Thainá Cortez Silva

Diversidade genômica do gastrópode *Littoraria flava* (King & Broderip, 1832) e estudo de padrões de fluxo gênico

Genomic diversity of the gastropod *Littoraria flava* (King & Broderip, 1832) and study of gene flow patterns

São Paulo

2020

Thainá Cortez Silva

Diversidade genômica do gastrópode *Littoraria flava* (King & Broderip, 1832) e estudo de padrões de fluxo gênico

Genomic diversity of the gastropod *Littoraria flava* (King & Broderip, 1832) and study of gene flow patterns

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências, na Área de Biologia (Genética).

Orientadora: Sónia Cristina da Silva Andrade

São Paulo

2020

Cortez, Thainá

Diversidade genômica do gastrópode *Littoraria flava* (King & Broderip, 1832) e estudo de padrões de fluxo gênico.

Dissertação (Mestrado) – Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. *Littorinidae* 2. Genética populacional 3. GBS (*Genotyping-by-Sequencing)* 4. Meta-população I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora

_____ _____

Prof(a). Dr(a). **Prof(a). Dr(a).**

_____ _____

Prof(a). Dr(a). **Profa. Dra. Sónia Cristina da Silva Andrade**

Dedicatória

Ao meu avô, Amilcar, pelo apoio

(*in memoriam).*

Agradecimentos

À minha orientadora, Sónia, pelos inúmeros ensinamentos dentro e fora da biologia, pela paciência e amizade. Obrigada por me mostrar os caminhos até aqui.

À minha mãe e ao meu irmão, por serem tão presentes em minha vida e pelo apoio diário incondicional.

Ao meu pai, por sempre acreditar e investir no meu potencial. Sem o seu suporte este trabalho certamente não existiria.

Ao meu melhor amigo e namorado, Henrique, por sempre incentivar e ansiar pelo meu crescimento. Você tornou essa jornada muito mais leve. Obrigada por emprestar a sua casa às vezes.

À toda minha família, por celebrar cada pequena vitória que tive neste tempo.

À Yasmin, por acompanhar essa jornada desde o início, me ajudando com tudo que pôde, desde com o uso de Photoshop até leituras e sugestões de diversas escritas.

À Cecili, pela paciência em ensinar, pelas coletas e amizade dentro e fora do laboratório. Sua cantoria às 20h no laboratório tornaram os trabalhos mais alegres.

À Camilla, por ser uma ótima companhia e parceira nos experimentos de reprodução, também pela leitura e sugestões desta dissertação.

Aos meus colegas de laboratório, Gabriel Sonoda, Rafael, Tammy, e Fagner, que tanto me ajudaram em diversas etapas deste mestrado, especialmente com o obscuro mundo da bioinformática.

Ao Tiago, pela ajuda com imagens, ideias e bancada. Obrigada também pela ótima companhia, ouvindo a cantoria da Cecili às 20h no laboratório.

Aos demais colegas do laboratório LDG que de alguma forma colaboraram para o meu crescimento e para a execução deste trabalho.

À Stephanie e ao Pachelle, pela ajuda com as coletas de campo no Rio de Janeiro e Espírito Santo.

Às minhas amigas Marcela, Aline e Izabel, por todo o apoio e incentivo desde a graduação.

Ao Paulo, técnico de laboratório, por tanto ter me ajudado e ensinado neste período.

Ao Instituto de Biociências, por fornecer toda a infraestrutura necessária para a execução deste projeto.

Ao Programa de Pós Graduação em Biologia (Genética), por oferecer a oportunidade e todas as condições necessárias para a realização deste trabalho.

Ao CEBIMar, por toda a infraestrutura e alojamento durante as coletas e execução de disciplinas e projetos paralelos que muito me acrescentaram.

À Coordenação de Formação de Ensino Superior (CAPES), pela concessão de bolsa de mestrado.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela bolsa de mestrado concedida e pelo apoio financeiro para a realização do projeto de pesquisa.

Finalmente, ao meu avô Amilcar, que não pôde acompanhar minha jornada como bióloga, a qual se deve, em grande parte, ao seu incentivo incondicional.

Sumário

Lista de Ilustrações

Lista de Tabelas

1. Introduction

1.1 Seascape genetics

Understanding how populations' connectivity is influenced by the marine environment and other abiotic characteristics is one of the major goals of seascape genetics (SELKOE *et al.*, 2016). The assumptions underlying the seascape genetics presume an association between genetic, spatial and temporal patterns. Comprehending how this association works is fundamental to untangle the evolution of natural populations (LIGGINS *et al.*, 2019). However, because direct observations are essentially unlikely for many species, marine populations connectivity is often indirectly inferred (HELLBERG, 2009; SELKOE *et al.*, 2016). An indirect method uses spatial distributions of alleles or/and phenotypic traits to evaluate the level or pattern of gene flow in a population. Evolutionary studies usually take two main approaches: (1) assessing the variation in genomic regions that determine phenotypic traits and thereby influence fitness using adaptive markers; and (2) evaluating the variation in DNA sequences presumed neutral, which may not affect fitness. These markers provide very different answers: while the adaptive markers can be used to understand the interaction between phenotype *versus* genotype and elucidate adaptation processes, neutral markers tell us about the genetic structure and connectivity of populations (CONOVER *et al.*, 2006; PALUMBI, 2003).

Most marine species have traditionally been assumed to have open populations that are interconnected by gene flow. This assumption comes from the apparent lack of dispersal barriers in marine systems and the fact that most marine invertebrates have planktonic larvae (BOHONAK, 1999; GROSBERG & CUNNINGHAM, 2001; PALUMBI, 1994). However, this paradigm has changed considerably since several studies revealed highly dispersal species presenting strong genetic structure on both large and small spatial scale (e.g. BEAUMONT, 1982; BUCKLIN, 2000; CARINI & HUGHES, 2006; HEIPEL *et al.*, 1999; KRAMARENKO & SNEGIN, 2015; LAUNEY, 2002; RYNEARSON & ARMBRUST, 2004; TAYLOR, 2003). These findings clearly suggest that genetic differentiation depends not only on the dispersal capacity, but also on other mechanisms such as behavior, hydrographical barriers to dispersal,

oceanic environment and life-history traits (e.g. high fecundity, population sizes and life cycle) (PALUMBI, 2003; WAPLES, 1998).

Studies over the nearshore environment emphasize some oceanographic features that often generate the afore-mentioned micro spatial scale variation, such as nutrients, temperature, pH, bathymetry, dissolved oxygen, phosphate and others (MENGE, 2000; NAVARRETE *et al.*, 2005). These complex mosaics of environmental conditions of many coastlines have considerable potential to create local adaptation by natural selection on small scales (FUNK *et al.*, 2012). Local adaptation can occur when some individuals with determined genotype frequencies have higher relative fitness than others from different habitats despite the random genetic drift and gene flow effects (KAWECKI & EBERT, 2004). Therefore, it is expected that the potential for local adaptation decrease in species with high levels of gene flow (CONOVER *et al.*, 2006; HEREFORD, 2009; PALUMBI, 1994; YAMADA, 1989). Local selective pressure may be substantial when considering the great environmental heterogeneity experienced by marine species, and local adaptation may be far more common in marine systems than assumed previously (COWEN, 2006; HEREFORD, 2009; YEAMAN & WHITLOCK, 2011; e.g. BURFORD *et al.*, 2014; SANDOVAL-CASTILLO *et al.*, 2018). It is expected that genomic regions under selection present higher variation when compared to selectively ‘neutral’ gene regions (CONOVER *et al.*, 2006). Nonetheless, the task of identifying signatures of selection and local adaptation in marine species is not trivial, and usually requires the combination of knowledge about gene function, landscape effects and species’ biology to provide concrete evidence of outlier loci adaptive roles (MANEL *et al.*, 2003).

In South America, the Brazilian coast extends over 8,000 km under the influence of an extensive set of oceanographic and ecological conditions. Many Brazilian population genetic studies test the presence of intrinsic biogeographical barriers, as well as the effects of environmental factors on dispersal along the coast range. These studies include many taxonomic groups, such as ascidian (DIAS *et al.*, 2006; NÓBREGA *et al.*, 2004; ROCHA *et al.*, 2012); bryozoa (MIRANDA *et al.*, 2018; VIEIRA *et al.*, 2012); mollusca (ANDRADE *et al.*, 2005; ANDRADE & SOLFERINI, 2007; JOSÉ & SOLFERINI, 2007; MOREIRA *et al.*, 2011); nemertea (ANDRADE *et*

al., 2011); polychaeta (PAIVA *et al.*, 2019; SEIXAS *et al.*, 2018); porifera (LAZOSKI *et al.*, 2001), and reef fishes (JOYEUX *et al.*, 2008; ROCHA, 2003; ROCHA *et al.*, 2002) across continental and island lands under different hypotheses. With the advance of bioinformatics tools in incorporating ecological and biotic factors into computational modeling, testing hypotheses about connectivity in the marine system has become increasingly accurate (CHAN *et al.*, 2011; KNOWLES, 2009). These approaches have improved the quality of genetic diversity inferences in Brazilian marine populations.

1.2 *Littorinidae*

The littorinids are a group of small gastropods that occurs on rocky shores and mangroves around the world, with shell lengths ranging from 0.5 to 4 cm. The group lives on intertidal zones and can feed algae or lichens (REID, 1989, 1996). In the majority of the species, the spawning releases hundreds of eggs, followed by a planktonic larval stage (RIOS *et al.*, 1994). The larvae can remain in the water column for weeks (REID, 1999; SCHELTEMA, 1971). From the littorinids, only the genus *Echinolittorina* Habe, 1956 and *Littoraria* Griffith & Pidgeon, 1834 occur on the Brazilian coast (REID, 1986). *Littoraria* is a group of 39 molluscs closely associated with mangroves, distributed in tropical or subtropical regions, mainly occupying mangrove trees, salt marshes, driftwood, and a minority in rocky shores (REID, 1999). The species *Littoraria flava* (KING & BRODERIP, 1832) and *Littoraria angulifera* are found in Brazil, being restricted to tropical zones and usually associated with mangroves or rocky shores near rivers (REID, 1986).

Littoraria flava (Fig. 1) shows a continuous distribution in the supratidal area of rocky shores and estuarine environments, is oviparous and dioecious. The spawning happens only when the animals are submerged, when hundreds of single egg capsules with approximately 30 µm of diameter are released. Each capsule keeps a single embryo, which will develop until the veliger phase (personal observation, unpublished data). Under laboratory conditions, this process takes approximately three days. Although its adult phase is relatively sedentary, the larvae disperse widely. The length of the larval phase is still unknown, but estimated between 3 and 10 weeks (REID, 1999; REID, 1986; RIOS *et al.*, 1994). The consequences of a planktonic stage

to gene flow dynamics in *L. flava* were studied using allozymes (ANDRADE *et al.*, 2005; ANDRADE & SOLFERINI, 2007). The authors found moderate structure on the macrogeographic scale, but even stronger variation across few meters, i.e. subpopulations in a single rocky shore. These results could be explained due to asynchronous spawning associated with recurrent colonization or to a possibility that allozymes may undergo natural selection (JANSON, 1987; JOHANNESSON & TATARENKOV, 1997; TATARENKOV & JOHANNESSON, 1999).

Fig. 1. *Littoraria flava* in its natural habitat. Praia de Gamboa rocky shore, Espírito Santo, Brazil. The specimens can be (A) solitary (B) close to each other or (C) in aggregations within rocky shore fissures.

