46* (3), 275-288.

Amor, M. D., Johnson, J. C., & James, E. A. (2020). Identification of clonemates and genetic lineages using next-generation sequencing (ddRADseq) guides conservation of a rare species, Bossiaea vombata (Fabaceae). *Perspectives in Plant Ecology, Evolution and Systematics* , *45* , 125544.

Arruda, W. Z., Campos, E. J., Zharkov, V., Soutelino, R. G., & da Silveira, I. C. (2013). Events of equatorward translation of the Vitoria Eddy. *Continental Shelf Research* , *70* , 61-73.

Avendaño, O., Roura, Á., Cedillo-Robles, C. E., González, Á. F., Rodríguez-Canul, R., Velázquez-Abunader, I., & Guerra, Á. (2020). Octopus americanus: a cryptic species of the O. vulgaris species complex redescribed from the Caribbean. *Aquatic Ecology* , *54* (4), 909-925.

Ávila, S. P., Cordeiro, R., Madeira, P., Silva, L., Medeiros, A., Rebelo, A. C., ... & Johnson, M. E. (2018). Global change impacts on large-scale biogeographic patterns of marine organisms on Atlantic oceanic islands. *Marine Pollution Bulletin* , *126* , 101-112.

Avni, E., Cohen, R., & Snir, S. (2015). Weighted quartets phylogenetics. *Systematic biology* , *64* (2), 233-242.

Barratt, C. D., Bwong, B. A., Jehle, R., Liedtke, H. C., Nagel, P., Onstein, R. E., ... & Loader, S. P. (2018). Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians. *Molecular Ecology* ,*27* (21), 4289-4308.

Barroso, C. X., Lotufo, T. M. D. C., & Matthews-Cascon, H. (2016). Biogeography of Brazilian prosobranch gastropods and their Atlantic relationships. *Journal of Biogeography* , *43* (12), 2477-2488.

Barry, P., Broquet, T., & Gagnaire, P. A. (2022). Age-specific survivorship and fecundity shape genetic diversity in marine fishes.*Evolution letters* , *6* (1), 46-62.

Boessenkool, S., Taylor, S. S., Tepolt, C. K., Komdeur, J., & Jamieson, I. G. (2007). Large mainland populations of South Island robins retain greater genetic diversity than offshore island refuges.*Conservation Genetics* , *8* (3), 705-714.

Bouth, H. F., Leite, T. S., Lima, F. D. D., & Oliveira, J. E. L. (2011). Atol das Rocas: an oasis for Octopus insularis juveniles (Cephalopoda: Octopodidae). *Zoologia (Curitiba)* , *28* (1), 45-52.

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference: a practical information-theoretic approach. Springer.

Canestrelli, D., Sacco, F., & Nascetti, G. (2012). On glacial refugia, genetic diversity, and microevolutionary processes: deep phylogeographical structure in the endemic newt Lissotriton italicus.*Biological Journal of the Linnean Society* , *105* (1), 42-55.

Caye, K., Deist, T. M., Martins, H., Michel, O., & Francois, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources* , *16* (2), 540-548.

Cheddadi, R., Vendramin, G. G., Litt, T., Francois, L., Kageyama, M., Lorentz, S., ... & Lunt, D. (2006). Imprints of glacial refugia in the modern genetic diversity of Pinus sylvestris. *Global Ecology and Biogeography* , *15* (3), 271-282.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics* , *30* (23), 3317-3324.

da Silva, E. J., Bezerra, L. E. A., & Martins, I. X. (2018). The tropical Octopus insularis (Mollusca, Octopodidae): a natural enemy of the exotic invasive swimming crab Charybdis hellerii (Crustacea, Portunidae). *Pan-American Journal of Aquatic Sciences* , *13* (1), 79-83.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , *27* (15), 2156-2158.

DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance of genetic diversity in conservation.*Molecular Ecology* , *30* (17), 4147-4154.