1.3 Phylogeography in marine systems

Phylogeographic studies goals include assessing the influence of geographic events and environmental factors over species ecology and evolutionary history, considering a wide range of spatial and temporal scales. When applied in a comparative context, these works are able to reveal patterns in entire communities or shared responses from different species due to a common demographic history (AVISE, 2012; KNOWLES, 2009). In broad terms, one of the most important

contributions of phylogeography is to emphasize non-equilibrium aspects of population structure and microevolution (AVISE, 2012). Since there is not a precise model of how genetic lineages are spatially configured, comparative assessments of many species have revealed a good deal about the nature of intraspecific evolution (AVISE, 2009). Several molecular markers have similarly been employed in phylogeographic investigations, such as microsatellites and mitochondrial sequences. Nonetheless, these markers have very different forms of hereditary transmission and dynamics (KARL *et al.*, 2012). The mitochondrial DNA (mtDNA) presents relatively rapid evolution and therefore a high level of polymorphism, which has led to its widespread use as a genetic marker for demographic history analyses (e.g. UTHICKE & BENZIE, 2003; LEE & BOULDING, 2007; CRANDALL *et al.,* 2007; BAKER *et al.,* 2008; LIU *et al.,* 2012). Because it is haploid and most likely uniparentally transmitted, the mtDNA has a smaller effective size and for this reason it is expected to present a more strongly effect of genetic drift than in nuclear loci (AVISE, 2012; KARL *et al.*, 2012). All these characteristics indicate mtDNA as a useful tool for investigation of demographic events, population structure and dynamics over time (AVISE, 2009, 2012).

Although the genetic approaches are particularly useful to understand both demographic processes and gene flow dynamics, they still present many challenges when it comes to marine populations: difficulties in collecting, markers development and the fact that most of the theoretical models hardly explain the data (SELKOE *et al.*, 2008). In this scenario, methods that generate genome-wide datasets via Next-Generation-Sequencing (NGS) have become a helpful tool by providing a large number of loci with high levels of polymorphism, the Single-Nucleotide-Polymorphisms (SNPs). These methods enable great accuracy to quantify genomic variation for both neutral and non-neutral signatures (NOSIL *et al.*, 2009; RELLSTAB *et al.*, 2015; STAPLEY *et al.*, 2010), facilitating the population genetics studies of non-model organisms (BENESTAN *et al.*, 2015; LIGGINS *et al.*, 2019; LUIKART *et al.*, 2003). Besides, it does not require any development or genetic information *a priori*. Even though the NGS techniques have been increasingly applied, the combination of neutral and adaptive markers for non-model organisms is yet far to be common

(EKBLOM & GALINDO, 2011; HELYAR *et al.*, 2011; SEEB *et al.*, 2011), being used more frequently in studies of economically important organisms such as bivalves (LAL *et al.*, 2016; VAN WYNGAARDEN *et al.*, 2017), lobsters (BENESTAN *et al.*, 2015) and fishes (DIBATTISTA *et al.*, 2017; LIMBORG *et al.*, 2012).

2. Objectives

The main goal of this study was to assess the demographic processes, population dynamics and environmental predictors underlying the connectivity among populations of *Littoraria flava*, a marine gastropod.

Based on the great larval dispersal capacity of *L. flava*, our hypothesis was that populations across the Brazilian coast are interconnected across large spatial distances due to high levels of gene flow. However, due to the heterogeneity of the intertidal environment, we expect that, at the microgeographic scale, the species presents higher genetic structuring, according to Andrade & Solferini (2007). To test these hypotheses, we:

1. Genetically characterized individuals of *L. flava* from 11 locations using SNPs and mtDNA markers;
2. Identified both neutral and adaptive sets of SNPs loci;
3. Assessed genetic diversity and population structure on a macro spatial scale using SNPs and mtDNA markers;
4. Assessed genetic variation on a micro spatial scale using SNPs;
5. Investigated the demographic history and;
6. Detected potential loci associated with local adaptation.

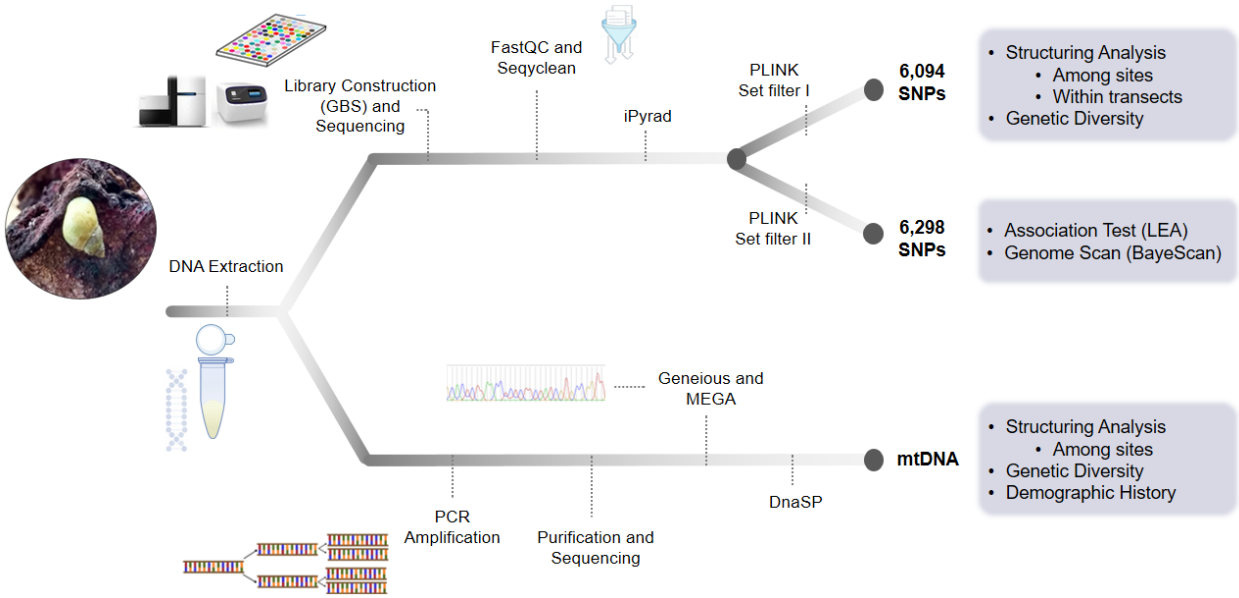
Finally, we discussed the implications of using different genetic markers for demographic inferences in non-model organisms' studies through a seascape genetics approach.

3. Materials and methods

The methodology included the (1) sampling and experimental design; (2) sequencing and filtering steps of mtDNA; (3) libraries construction and SNPs quality filtering; (4) analyses of genetic diversity and variation in macro (mtDNA and SNPs) and micro (SNPs) spatial scales; (5) analysis of the demographic history

(mtDNA) and finally (6) genome scan and seascape association test (SNPs). These methodological steps are shown in Fig. 2 and detailed in the following sections.

Fig. 2. Methodological steps used in this study. For each set of genetic markers, analyses of demographic history, population variation, genetic diversity and environmental selection were conducted by combining different approaches.

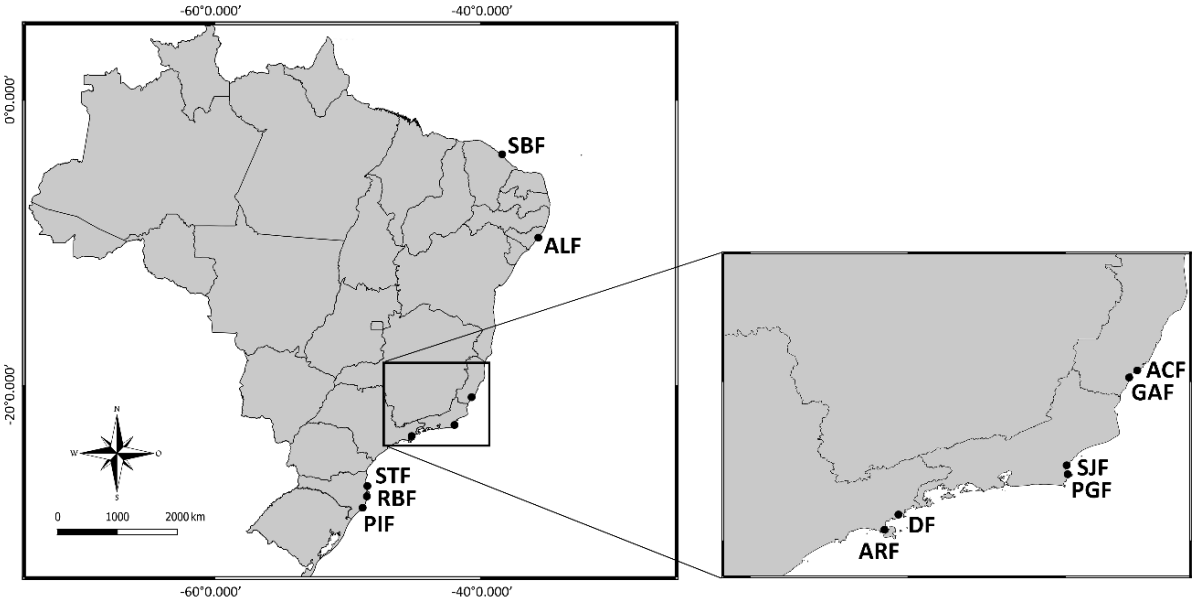


3.1 Sampling and experimental design

Ninety-two individuals of *Littoraria flava* were collected in the supralittoral from 11 localities distributed across the Northeast, Southeast and South regions of the Brazilian coast (Fig. 3). In six out of the eleven localities, the samples were collected along horizontal transects towards the sea, following the experimental design of Andrade & Solferini (2007) (Table 1). The distances among the sites were established according to $Dn = 2^{n-1}$, where D is the distance in meters and n is the sequence number of the site. The periwinkles were collected within 1 m^2 around three marked sites, which could be 0, 2, 4, 8, 16, 32 and 64 m along the transect. The three sites were chosen according to the abundance of the species. This approach allowed us to investigate both genetic structure in macro and micro spatial scales, i.e. subpopulations within a rocky shore.

For the demographic history investigation through mtDNA sequencing, at least two individuals from each locality were chosen, ignoring the sites within transects, totalizing 63 specimens distributed across the same 11 locations (Table 1). The local water surface temperature and salinity were registered, and the associated fauna and flora were preserved in 96% ethanol for subsequent environmental association tests.

Fig. 3. Sampling locations along the Brazilian coast.



SBF - Sabiaguaba; ALF - Alagoas; ACF - Anchieta; GAF - Gamboa; SJF - São João; PGF - Praia da Gorda; DF - Praia Dura; ARF - Araçá; STF - Santo Antônio; RBF - Ribeirão da Ilha; PIF - Ponta de Ribeirão da Ilha.

3.2 DNA extraction

The collected individuals were stored in liquid nitrogen until DNA extraction. Genomic DNA was extracted following the protocol from Doyle & Doyle (1987). The DNA integrity was checked with 1% agarose gel electrophoresis and quantified using dsDNA BR Assay kit (Invitrogen™) on Qubit v3 fluorimeter. For the subsequent steps, the DNA samples were adjusted to a final concentration of 20 ng/µL in a final volume of 30 µL using pure and DNP or RNase-free distilled water (Invitrogen™ UltraPure DNase / RNase-Free Distilled Water).

Table 1. Sampling locations and sample sizes used for each molecular marker.

Location	Transect	Code	Lat. (ºS)	Long. (ºW)	SNPs	mtDNA
Northeast						
Sabiaguaba, CE	NA	SBF	3°47'24"	38°25'23"	11	7
Alagoas, AL	NA	ALF	9°36'57"	35°44'13"	6	5
Southeast						
Anchieta, ES	4, 8, 64m	ACF	20°48'37"	40°39'39"	8	2
Praia de Gamboa, ES	0, 2, 32m	GAF	20°53'19"	40°45'55"	8	3
Barra de São João, RJ	0, 8, 32m	SJF	22°35'55"	41°59'25"	10	7
Praia da Gorda, RJ	0, 8, 32m	PGF	22°43'48"	41°58'2"	10	5
Praia Dura, SP	4, 16, 64m	DF	23°29'32"	45°09'55"	9	10
Araçá, SP	4, 16, 64m	ARF	23°48'47"	45°24'31"	8	10
South						
Santo Antônio, SC	NA	STF	27°30'46"	48°30'57"	6	7
Ribeirão da Ilha, SC	NA	RBF	27°42'45"	48°33'40"	14	5
Ponta de Rib. da Ilha, SC	NA	PIF	27°49'54"	48°34'14"	2	2

Transect - indicates if the sampling were performed along horizontal transects and the corresponding sites; SNPs - number of individuals sent to library construction; mtDNA - number of individuals chosen for mitochondrial sequencing.

3.3 Mitochondrial markers (mtDNA)

The mitochondrial regions cytochrome oxidase c subunit I (COI) and 16S ribosomal RNA (16SrRNA) were amplified through polymerase chain reactions (PCR) using the primer pairs LCO1490/HCO2198 (FOLMER *et al.*, 1994) 5-GGT CAA CAA ATC ATA AAG ATA TTGG-3 and 5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3', and 16SH/16SR (PALUMBI *et al.*, 1991) 5'-CGC CTG TTT ATC AAA AAC AT-3' and 5'-CCG GTC TGA ACT CAG ATC ACGT-3', respectively. The PCRs were performed with Taq PCR Master Mix (Qiagen®) containing 2.5 units of Taq DNA Polymerase, 1x QIAGEN PCR Buffer, 1.5 mM MgCl2, 200 µM of each dNTP, 0.35 µM of each oligonucleotide, 100 ~ 150 ng of DNA and ultrapure water to complete the reaction volume. The PCR followed the conditions: denaturation at 94°C for 1 min, 35 cycles of denaturation at 94°C for 1 min; annealing temperature for 1 min, extension at 72°C for 1 min; and extension at 72°C for 10 min. Annealing temperatures were 52.5ºC and 51ºC for COI and 16SrRNA, respectively. PCR products were purified according to an

adapted protocol with polyethylene glycol solution 15% (PEG, http://labs.icb.ufmg.br/lbem/protocolos/peg.html) and then amplified for sequencing using BigDye Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems) at the Myleus Sequencing Facility (Belo Horizonte, Brazil).

The chromatograms were analyzed with Geneious 9.1.8 (Biomatters Limited), which performed BLAST searches on NCBI to check for contaminations or sequencing errors. The software also read the frames to identify stop codons and, finally, produced the consensus sequence of each individual. Individuals consensus sequences were aligned using MEGA (KUMAR *et al.*, 2016). The 16SrRNA and COI data were concatenated with the DnaSP v.5.10.01 (LIBRADO & ROZAS, 2009) into a single mitochondrial haplotype.

3.4 Library construction with Genotyping-by-Sequencing (GBS)

Individual libraries were produced through the Genotyping-by-Sequencing (GBS) method following the protocol from Elshire *et al.* (2011). Genomic DNA of each sample was digested with PstI restriction enzyme (5' CTGCAG 3') (DE DONATO *et al.*, 2013) and ligated to barcode and common adaptors with appropriate sticky ends. The products were grouped into sets of 40 to 53 samples and then amplified by PCR using generic primers matching the common adaptors following the PCR conditions: 5 minutes at 72 °C, 30 seconds at 98 °C, 18 cycles of 10 seconds at 98 °C, 30 seconds at 65 °C and 30 seconds at 72 °C and an extension step of 5 minutes at 72 °C. The presence of spare adapters and the size of the DNA fragments were assessed by quantification on the Agilent 2100 Bioanalyzer (Agilent Technologies) with the Agilent DNA 1000 kit and by qPCR on Light Cycler 480II (Roche) with Kapa Biosystems kit. Finally, the libraries were sequenced in four separated lanes: EM01 (7 individuals), EM02 (46 individuals), EM06 (7 individuals) and EM17 (32 individuals). The sequencing were performed on the HiSeq 2500 (Illumina®) platform of the Center for Functional Genomics Applied to Agriculture and Agroenergy (Animal Biotechnology Laboratory, LZT/ESALQ/USP, Piracicaba/SP).

3.4.1 SNPs filtering and quality control

The first quality control on the raw sequences was implemented with FastQC v.0.11.8 (S. Andrews, 2010) from the BaseSpace platform (Illumina®, California, USA). The SeqyClean pipeline v.1.10.07 (ZHBANNIKOV *et al.*, 2017) discarded sequences smaller than 50 bp and removed adapter sequences, vectors and oligonucleotides detected based on the UniVec database (NCBI, ftp://ftp.ncbi.nlm.nih.gov/pub/UniVec/). The program iPyrad v.0.7.28 (EATON, 2014) filtered bases with quality scores lower than 20 in a phred scale, assigned reads to individual samples, edited and then clustered reads into consensus sequences through paralog identification. The clustering step assumed a minimum of 90% similarity. The quality parameters discarded consensus sequences with more than five ambiguous bases, eight heterozygous bases and two alleles per site. To evaluate the density of missing data per locus across the samples and the efficacy of the filtering parameters, the final dataset was analyzed through a matrix occupancy (DE MEDEIROS & FARRELL, 2018). The resulting VCF file (Variant Call Format) was converted to other program-specific input formats using PGDSpider v.2.1.15 (LISCHER & EXCOFFIER, 2012).

The software PLINK (PURCELL *et al.*, 2007) was used to create two datasets from different filtering steps. In the first set (hereafter, filter set I), all variants with linkage disequilibrium (LD) with a correlation higher than 0.5 ($r^2 > 0.5$) were removed. The resulting SNPs were used to determine population structure, genetic diversity and demographic history. In the second set of filtering steps (hereafter, filter set II), no LD filtering was performed. This set was used to identify putative adaptive loci. Since a neutral locus may be tightly linked to a locus under selection (BARTON, 2011; CONOVER *et al.*, 2006; MUSTONEN & LÄSSIG, 2009), the removal of SNPs in linkage disequilibrium minimizes the possible effects of selective forces in the gene flow analyses (HOLDEREGGER *et al.*, 2008). On the other hand, by adding these SNPs in the second dataset, the chances of detecting the outliers SNPs might increase. For both datasets, missing genotypes (geno) with a frequency higher than 20%, and SNPs with minimum allele frequency (MAF) lower than 1% were removed.

3.5 Genetic diversity and demographic analyses

3.5.1 Mitochondrial markers

The parameters used to quantify the genetic diversity were the number of polymorphic sites (S), haplotype diversity (h) (NEI, 1987), nucleotide diversity (π) (TAJIMA, 1989) and the mean number of pairwise differences (*k*) (TAJIMA, 1983) per site. The demographic processes over time were assessed using Tajima's D (TAJIMA, 1989), Fu's F_S (FU, 1997) tests and mismatch distribution analyses (HARPENDING *et al.*, 1993) for each locality using 10,000 permutations. All tests were conducted in Arlequin v.3.5 software (EXCOFFIER & LISCHER, 2010).

3.5.2 SNPs

Multi-locus estimates of expected heterozygosity (H_E), observed heterozygosity (H_O), nucleotide differences (θ_S), nucleotide diversity (θ_π) and the fixation index F_{IS} (WEIR & COCKERHAM, 1984) were calculated with Arlequin. The Bartlett test evaluated the variance differences among observed and expected heterozygosity, and its significance was tested with 10,000 permutations (p-value < 0.05).

3.6 Population structure analysis

3.6.1 Mitochondrial markers

Based on the mitochondrial haplotype, a minimum spanning network (MSN) was constructed in PopART v.1.7 (LEIGH & BRYANT, 2015), which implements the same statistics method from TCS (CLEMENT *et al.*, 2002) to infer the most parsimonious branch connections at the 95% confidence level between haplotype pairs.

The software Arlequin calculated the genetic differentiation by using the unbiased F_{ST} estimator θ (WEIR & COCKERHAM, 1984). The significance of the observed F_{ST} was determined by running 10,000 permutations (AMOVA, 100,000 MCMC steps, $p < 0.05$). To avoid mixing different populational units, the AMOVA followed hierarchical levels, considering 1) individuals from one location and 2) from a region (South, Southeast and Northeast). To test whether the population structure follows a model of isolation by distance (IBD), a Mantel test was implemented with adegenet package v.1.4 (JOMBART, 2008) from R (R CORE TEAM, 2013) using 10,000 permutations and following the same hierarchical levels from AMOVA. Based

on the localities coordinates, the geographical distances were transformed into Euclidean distances using the dist function of R.

3.6.2 SNPs

The remaining SNPs from set filter I were used for the clustering analyses, which evaluated, separately (1) all individuals from the 11 localities (macrogeographic scale) and (2) individuals within each transect (microgeographic scale).

Both Bayesian method from STRUCTURE v.2.3.4 (PRITCHARD *et al.*, 2000) and the non-model-based method from Discriminant Analysis of Principal Components (DAPC) were implemented in the clustering analyses. The populations (K) were allowed to vary from 1 to 11. The STRUCTURE settings assumed admixture, correlated allele frequencies and none geographic information *a priori*. The program ran each K-value 20 times with a burnin of 50,000 followed by 1,000,000 Markov Chains Monte Carlo (MCMC) iterations. The most likely number of genetic groups was chosen based on ΔK using STRUCTURE HARVESTER (EARL & VONHOLDT, 2012). The DAPC was performed using the adegenet package, also without providing the geographic information *a priori*. The function *find.clusters* indicated the optimal number of groups according to the Bayesian information criterion (BIC) method. To avoid retaining too many discriminant functions and consequent biases of clusters inferences, the appropriate number of discriminant functions (n = 20) was chosen based on the optimal α (JOMBART *et al.*, 2010). The *loading.plot* function from adegenet identified the SNPs with contribution higher than 0.1% for the inferred structuring pattern.

The genetic differentiation was calculated by Arlequin with AMOVA based on the unbiased F_{ST} estimator θ (WEIR & COCKERHAM, 1984), using the same methodology described in section 3.6.1. For the macro spatial analysis, the AMOVA considered the same hierarchichal levels used for mitochondrial markers. For the micro spatial scale, within transects, each site was used as a populational unit. The Mantel test from adegenet checked the IBD model by using 10,000 permutations and the same hierarchical levels used in AMOVA for macro and micro spatial scales.