Duda Jr, T. F., Kohn, A. J., & Matheny, A. M. (2009). Cryptic species differentiated in Conus ebraeus, a widespread tropical marine gastropod.*The Biological Bulletin* , *217* (3), 292-305.

Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* , *36* (8), 2592-2594.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity.*Nature Reviews Genetics* , *17* (7), 422-433.

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* , *40* , 481-501.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* , *10* (3), 564-567.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.*Genetics* , *131* (2), 479-491.

Filatov, D. A., Bendif, E. M., Archontikis, O. A., Hagino, K., & Rickaby, R. E. (2021). The mode of speciation during a recent radiation in open-ocean phytoplankton. *Current Biology* .

Fisher, M. C., Helser, T. E., Kang, S., Gwak, W., Canino, M. F., & Hauser, L. (2022). Genetic structure and dispersal in peripheral populations of a marine fish (Pacific cod, Gadus macrocephalus) and their importance for adaptation to climate change. *Ecology and Evolution* , *12* (1), e8474.

Flores-Valle, A., Pliego-Cardenas, R., Jimenez-Badillo, M. D. L., Arredondo-Figueroa, J. L., & Barriga-Sosa, I. D. L. A. (2018). First record of Octopus insularis Leite and Haimovici, 2008 in the octopus fishery of a marine protected area in the Gulf of Mexico. *Journal of Shellfish Research* , *37* (1), 221-227.

Fonseca, R. D., Campos, P., Iglesia, A. R. D. L., Barroso, G., Bergeron, L., Nande, M., ... & Castro, L. F. C. (2021). Low coverage whole genome sequencing reveals the underlying structure of European sardine populations. *Authorea* .

Francisco, P. M., Mori, G. M., Alves, F. M., Tambarussi, E. V., & de Souza, A. P. (2018). Population genetic structure, introgression, and hybridization in the genus Rhizophora along the Brazilian coast.*Ecology and evolution* , *8* (6), 3491-3504.

Frankham, R. (1995). Conservation genetics. *Annual review of genetics* , *29* (1), 305-327.

Frankham, R. (1997). Do island populations have less genetic variation than mainland populations?. *Heredity* , *78* (3), 311-32

Frichot, E., & Francois, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* , *6* (8), 925-929.

Gaither, M. R., Bernal, M. A., Fernandez-Silva, I., Mwale, M., Jones, S. A., Rocha, C., & Rocha, L. A. (2015). Two deep evolutionary lineages in the circumtropical glasseye Heteropriacanthus cruentatus (Teleostei, Priacanthidae) with admixture in the south-western Indian Ocean.*Journal of Fish Biology* , *87* (3), 715-727.

Galtier, N., Nabholz, B., Glemin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal.*Molecular ecology* , *18* (22), 4541-4550.

Giarla, T. C., & Esselstyn, J. A. (2015). The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology* , *64* (5), 727-740.

Giglio, V. J., Pinheiro, H. T., Bender, M. G., Bonaldo, R. M., Costa-Lotufo, L. V., Ferreira, C. E., ... & Francini-Filho, R. B. (2018). Large and remote marine protected areas in the South Atlantic Ocean are flawed and raise concerns: Comments on Soares and Lucas (2018). *Marine Policy* , *96* , 13-17.

Guo, B., DeFaveri, J., Sotelo, G., Nair, A., & Merila, J. (2015). Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC biology* , *13* (1), 1-18.

Guo, B., Li, Z., & Merila, J. (2016). Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Molecular ecology* , *25* (12), 2833-2852.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genet* , *5* (10), e1000695.

Haimovici, M., Leite, T. S., Marinho, R. A., Batista, B., Madrid, R. M., Oliveira, J. E., Lima, L. F. D., & Candice, L. (2014). As pescarias de polvos do nordeste do Brasil. In M. Haimovici, J. M. Andriguetto Filho & P. S. Sunye (Eds.), *A pesca marinha e estuarina no Brasil: estudos de caso* (pp. 147–160). FURG.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages.*Nature* , *405* (6789), 907-913.