3.7 Seascape genetics analyses and putative adaptive loci

In order to identify outlier SNPs, the markers from set filter II were submitted for the genome scan in BayeScan v.2.0 (FOLL & GAGGIOTTI, 2008). The program uses a Bayesian approach, which decomposes the F_{ST} values into a locus-specific component (α, selection effect) and a population-specific component (β, demographic effect). If the observed pattern of diversity is only explained with the locus-specific component, the software assumes departure of the neutrality. The program conducted twenty pilot runs of 50,000 iterations followed by 100,000 simulations with a prior odd of 10 and a 5% false discovery rate (FDR), which set the neutral model being 10 times more likely than the selection model.

Possible associations between SNPs and environmental variables were identified with the latent factor mixed models tests (LFMM) (FRICHOT *et al.*, 2013; RELLSTAB *et al.*, 2015) from R package LEA (FRICHOT & FRANÇOIS, 2015). For that, a principal component analysis (PCA) first detected the most representative environmental predictors through the study area. This analysis was performed using the *princomp* function from stats R package and both continental and oceanographic bioclimatic variables (Table 1S) from WorldClim plus Bio-Oracle databases. Only the variables showing strong representations on the first PCA axes were attached to LFMM test. To avoid the increasing of both type I and II errors, the program ran multiple k-values (K ± 2, where K is the optimal number of groups according to STRUCTURE and DAPC), which candidates SNPs were only those detected across all runs. For each K, the LFMM conducted five runs per environment variable with a burn-in of 5,000 followed by 20,000 iterations. The *p-value*s were adjusted based on the median z-score to increase the power of the LFMM statistic test (STORFER *et al.*, 2018).

The loci containing SNPs detected on BayeScan, LFMM and DAPC were blasted against a transcriptome of *Littoraria flava* annotated in November 2019 (unpublished data) using the SwissProt from UniProt (https://www.uniprot.org/) database. As the GBS reads were only ~80 bp in length, this step helped to reduce the number of false positives found when performing a blast search of these query sequences. The blast was performed using a BLASTN search tool from BLAST v.2.9.0

(CAMACHO *et al.*, 2009) with an e-value threshold of 10^{-3}. Gene ontology (GO) annotation terms attributed to the transcriptome were then associated with the candidate SNPs.

4. Results

4.1 Obtained data

4.1.1 Mitochondrial markers

For all the 63 individuals, the COI and 16SrRNA length was of 626 bp and 487 bp, respectively. There were 31 variables sites in COI gene, where 11 of them were parsimony-informative, and 20 singletons. The 16SrRNA presented 19 variable sites with 3 parsimony-informative and 16 singletons. After concatenating the COI and 16SrRNA for the following analyses, the mitochondrial haplotype presented 1,113 bp with 46 variable sites and 13 parsimony-informative. Thirty-three singletons were identified.

4.1.2 SNPs

The four sequenced lanes resulted in 322,479,123 reads of *L. flava*. The Seqyclean pipeline removed ~29% of the reads (Table 2S). From the remaining 227,389,910 sequences, the first quality filter done by iPyrad retained 19,133 SNPs within 2,249 loci with approximately 12.67% of missing data per locus (Table 2, Fig. 1S). Seven samples were removed due to the amount of missing data (> 35% per individual). The average of reads per sample were 2,748,578, ranging from 203,705 to 10,356,692 (Table 3S).

The filter set I of PLINK resulted in 6,094 SNPs, whereas the filter set II retained 6,298 SNPs. Both datasets included 1,572 loci of 85 individuals. In this data, a locus represents a fragment resulted from the sequencing, and a SNP is the variable site, i.e. a base pair, which varies among individuals. In other words, a single locus can contain one or multiple SNPs.

Table 2. Filtering steps of Seqyclean, iPyrad and PLINK. In Seqyclean, the filters were applied at the read level; in iPyrad, at both read and locus levels, and in PLINK, at the SNP level.

Seqyclean	
Total pre filtered reads (4 lanes)	322,479,123
	Removed reads count (%)
Quality and length criteria	95,098,153 (29.49%)
Retained reads (%)	227,380,970 (99.15%)
iPyrad	
Total pre filtered loci	136,225
	Removed loci count (%)
Duplicated loci	16,711 (12.27%)
Indels per locus > 8	166 (0.12%)
SNPs per locus > 15	3,799 (2.79%)
Heterozygous sites per locus > 0.50	167 (0.12%)
Represented samples < 70%	112,024 (82.24%)
Alleles per individual > 2	1,109 (0.81%)
Retained loci	2,249
PLINK	
Initial potential SNPs	19,133
	Removed SNPs count (%)
Filter set I (MAF < 1%; geno > 20%, LD > 0.5)	13,039 (68.15%)
Filter set II (MAF < 1%; geno > 20%)	12,835 (67.08%)
Retained SNPs for structure analysis – Dataset I	6,094
Retained SNPs for selection analysis – Dataset II	6,298

MAF - Minimum allele frequency; geno – missing genotype; LD - linkage disequilibrium.

4.2 Genetic diversity and demographic history

4.2.1 Mitochondrial markers

The diversity analysis revealed moderate to high haplotype diversity (h) in all localities, with a mean of 89%, ranging from 67% (Gamboa) to 100% (Anchieta and Ponta de Ribeirão da Ilha) (Table 3). Overall nucleotide diversity (π) ranged from 0.001 (Gamboa) to 0.005 (Sabiaguaba), with most localities presenting 0.002 (Alagoas, Anchieta, Praia Dura and Santo Antônio). The average number of pairwise differences (k) showed high variation within localities, varying from 1.333 (Gamboa) to 5.40 (Praia da Gorda). The localities with the highest number of haplotypes were from the Southeast and South regions (Praia Dura, Araçá and Santo Antônio). Sabiaguaba presented the highest amount of polymorphic sites.

Table 3. Genetic diversity indexes based on the mitochondrial haplotypes.

Locality	*H*	*S*	*h*	*π* (± SD)	*k* (± SD)
SBF	6	16	0.952	0.005 (0.002)	4.571 (2.553)
ALF	4	8	0.900	0.002 (0.002)	3.200 (1.979)
ACF	2	3	1.000	0.002 (0.003)	3.000 (2.449)
GAF	2	2	0.667	0.001 (0.001)	1.333 (1.098)
SJF	6	11	0.952	0.004 (0.002)	4.380 (2.459)
PGF	4	12	0.900	0.004 (0.003)	5.400 (3.130)
DF	7	11	0.867	0.002 (0.001)	2.866 (1.644)
ARF	7	12	0.911	0.003 (0.001)	3.466 (1.930)
STF	7	9	0.952	0.002 (0.001)	3.142 (1.847)
RBF	3	8	0.700	0.004 (0.002)	3.600 (2.189)
PIF	2	3	1.000	0.003 (0.003)	3.000 (2.444)

H – number of haplotypes; S – number of polymorphic sites; h – haplotype diversity; π – nucleotide diversity; k – average number of pairwise differences; SD – standard deviation. Abbreviations as in Table 1.

The three different tests of populational events (Tajima's D, Fu's F_S and mismatch distribution) showed contrasting patterns. Both Fu's F_S and the Harpending's raggedness index did not present significant signs of expansion or

retraction for any location (Table 4 and Fig. 4). Although the Tajima's D also revealed most localities in equilibrium between mutation and genetic drift, the Northeastern localities (Sabiaguaba and Alagoas) presented statistically significant negative D values ($p < 0.05$), suggesting a population expansion after a recent bottleneck.

Table 4. Neutrality tests and mismatch distribution based on the mtDNA.

Locality	*D*	F_S	*R*	*tau*	θ_0	θ_1
SBF	**-1.665****	-1.311	0.057	4.433	0.772	29.921
ALF	**-1.174***	-0.226	0.090	6.222	0.001	5.831
ACF	0.000	1.099	*β*	-	-	-
GAF	0.000	1.061	1	2.289	0	11.011
SJF	-0.131	-1.399	0.138	4	1	3,414.978
PGF	-0.452	0.612	0.150	9.246	0	13.119
DF	-1.177	-2.178	0.068	3.813	0.005	7.397
ARF	-0.825	-1.657	α	-	-	-
STF	-0.768	-2.128	α	-	-	-
RBF	-0.440	1.674	0.470	9.090	0	4.275
PIF	0.000	1.099	*β*	-	-	-

D - Tajima's D results; FS - Fu's F_S results; R - Harpending's Raggedness index; tau - demographic expansion factor; θ_0 – function of population size before expansion; θ_1 – function of population size after expansion. Values in bold are significant. Abbreviations as in Table 1.

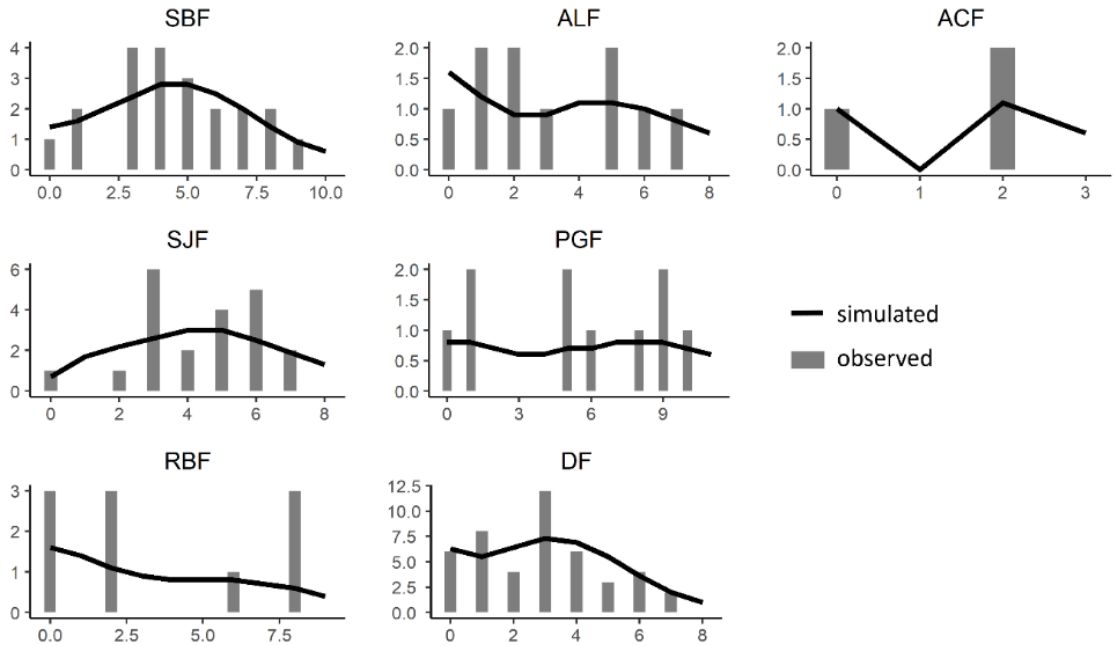
α - The last-squares procedure to fit model and observed distribution did not converge after 2,000 steps;

β – Number of samples is too small;

** Statistically significant values ($p < 0.05$)*

*** Statistically significant values ($p < 0.01$)*

Fig. 4. Mismatch distributions based on the mtDNA. The bars represent the observed pairwise distances, and the line represent the distribution expected under a model of recent demographic change. Abbreviations as in Table 1.



4.2.2 SNPs – dataset I

The number of polymorphic sites per location ranged from 7.79% (PIF) to 23.21% (SBF) (Table 5). The observed heterozygosity across all loci was significantly lower than the expected in 5 out of 11 localities, including three transects ($p < 0.05$). Most of the average F_{IS} values across loci were statistically significant within each locality, ranging from 0.084 (DF) to 0.167 (RBF). The lowest nucleotide differences was found in Santo Antônio ($\theta_S = 114.576$), whereas the highest belonged to Barra de São João ($\theta_S = 333.162$). For the nucleotide diversity (θ_π), the lowest value belonged to Praia de Gamboa ($\theta_\pi = 82.309$), and the highest to Ponta de Ribeirão da Ilha ($\theta_\pi = 241.161$). When grouped according to the Brazilian coast regions, the Southeastern localities exhibited the highest amount of polymorphic sites, followed by the Northeastern locations. All the regions presented significant heterozygote deficiency. The results of nucleotide differences and nucleotide diversity exhibited the same pattern: the higher values in the Southeastern localities, and the lowest, in the Southern.

Table 5. Genetic diversity indexes based on the 6,094 SNPs loci.

Locality	PS (%)	H_O	H_E	F_{IS}	θ_S	θ_π
SBF	23.210	0.161	0.176	**0.089**	305.591	196.135
ALF	16.201	0.223	0.262	0.097	296.362	234.222
ACF	18.032	**0.192**	0.214	0.051	118.133	83.961
GAF	18.515	**0.177**	0.212	**0.112**	117.233	82.309
SJF	22.884	0.169	0.194	**0.097**	333.162	229.664
PGF	21.478	**0.164**	0.190	**0.107**	270.034	182.257
DF	19.466	0.180	0.205	**0.084**	296.254	102.288
ARF	18.713	0.192	0.222	**0.140**	297.14	218.756
STF	15.503	0.196	0.260	**0.157**	114.576	90.072
RBF	16.831	**0.189**	0.236	**0.167**	160.054	119.902
PIF	7.795	**0.500**	0.540	0.047	243.812	241.161
NE	29.142	**0.127**	0.140	**0.091**	323.072	185.274
SE	44.241	**0.075**	0.086	**0.103**	193.854	84.741
S	27.067	**0.123**	0.147	**0.148**	106.018	61.802

PS = percentage of polymorphic sites ($p < 0.05$); H_O = observed heterozygosity; H_E = expected heterozygosity; F_{IS} = fixation index; θ_S = nucleotide differences; θ_π = nucleotide diversity; NE = Northeast; SE = Southeast; S = South. In bold, significant values of F_{IS} and differences among H_E and H_O ($p < 0.05$). Abbreviations as in Table 1.

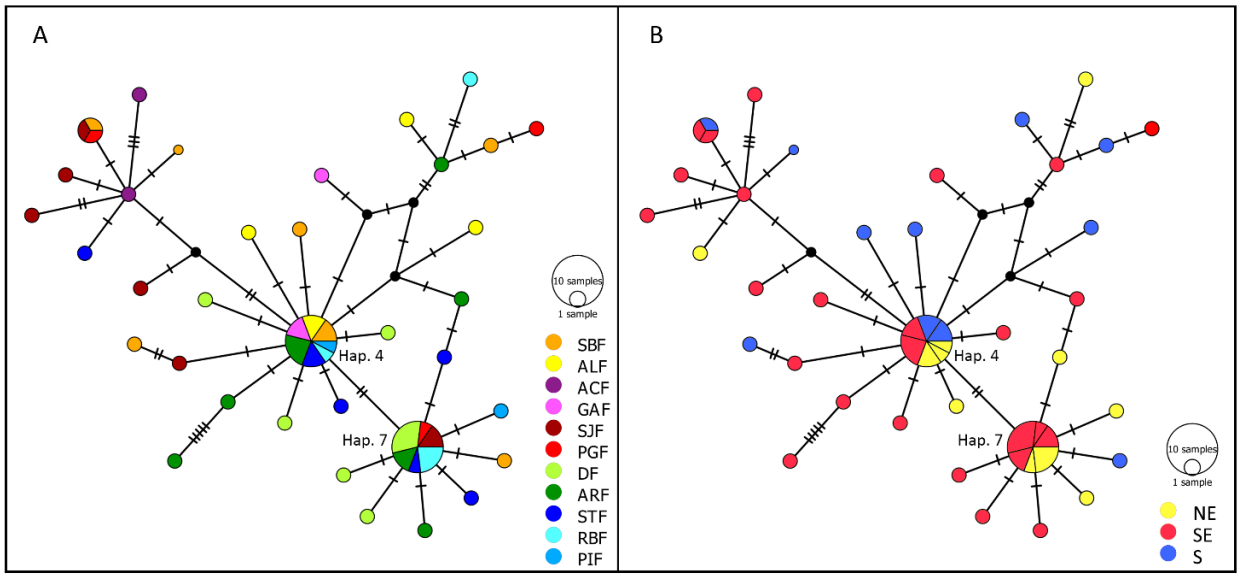
4.3 Population structure

4.3.1 Mitochondrial markers

The haplotype network contained 35 haplotypes with 46 polymorphic sites. The MSN revealed a structure with two central haplotypes (4 and 7) shared among several samples and singletons diverging by only a few substitutions (Fig. 5). Only three haplotypes (3, 4 and 7) contained more than two individuals (Table 4S). The most diverse haplotype (hap. 4) had individuals from 5 states (Fig. 5A). The mutational steps varied from one to five. The two haplotypes that diverged by five mutational steps are from the same location (ARF, São Paulo). There is not a clear geographic pattern on the distribution of the haplotypes across localities, even when the samples are grouped according to the regions (Fig. 5B). The Northeastern localities presented

the highest percentage of unique haplotypes (75%), while the South, the lowest percentage (~42%).

Fig. 5. Minimum spanning network (MSN) among *Littoraria flava* haplotypes. Each node represents a haplotype and the size reflects its frequency. The black nodes represent inferred ancestral nodes. The colors nodes indicates (A) the localities and (B) the regions along the Brazilian coast. Abbreviations as in Table 1.



When considering each location as a populational unit, the average fixation index F_{ST} was significant (F_{ST} = 0.077, F_{SC} = 0.076, $p < 0.05$), with 92.3% of the total variance within the locality (Table 6). The groups, i.e. the regions Northeast, Southeast and South, contained the lowest percentage of variation. When considering all individuals from a region as a populational unit, the AMOVA did not reveal a significant differentiation (F_{ST} = 0.024, p = 0.1283).

The pairwise F_{ST} among the 11 localities ranged from 0 to 0.52 (Fig. 6A and Table 5S), but only 6 out of 55 indexes were significantly different from zero, originated from the Northeast and Southeast regions. The pairwise F_{ST} calculated among the three regions did not exhibited any significant value (Table 5S). The Mantel's test was not significant for the correlation between genetic and geographic distance for both hierarchical levels considering the 11 localities (r^2 = 0.0007, p = 0.3527) or three regions (r^2 = -0.9937, p = 1) (Fig. 2S).

Table 6. AMOVA results from both mtDNA and SNPs markers of *Littoraria flava*. Each population represents a locality and each group correspond to a region from Brazilian coast (Northeast, Southeast and South).

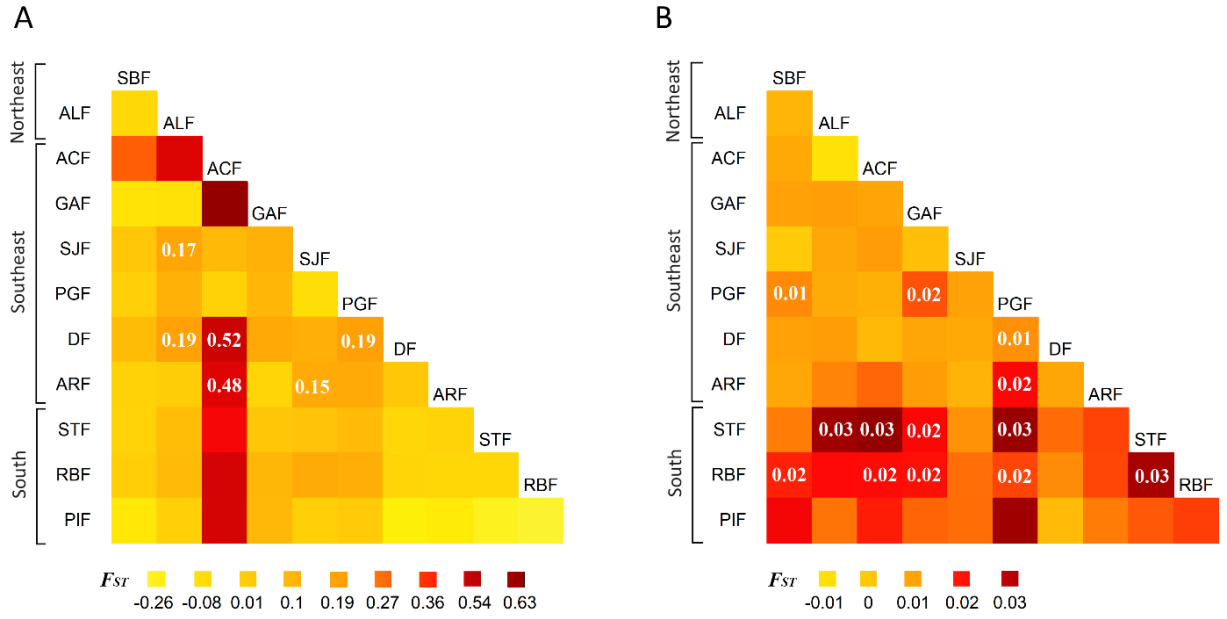
	mtDNA			ITS		
Source of Variation	df	SS	% Variation	df	SS	% Variation
Among groups	2	5.516	0.10	2	106.803	0.22
Among populations	8	20.807	7.57	8	391.906	0.40
Among individuals	53	93.519	92.32	159	6,627.279	99.38
Total	63	119.842		169	7,125.988	
	F_{ST} = 0.077*			F_{ST} = 0.014**		
	F_{SC} = 0.076*			F_{SC} = 0.011*		

d.f = degrees of freedom; SS = some of squares; % Variation = percentage of variation within each source.

** Statistically significant values (p < 0.05)*

*** Statistically significant values (p < 0.01)*

Fig. 6. Heatmaps of F_{ST} calculated from (A) mtDNA and (B) SNPs markers among sampled locations of *Littoraria flava*. The colour code illustrates the F_{ST} value. Only the significant indexes are shown (p < 0.05). Abbreviations as in Table 1.