Higgins, K. L., Semmens, J. M., Doubleday, Z. A., & Burridge, C. P. (2013). Comparison of population structuring in sympatric octopus species with and without a pelagic larval stage. *Marine Ecology Progress Series* , *486* , 203-212.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation.*Molecular biology and evolution* , *35* (2), 518-522.

Hoffmann, A. A., & Sgro, C. M. (2011). Climate change and evolutionary adaptation. *Nature* , *470* (7335), 479-485.

Jereb, P., Roper, C. F. E., Norman, M. D., & Finn, J. K. (2014)*Cephalopods of the world: An annotated and illustrated catalogue of cephalopod species known to date. Octopods and vampire squids.FAO Species Catalogue for Fisheries Purposes. No. 4, Vol. 3* . Rome: FAO.

Johannesson, K., Ring, A. K., Johannesson, K. B., Renborg, E., Jonsson, P. R., & Havenhand, J. N. (2018). Oceanographic barriers to gene flow promote genetic subdivision of the tunicate Ciona intestinalis in a North Sea archipelago. *Marine biology* , *165* (8), 1-10.

Jorgensen, H. B., Hansen, M. M., Bekkevold, D., Ruzzante, D. E., & Loeschcke, V. (2005). Marine landscapes and population genetic structure of herring (Clupea harengus L.) in the Baltic Sea. *Molecular Ecology* , *14* (10), 3219-3234.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* , *14* (6), 587-589.

Lande, R. (1995). Mutation and conservation. *Conservation biology* , *9* (4), 782-791.

Leite, T. S., Haimovici, M., Molina, W., & Warnke, K. (2008). Morphological and genetic description of Octopus insularis, a new cryptic species in the Octopus vulgaris complex (Cephalopoda: Octopodidae) from the tropical southwestern Atlantic. *Journal of Molluscan Studies* , *74* (1), 63-74.

Leite, T. S., Haimovici, M., Mather, J., & Oliveira, J. L. (2009a). Habitat, distribution, and abundance of the commercial octopus (Octopus insularis) in a tropical oceanic island, Brazil: Information for management of an artisanal fishery inside a marine protected area. *Fisheries research* , *98* (1-3), 85-91.

Leite, T. S., Haimovici, M., & Mather, J. (2009b). Octopus insularis (Octopodidae), evidences of a specialized predator and a time-minimizing hunter. *Marine Biology* , *156* (11), 2355-2367.

Leite, T. S., Batista, A. T., Lima, F. D., Barbosa, J. C., & Mather, J. (2016). Geographic variability of Octopus insularis diet: from oceanic island to continental populations. *Aquatic Biology* , *25* , 17-27.

Lima, F. D., Berbel-Filho, W. M., Leite, T. S., Rosas, C., & Lima, S. M. (2017). Occurrence of Octopus insularis Leite and Haimovici, 2008 in the Tropical Northwestern Atlantic and implications of species misidentification to octopus fisheries management. *Marine Biodiversity* , *47* (3), 723-734.

Lima, F. D., Angeles-Gonzalez, L. E., Leite, T. S., & Lima, S. M. (2020). Global climate changes over time shape the environmental niche distribution of Octopus insularis in the Atlantic Ocean. *Marine Ecology Progress Series* , *652* , 111-121.

Lima, F. D., Leite, T. S. & Lima, S. M. (2022). Seamounts and oceanic currents drive the population structure of Octopus insularis in the Southwest Tropical Atlantic. *Under review*

Lischer, H. E., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* , *28* (2), 298-299.

Lopes, P. F. M., Andrade, L. C. A., Pennino, M. G., & Leite, T. S. (2021). The inter-annual fishing variability in *Octopus insularis*(Leite & Haimovici 2008) as a result of oceanographic factors.*Fisheries Oceanography* , *30* (5), 515–526. doi: 10.1111/fog.12534

Lotze, H. K., Lenihan, H. S., Bourque, B. J., Bradbury, R. H., Cooke, R. G., Kay, M. C., ... & Jackson, J. B. (2006). Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science* ,*312* (5781), 1806-1809.