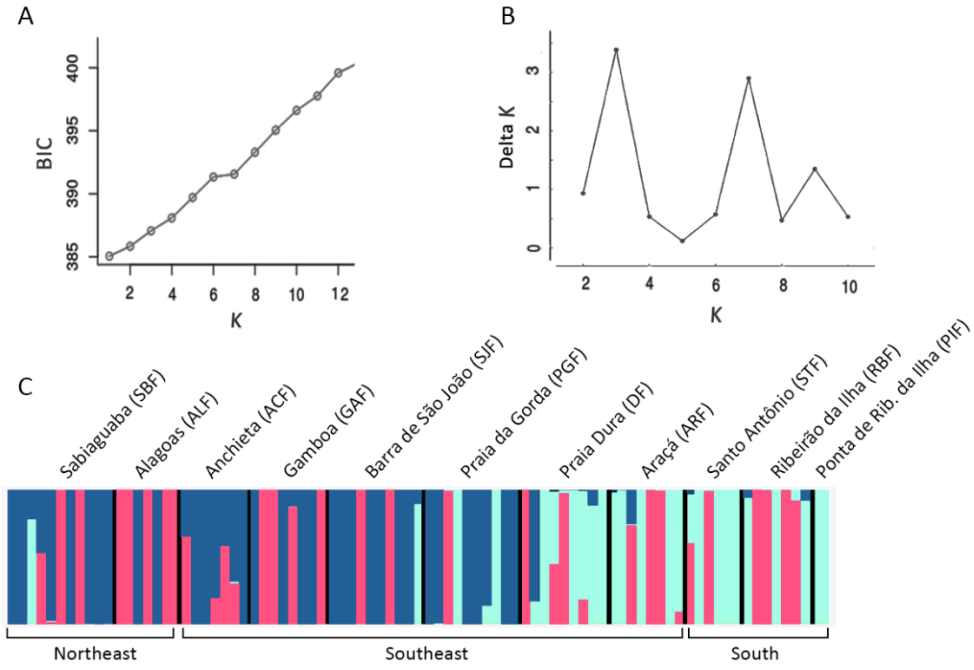


4.3.2 SNPs – dataset I

4.3.2.1 Among localities (macrogeographic scale)

The clustering approaches from STRUCTURE (K = 3) and DAPC (K = 1) did not produce the same patterns (Fig. 7A and 7B). Whereas the DAPC assigned all individuals into one single cluster, the STRUCTURE showed a clear distinction between three clusters: the first is mainly composed by individuals from the Northeast and Southeast regions; the second, by individuals from the Southeast and South regions; and the third, by individuals from the three regions, with no predominance of any of them (Fig. 7C). When assuming K = 3, the DAPC indicated that 227 SNPs with a contribution greater than 0.1%. The largest contribution found was ~0.46% (Fig. 3S).

Fig. 7. Inferred population structure based on 6,094 SNPs. (A) The Bayesian Information Criterion (BIC) for each value of K inferred by DAPC (K = 1). (B) The delta (Δ) value for each K inferred by STRUCTURE (K = 3). (C) Results of genetic assignment from STRUCTURE (K = 3). Each vertical bar corresponds to one individual's probability of belonging to the cluster with that color. The black lines separate the localities.



The global F_{ST} obtained from AMOVA revealed a low but significant structure when assuming each locality as a populational unit (F_{ST} = 0.014 and F_{SC} = 0.011, $p < 0.05$, Table 6). Most of the variation found was explained by differences between individuals within a locality (~99.38%), contrasting with the lowest variation contained among groups (0.22%). When assuming three populations, the fixation index F_{ST} was lower, but still significant (F_{ST} = 0.006, p = 0.008).

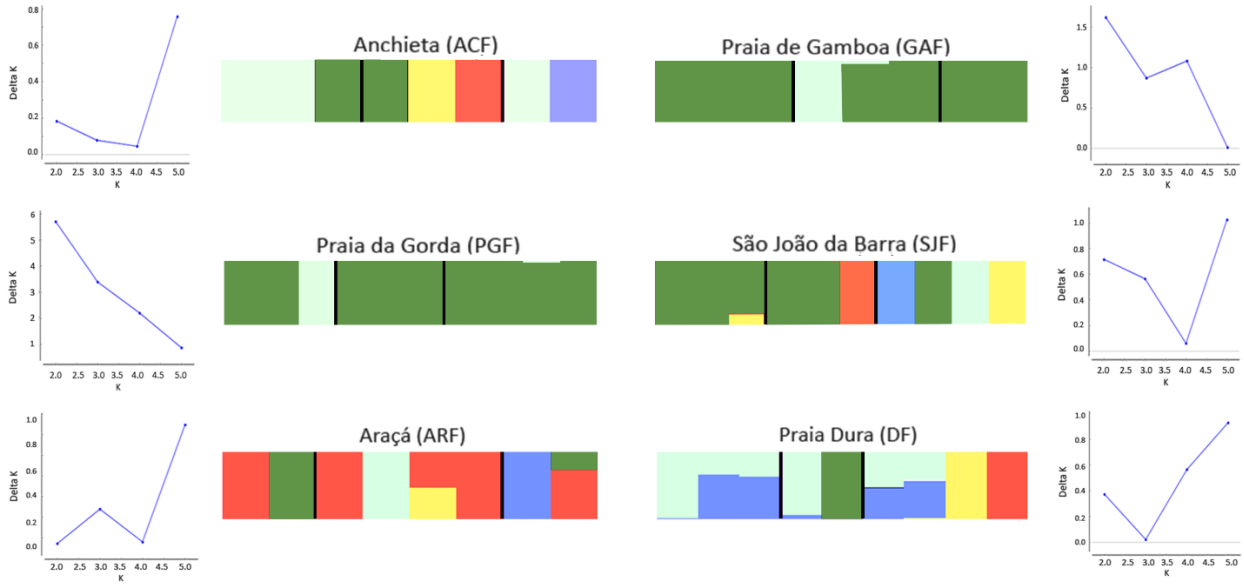
The pairwise F_{ST} among the 11 localities revealed 13 out of 55 statistically significant comparisons ($p < 0.05$, Fig. 6B and Table 5S). The analysis indicated STF and ACF as the most distant populations (F_{ST} = 0.03323, $p < 0.05$). Furthermore, SJF and PIF did not differ from any other population, whereas PGF showed the highest amount of significant F_{ST} values. The two Southern localities, RBF and STF, despite their geographical proximity, showed significant variation, being also the populations with highest frequencies of significant F_{ST} along with PGF. When grouped by region, the South differed significantly from the other regions, with the highest value between South and Northeast (F_{ST} = 0.01279, $p < 0.05$). The Mantel's test revealed that distribution of genetic variation did not correspond to the IBD model for both populational units tests (11 localities: r^2 = 0.1024, p = 0.2814 and 3 regions: r^2 = 0.5425, p = 0.3314) (Fig. 2S).

4.3.2.2 Within transects (microgeographic scale)

The STRUCTURE and DAPC clustering analyses based on SNPs from individuals sampled along transects showed no signs of genetic structure (Fig. 8). The DAPC identified one single cluster for all sites and was unable to differentiate the three sites when informed K = 3 *a priori*. The STRUCTURE detected two clusters (K = 2) for GAF and PGF and five clusters (K = 5) for ACF, ARF, DF and SJF. However, there was not a clear geographic pattern for any locality.

The pairwise F_{ST} among the sites within transects did not reveal any statistically significant variation (Table 6S). In other words, the individuals from the established sites on the rocky shore did not constitute subpopulations. The Mantel test was not significant ($p < 0.05$), therefore rejecting a possible distance isolation pattern within each transect (Fig. 4S).

Fig. 8. Genetic structuring results of micro-geographic scale analysis. The Delta (Δ) value inferred by STRUCTURE for each locality is on the side of its clustering plot. The black lines separate the three sites within the transect.

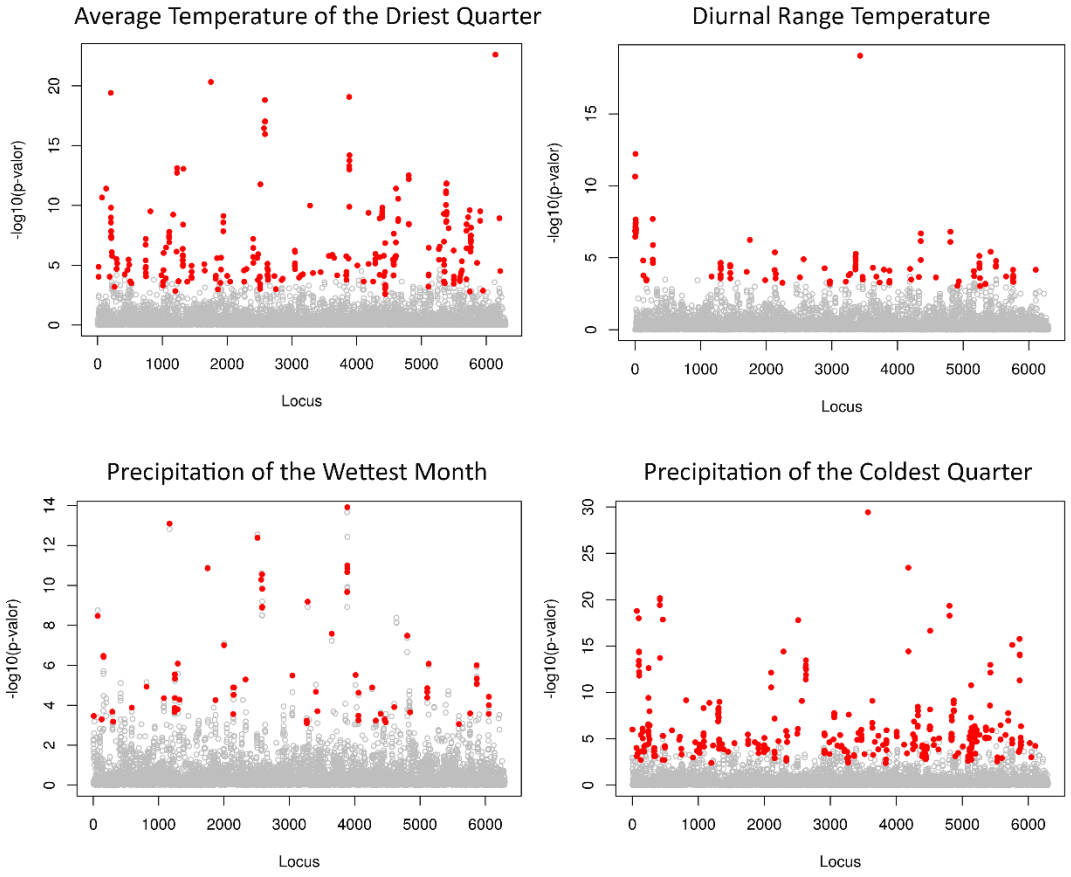


4.4 Seascape genetics analyses and putative adaptive loci – dataset II

The PCA indicated average temperature of the driest quarter, diurnal range temperature, precipitation of the wettest month and precipitation of the coldest quarter as the predictors that mostly explained the total variance across the study area (~92%, Fig. 5S). The LFMM identified 248 SNPs potentially associated with the average temperature of the driest quarter; 101 to the diurnal range temperature, 73 to the wettest month precipitation and 329 to precipitation of the coldest quarter ($p < 0.05$, Fig. 9 and Fig. 10A). The genome scan from BayeScan detected 78 SNPs potentially under selection (Fig. 6S). Overall, the SNPs identified by LEA, DAPC and BayeScan were within 433 loci, which 225 (~52%) were annotated (Table 7S). Many SNPs appeared for two or more predictors, in which 87 were exclusively associated with temperature and 93 to precipitation variation. The most common functions of loci with SNPs associated with temperature were cilia/flagella (loci 58991, 92017, 121390 and 121390), dynein (loci 20751, 128341 and 128341) and myosin/myogenesis (loci 1433, 19523, 66429 and 37040) (Table 7S). These same functions, however, appear only a few times in loci with SNPs associated with precipitation (myosin and dynein only two

times; cilia, one time), whereas zinc finger protein (loci 82767, 126785, 19244, 62977 and 68365) is the most common function. From the five loci commonly found in all environmental predictors (loci 20550, 32443, 45101, 46059 and 61753), only one was blasted (locus 46059), which seemed to play a role as a ribonuclease III.

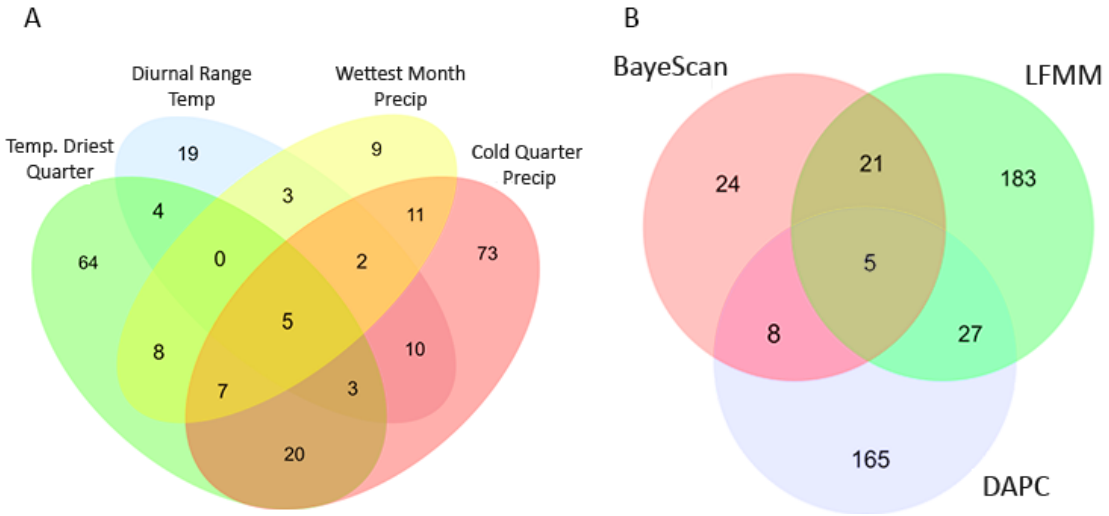
Fig. 9. Manhattan plot from the association test of individual SNPs in LEA. Each point on the x-axis represents a single SNP with its respective transformed p-value on the y-axis. Only SNPs with statistically significant values across all runs are red.



From the 58 blasted SNPs recognized by BayeScan, 26 were also found on LFMM (Fig. 10B). Many GOs associated with these common SNPs were dynein chain (loci 63409, 128341, 63409). From the SNPs identified exclusively in DAPC, the most common function were zinc finger protein (loci 126785, 63410, 47674, 62977 and 63211), followed by multiples GOs terms that appeared two times, as ubiquitin hydrolases (loci 62752 and 4300), vitamin D3 receptor (loci 10201 and 41207) and

centrosomal protein (loci 46508 and 6087), for example. The three methods implemented (BayeScan, DAPC and LFMM) commonly identified 5 outlier loci (loci 126785,12902, 44975, 63680 and 7063) (Fig. 10B), which blast attributed functions related to zinc finger protein, Rho GTPase, Tripartite motif-containing protein 3, NLR family CARD domain and Caprin-2 protein, respectively.

Fig. 10. Venn diagram of the intersection of loci containing putative adaptive SNPs identified on (A) LFMM for each environmental predictor selected by PCA and (B) on BayeScan, LFMM and DAPC.



5. Discussion

From our initial hypotheses, the data corroborated with the assumption of interconnected populations across large spatial distances due to high levels of gene flow. Nonetheless, there was not variation within transects according to the expected by the study of Andrade & Solferini (2007). By employing the Genotyping-by-Sequencing technic, we were able to genotype thousands of genetic markers to clarify the processes and environmental factors underlying the population structure and diversity of a non-model marine organism.

Demographic history

Some of the advantages of using mtDNA to infer populational events included (1) easy isolation and assay, (2) low cost, (3) a simple genetic structure lacking features such as repetitive DNA, transposable elements and introns, (4) rapid evolution and (5) absence of recombination (AVISE *et al.*, 1987). Its application is quite common in the investigation of population events, which assesses the evolutionary relationships between haplotypes (SLATKIN & HUDSON, 1991). Both analyses of the distribution of pairwise differences, i.e. Tajima's D, Fu's F_S and mismatch distribution, provided a method for inferring demographic events, such as population expansion and retraction (HARPENDING *et al.*, 1993; ROGERS & HARPENDING, 1992; SLATKIN & HUDSON, 1991). Our neutrality analysis revealed significant Tajima's D values for the Northeastern localities, consistent with an excess of rare variants probably due to population growth (FORD, 2002; RAMÍREZ-SORIANO *et al.*, 2008) (Table 4). On the other hand, the mismatch distribution and the Fu's results did not support the population expansion model. These divergences from different methods could be explained by the greater statistical power of the Fu's F_S test when compared to Tajima's (RAMÍREZ-SORIANO *et al.*, 2008) and by the most conservative method of the mismatch distribution (RAMOS-ONSINS & ROZAS, 2002). An alternative explanation for these results considers a scenario with range expansion rather than population (deme) expansion, where the demes would exchange a high number of migrants (MAGOULAS *et al.*, 2006).

The haplotypes distribution supports a high level of gene flow, where there are a few common haplotypes present in at least two regions of the Brazilian coast, and many rare haplotypes differing by one to five mutational steps (Fig. 5). A very similar pattern was reported for the bivalves *Mytilus californianus* and *Mytilus trossulus* (MARKO *et al*., 2010), which also lives in rocky shores and has planktonic development. Both results are consistent with a rapid demographic expansion in the past. Although our inferences from genetic data do not reveal the actual demographic processes or spatial distribution of *L. flava* populations in the past, the results suggest that the species has experienced some type of departure in the populational equilibrium. Because of the high degree of small-scale spatial heterogeneity in the

intertidal habitats, the demographic histories of species living in these environments might reflect responses to climate change over long periods (HELMUTH *et al.,* 2002). Nonetheless, to a deeper investigation about these predictions, a characterization of the intrapopulation coalescence time (i.e. time since the start of a population expansion) and past changes in effective population size (N_e) is required (e.g. CRANDALL *et al*., 2007; HURTADO *et al*., 2007; MARKO *et al*., 2010). Despite the great current knowledge of the biotic and abiotic factors influencing local abundance and distribution of species on rocky shores, the combining effects of these features over large temporal scales still remains enigmatic (HART & MARKO, 2010).

Small spatial scale variation (within transects)

None of the transects revealed significant variation among the stablished sites (Fig. 8), diverging from the results of Andrade & Solferini (2007), which have found more structuring on a microgeographical scale than on a large-scale. These divergences might be due the different mutation rates of genetic markers (ANNE, 2006; SCHLÖTTERER, 2004; SUNNUCKS, 2000) or because allozymes are usually involved in metabolic functions and therefore may be undergoing some type of natural selection, resulting in a variation on a microscale (CARINI & HUGHES, 2006; JANSON, 1987; JOHANNESSON *et al.*, 1995, 2004; JOHANNESSON & TATARENKOV, 1997; KRAMARENKO & SNEGIN, 2015; TATARENKOV & JOHANNESSON, 1999). Besides, genetic heterogeneity on a local level can be produced by the differences in the genetic composition of larvae that settles in an area; natural selection acting on larvae before the settlement; or variance in reproductive success among adults (HEDGECOCK, 1994; HEDGECOCK *et al*., 1994).

The lack of populational subdivision, from meters to 10 km, is common in species with a dispersal stage (GOLDSON *et al*., 2001; KYLE & BOULDING, 2000), which agrees with our findings and with the development of *L. flava*. The fine-scale genetic structure should reflect a complex set of features, such as life-history traits, larvae features and local adaptation (PALUMBI, 2003). An alternative explanation for the lack of subdivided populations among the established sites within transects is the number of sampled individuals here, which may not be representative enough to

capture the genetic variation on a small spatial scale, considering the abundance of the species on rocky shores.

Large scale variation and gene flow dynamic

By evaluating the structuring patterns of SNPs and mtDNA from the same samples, we were able to compare the resolution of each marker result. Both datasets produced structure patterns divergent to those previously found with allozymes (ANDRADE & SOLFERINI, 2007), showing how differently these markers behave in this species. Although the mtDNA-based analyses were clearly informative, especially for demographic history investigation, the results from SNPs set markers appear to have better accuracy, revealing some significant inferences that were not evident from mtDNA. This discrepancy can occur due to stochastic factors affecting mtDNA evolution, such as the distribution of the haplotypes following a post-bottleneck expansion considering the mtDNA inheritance mechanism (HOELZEL *et al.*, 2002; MOURA *et al.*, 2014).

Both mtDNA and SNPs, on macro and micro spatial scale, showed no significant correlation among geographic and genetic distances, i.e. our data did not correspond to the IBD model (Fig. 2S and Fig. 4S). Deviations from IBD are frequently reported for marine animals, where geographically distant locations present small or no genetic variation. This pattern might result from the combination of several abiotic and biotic features, which has a higher impact than the geographic distance *per se* (SIEGEL *et al.*, 2008; WHITE *et al.*, 2010). It could also result from a selection pressure during the larval stage or post-settlement period (JOHNSON & BLACK, 1984; SCHMIDT & RAND, 2001; SHIMA & SWEARER, 2009, 2010).

The Bayesian method implemented on STRUCTURE revealed to be more efficient than DAPC at detecting the subtle population clustering of *L. flava*. While the program detected three clusters (K = 3) along the Brazilian coast, the DAPC did not reveal any genetic structure (K = 1, Fig. 7). This contrasts with previous demonstrations where DAPC was more accurate in identifying distinct genetic groups (BENESTAN *et al.*, 2015; JOMBART *et al.*, 2010; KANNO *et al.*, 2011). Nonetheless, simulations showed that when correlated allele frequency model is assumed on STRUCTURE, low levels of population differentiation, with F_{ST} between 0.02 and 0.03,

are usually correctly identified (LATCH *et al.*, 2006), being a more efficient clustering method for shallow populational variation (PRITCHARD *et al.*, 2000; ROSENBERG *et al.*, 2005).

Although the mitochondrial markers revealed higher pairwise F_{ST} values than SNPs, both markers presented low amount of significant structuring (Fig. 6). This pattern may suggest large effective population sizes or high levels of genetic connectivity among locations (MARKO & HART, 2011), which is usually assumed for high-dispersal species (PALUMBI, 1994; BOHONAK, 1999). Nonetheless, we found significant heterozygote deficiency and high positive F_{IS} estimates in several localities (Table 5). These patterns were already reported for marine invertebrates (ADDISON & HART, 2005; COSTANTINI *et al.*, 2007; KNUTSEN *et al.*, 2003), including *L. flava* (ANDRADE *et al.*, 2005), which could be caused by natural selection, Wahlund effect, inbreeding or null alleles (DAVID *et al.*, 1997; RAYMOND *et al.*, 1997; WHITAKER, 2004; ZOUROS & FOLTZ, 1983).

Since we only removed SNPs in LD and did not perform a selection test over the dataset I, some regions experiencing natural selection may remain. However, even in that case, the outliers would be able to produce heterozygosity deficiency in a few SNPs, not on entire populations, as observed in our data (GAFFNEY, 1990; LEWONTIN & KRAKAUER, 1973). The Wahlund effect occurs when considering a pool of subpopulations with different genotype frequencies as a single population, generating heterozygote deficiency (CROW & KIMURA, 1970; HARTL & CLARK, 2006; WAHLUND, 1928). It could be a plausible explanation to our results if there was a chaotic recruitment of cohorts from different origins or if many breeding groups composed each population (ANDRADE & SOLFERINI, 2007). Inbreeding, else way, should result in a uniform heterozygote deficiency among populations in certain loci (GAFFNEY, 1990; LEWONTIN & KRAKAUER, 1973). Both inbreeding and Wahlund effects might not be maintained for many generations because of the highly dispersal larva of *L. flava*. However, even if these mechanisms are not entirely responsible for our results, we cannot exclude the possibility that they are in some way shaping the population dynamics of *L. flava* (ANDRADE *et al.*, 2003, 2005).