Ludt, W. B., & Rocha, L. A. (2015). Shifting seas: The impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography* , *42* (1), 25-38.

Macieira, R. M., Simon, T., Pimentel, C. R., & Joyeux, J. C. (2015). Isolation and speciation of tidepool fishes as a consequence of Quaternary sea-level fluctuations. *Environmental Biology of Fishes* , *98* (1), 385-393.

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: population inference from RADseq data. *Molecular biology and evolution* , *35* (5), 1284-1290.

Martins, N., Macagnan, L., Cassano, V., & Gurgel, C. (2021). Barriers to gene flow along the Brazilian coast: a synthesis and data analysis. Authorea Preprints.

Mazzei, E. F., Pinheiro, H. T., Simon, T., Moura, R. L., Macieira, R. M., Pimentel, C. R., ... & Joyeux, J. C. (2021). Mechanisms of dispersal and establishment drive a stepping stone community assembly on seamounts and oceanic islands. *Marine Biology* , *168* (7), 1-11.

Meisner, J., Liu, S., Huang, M., & Albrechtsen, A. (2021). Large-scale inference of population structure in presence of missingness using PCA.*Bioinformatics* , *37* (13), 1868-1875.

Melis, R., Vacca, L., Cuccu, D., Mereu, M., Cau, A., Follesa, M. C., & Cannas, R. (2018). Genetic population structure and phylogeny of the common octopus Octopus vulgaris Cuvier, 1797 in the western Mediterranean Sea through nuclear and mitochondrial markers.*Hydrobiologia* , *807* (1), 277-296.

Mill, G. N., da Costa, V. S., Lima, N. D., Gabioux, M., Guerra, L. A. A., & Paiva, A. M. (2015). Northward migration of Cape Sao Tome rings, Brazil. *Continental Shelf Research* , *106* , 27-37.

Mills, L. S. (2012). *Conservation of wildlife populations: demography, genetics, and management* . John Wiley & Sons.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era.*Molecular biology and evolution* , *37* (5), 1530-1534.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. (2011). How many species are there on Earth and in the ocean?. *PLoS biology* , *9* (8), e1001127.

Moritz, C., & Agudo, R. (2013). The future of species under climate change: resilience or decline?. *Science* , *341* (6145), 504-508.

Neves, J. M., Lima, S. M., Mendes, L. F., Torres, R. A., Pereira, R. J., & Mott, T. (2016). Population structure of the rockpool blenny Entomacrodus vomerinus shows source-sink dynamics among ecoregions in the tropical Southwestern Atlantic. *PloS one* , *11* (6), e0157472.

Oliveira, L. R. D., Fraga, L. D., Ott, P. H., Siciliano, S., Lopes, F., Almeida, R., ... & Bonatto, S. L. (2019). Population structure, phylogeography, and genetic diversity of the common bottlenose dolphin in the tropical and subtropical southwestern Atlantic Ocean.*Journal of Mammalogy* , *100* (2), 564-577.

Palumbi, S. R. (1992). Marine speciation on a small planet. *Trends in Ecology & Evolution* , *7* (4), 114-118.

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annual review of ecology and systematics* ,*25* (1), 547-572.

Pauly, D., & Zeller, D. (2016). Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining.*Nature communications* , *7* (1), 1-9.

Peijnenburg, K. T. C. A., Fauvelot, C., Breeuwer, J. A. J., & Menken, S. B. J. (2006). Spatial and temporal genetic structure of the planktonic Sagitta setosa (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology* ,*15* (11), 3319-3338

Peluso, L., Tascheri, V., Nunes, F. L., Castro, C. B., Pires, D. O., & Zilberberg, C. (2018). Contemporary and historical oceanographic processes explain genetic connectivity in a Southwestern Atlantic coral.*Scientific reports* , *8* (1), 1-12.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one* , *7* (5), e37135.

Pinheiro, H. T., Mazzei, E., Moura, R. L., Amado-Filho, G. M., Carvalho-Filho, A., Braga, A. C., ... & Joyeux, J. C. (2015). Fish biodiversity of the Vitoria-Trindade Seamount Chain, southwestern Atlantic: an updated database. *PloS one* , *10* (3), e0118180.

Pinheiro, H. T., Bernardi, G., Simon, T., Joyeux, J. C., Macieira, R. M., Gasparini, J. L., ... & Rocha, L. A. (2017). Island biogeography of marine organisms. *Nature* , *549* (7670), 82-85.

Portik, D. M., Leache, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., ... & Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular ecology* , *26* (19), 5245-5263.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses.*The American journal of human genetics* , *81* (3), 559-575.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL*https://www.R-project.org/*.

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation biology* , *17* (1), 230-237.

Riginos, C., & Liggins, L. (2013). Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography compass* ,*7* (3), 197-216.

Rosa, R., Pissarra, V., Borges, F. O., Xavier, J., Gleadall, I., Golikov, A... Villanueva, R. (2019). Global patterns of species richness in coastal cephalopods. *Frontiers in Marine Science* ,*6* , 1–16. doi: 10.3389/fmars.2019.00469

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution* , *34* (12), 3299-3302.

Sanford, E., & Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Annual review of marine science* , *3* , 509-535.

Sauer, W. H., Gleadall, I. G., Downey-Breedt, N., Doubleday, Z., Gillespie, G., Haimovici, M., ... & Pecl, G. (2021). World octopus fisheries. *Reviews in Fisheries Science & Aquaculture* ,*29* (3), 279-429

Simon, T., Pinheiro, H. T., Santos, S., Macieira, R. M., Ferreira, Y. S., Bernardi, G., ... & Joyeux, J. C. (2021). Comparative phylogeography of reef fishes indicates seamounts as stepping stones for dispersal and diversification. *Coral Reefs* , 1-11.

Sissini, M. N., de Barros Barreto, M. B. B., Szechy, M. T. M., de Lucena, M. B., Oliveira, M. C., Gower, J., ... & Horta, P. A. (2017). The floating Sargassum (Phaeophyceae) of the South Atlantic Ocean–likely scenarios. *Phycologia* , *56* (3), 321-328.

Sodeland, M., Jentoft, S., Jorde, P. E., Mattingsdal, M., Albretsen, J., Kleiven, A. R., ... & Knutsen, H. (2022). Stabilizing selection on Atlantic cod supergenes through a millennium of extensive exploitation.*Proceedings of the National Academy of Sciences* , *119* (8).

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics* ,*14* (6), 404-414.

Spielman, D., Brook, B. W., & Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them.*Proceedings of the National Academy of Sciences* , *101* (42), 15261-15264.

Timm, L. E., Bracken-Grissom, H. D., Sosnowski, A., Breitbart, M., Vecchione, M., & Judkins, H. (2020). Population genomics of three deep-sea cephalopod species reveals connectivity between the Gulf of Mexico and northwestern Atlantic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers* , *158* , 103222.

Volk, D. R., Konvalina, J. D., Floeter, S. R., Ferreira, C. E., & Hoffman, E. A. (2021). Going against the flow: Barriers to gene flow impact patterns of connectivity in cryptic coral reef gobies throughout the western Atlantic. *Journal of Biogeography* , *48* (2), 427-439.

Watson, R. A., & Tidd, A. (2018). Mapping nearly a century and a half of global marine fishing: 1869–2015. *Marine Policy* , *93* , 171-177.

White, T. A., & Searle, J. B. (2007). Factors explaining increased body size in common shrews (Sorex araneus) on Scottish islands. *Journal of Biogeography* , *34* (2), 356-363.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* , *8* (1), 28-36.