Null alleles are alleles that do not amplify during genotyping. Possible reasons may be deletion, polymorphisms in the sequence where the primer or the restriction enzyme should anneal, and triallelic sites. When originated by polymorphism, null alleles result in missing data. In this scenario, the heterozygotes would be indistinguishable from the expected homozygotes on most of the genotyping platforms, being counted as monomorphic sites (CARLSON *et al.*, 2006; CROOKS *et al.*, 2013). Monomorphic sites were found only in Anchieta (ACF). In addition, all the SNPs with missing data rates greater than 35% were removed since they did not match our quality criteria. Therefore, although the null allele hypothesis could fit in Anchieta, it is unlikely that the heterozygosity deficit of other populations is also produced by null alleles.

An alternative hypothesis for our findings is a metapopulation model, where the history of extinctions would produce low heterozygosity rates, and the frequent population turnover would decrease genetic variation among local populations (GILPIN, 1991; SMEDBOL *et al.*, 2002). The dynamic of a species can be described using metapopulation theory if (1) patches contain subpopulations rather aggregations formed only by the movement of individuals; (2) the dynamics of local populations are not synchronous; (3) patches are linked by dispersal with the possibility of fouding new populations and (4) there is a risk of local extinction of a patch (GRIMM *et al.*, 2003; HANSKI, 1999; KRITZER & SALE, 2010). Even though we did not find evidences of subpopulations according to our designed spatial scale, under a metapopulation scenario a patch could be composed of individuals genetically different on a temporal scale, i.e. originated from different settlements periods. Andrade & Solferini (2007) found significant temporal variation among individuals living in the same rocky shore. Moreover, based on the size of the individuals, previous work (unpublished data) also showed an asymmetric abundance of adults and juveniles of *L. flava* on distinct rocky shores during the same period, indicating asynchronous rates of new recruits arrival. If asynchronous colonization is true for this species, a local extinction at any time might also happen. However, only a deeper study involving small-scale systems during a certain period would be able to explicitly investigate the extinction and recolonization of local populations (SMEDBOL *et al.*,

2002). Therefore, despite we could not assume that the observed results actually mirror a metapopulation model, our data seems to corroborate with this hypothesis, where the interconnected populations are composed of heterogeneous larval and recruit cohorts, resulting in the unrecognizable pattern of local variation. In addition to the metapopulation model, other phenomena may produce the same results. As suggested by Andrade & Solferini (2007), the high and heterogeneous F_{IS} across populations could be observed if our samplings represent a small portion of the population and the reproductive rates are smaller than a unit, i.e. not totally panmictic population.

The significant variation found in the Southern localities (Table 5S) could be explained by oceanographic factors promoting larval retention in this region (URREGO-BLANCO & SHENG, 2014). Besides, it is possible that the larvae behavior may not be compatible with the oceanographic conditions, or that it may not survive during the transit on the water column (COWEN, 2000). However, only oceanographic modeling including the pelagic larval duration (PLD) and other seascape predictors would provide accurate evidences for these hypotheses (e.g. GALINDO *et al.*, 2010).

Seascape genetics

Detecting molecular markers of adaptive relevance can be done by (1) identifying polymorphisms with significantly higher genetic differentiation among populations than is expected under neutrality or (2) correlating the presence/absence of alleles to environmental data (HOLDEREGGER *et al.,* 2008). The first method, often called genome scan, uses simulations of neutral evolution to compute fixation index F_{ST} and then compare the results with the empirical data (LUIKART *et al.*, 2003; STORZ, 2005). This procedure does not provide information about an association of molecular markers with ecological factors unless the allele frequencies of outlier loci are correlated with environmental data *a posteriori*. The second approach provides direct clues of which ecological factor is acting as a selective force (LUIKART *et al.*, 2003). This method correlates the environmental data with an allele distribution model. In other words, it uses both genetic and geo-referenced environmental data of each individual. As the allele distribution is individual-based, the analysis is independent of sample size per location. Although this second method provides a more direct link

between molecular and environmental data, the pre-selection of variables potentially adaptive is not trivial and should include predictors that influence the species' survival and performance (HOLDEREGGER *et al.,* 2008). Here, we applied both methods for detecting outlier SNPs, which allowed identifying and combining the results to recognize SNPs that not only have a high degree of differentiation, but also a correlation with environmental variables.

Because linkage disequilibrium (LD) exists when the combination of two or more alleles is significantly greater than would be expected at random, any evolutionary mechanisms causing a departure from the equilibrium expectation will leave a signature of LD in the genome (BARTON *et al.*, 2007). Natural selection, for instance, can be detected through reduced diversity in specific genome regions due to a rapid fixation of a mutation and consequent decrease of the polymorphism on linked loci in its neighborhood (BARTON, 2011; MUSTONEN & LÄSSIG, 2009). From a genetic perspective, outlier loci should be identified before performing population inferences, since selection can create artifacts in estimates of migration rates, substructure and population differentiation (HOLDEREGGER *et al.,* 2008). For this reason, to avoid any bias in the results, the dataset submitted for population analyses did not include SNPs in LD, contrary to the dataset used in BayeScan and LFMM, which might increase the outlier detection.

Despite the inclusion of environment data collected at the sampling time — salinity, sea surface temperature, and associated fauna and flora — in the association tests, none of them were the most variable predictors throughout the study area. Instead, the highest environmental heterogeneity among the sampled localities included temperature and precipitation variation, which explained almost the entire variation across the sampled locations (~92%, Fig. 5S). Several studies have indicated that thermal stress varies in space and time, and that is an important determinant of organisms distribution in the intertidal zone (BUCKLEY *et al.*, 2001; MENGE *et al.*, 2007; STILLMAN & SOMERO, 2000; WETHEY, 1983; WILLIAMS & MORRITT, 1995). A long exposure to high temperatures can generate significant physiological consequences to animals (BUCKLEY *et al.*, 2001; DAHLHOFF *et al.*, 2001; HELMUTH & HOFMANN, 2001; ROBERTS *et al.*, 1997; SNYDER *et al.*, 2001;

TOMANEK & SOMERO, 1999). On the other hand, precipitation is closely related to the tidal height, which affects the wave action and consequent submersion and distribution of animals in rocky shore communities (BUSTAMANTE *et al.*, 1997; HARLEY & HELMUTH, 2003; WILLIAMS & MORRITT, 1995). The duration and levels of submersion result in a variety of potential stresses, including a higher rate of predation (ROBLES *et al.*, 2001), thermal stress due to emersion (ROBERTS *et al.*, 1997; TOMANEK & SOMERO, 2000) and feeding time (BAYNE *et al.*, 1988). Although animals living higher on the shore, such as members of the *Littorinoidea* superfamily, are usually more tolerant than those living nearer to water (e.g. BRITTON, 1995; DAVENPORT & DAVENPORT, 2005; FRAENKEL, 1968; MCMAHON, 1990), both thermal and rainfall variation can greatly affect the survival of adults and larvae (FUCHS *et al.*, 2010; MINTON & GOCHFELD, 2001; PRZESLAWSKI, 2005). Therefore, because of its biological relevance afore detailed, both predictors would be able to generate local adaptation (ACKERMAN *et al.*, 2013; MILANO *et al.*, 2014). In marine invertebrates, responses to complex mosaics involving other environmental features besides temperature and precipitation, such as salinity, substrate gradient, larval behavior and duration, predation and competition, for example (BENESTAN *et al*., 2016; BERGER & KHARAZOVA, 1997; COWDEN *et al*., 1984; MENGE, 2000; MURAEVA *et al*., 2016; RIASCOS *et al*., 2009; ROLÁN-ALVAREZ, 2007; SOKOLOVA & BOULDING, 2004), has been documented.

Since we consider as potential adaptive SNPs only those intersected across all the K-values from LFMM, we believe that the association test was able to capture signs of adaptation (FRANÇOIS *et al.*, 2016). Besides, the previously annotated transcriptome of *L. flava* used as a reference for BLAST here certainly improved the results accuracy, given its greater loci number and unigenes length, which may favor an increased of hits chance.

From the loci exclusively associated with temperature, the most common functions were related to cilia/flagella (Table 7S). In gastropods, these structures can be found in the foot epithelial tissue in order to promote better adhesion and mobility to the substrate, in addition to facilitate the movement through the mucus (TONAR & MARKOŠ, 2004). Futhermore, the veliger larvae stage, found in gastropods and other

mollusks, has a velum covered by long cilia that not only act helping to capture of food in suspension, but also in the locomotion across the water column (ROMERO *et al.*, 2010). Derived forms from myosin were also found in this dataset, but its functional relevance in mollusks is still poorly understood. Weiss *et al.* (2006) proposed that interactions of a specific unconventional myosin domain are likely involved in the complex regulation of mollusk shell formation, while other findings showed that myosin is also important for cytokinesis and polar lobe formation in embryos of gastropod *Ilyanassa obsoleta* (HEJNOL & PFANNENSTIEL, 1998). Transcripts obtained from viable embryos of the bivalve *Crepidula navicella* revealed numerous types of myosin bein expressed to developing retractor muscles (LESOWAY *et al.*, 2016), which corroborate to previous predictions about its functional significance in mollusks foot muscle (WATABE *et al.*, 1990). Because of the relevance of all these functions for the species survival, we can not exclude the possibility of potential selection acting in this SNPs. Unfortunately, we were not able to find any SNP associated with Heat Shock Proteins (HSP) in our data (HOFMANN, 1999; TOMANEK & SOMERO, 1999).

Some loci containing putative adaptive SNPs commonly identified by LFMM and BayeScan (Fig. 10) were associated with dynein. Interestingly, the dynein constitutes internal cytoskeletal structures called axonemes, which may present a role in ciliary and flagellar bend (ALBERTS *et al.*, 2002; GIBBONS, 1981; INABA, 2003), previously described to a littorinid (BUCKLAND-NICKS & CHIA, 1981). These features might be associated to ocelli microstructure (HOWARD & MARTIN, 1984), veliger locomotion, sensory organ and neural control (ARKETT *et al.*, 1987; BRAUBACH *et al.*, 2006; PAGE, 2002), adhesion to the substrate and movement on it (TONAR & MARKOŠ, 2004), feeding mechanisms (CHAPARRO *et al.*, 2002; ROMERO *et al.*, 2010) and sperm motility in gastropod (BOJAT *et al.*, 2002; SHIBA *et al.*, 2014). Although these mechanisms are not well-known in *L. flava*, it is possible that these features have been experiencing environmental adaptation and/or natural selection, since several studies have already mentioned their biological importance for the gastropod growth and survival (CHOI *et al.*, 2004; GERLACH, 2007; LIMA & PECHENIK, 1985; PRZESLAWSKI, 2004, 2005; SCHELTEMA, 1967; THIVAKARAN & KASINATHAN, 1990).

6. Conclusions

In this study, we applied two different molecular markers to uncover the genetic diversity patterns in a widespread marine gastropod, *Littoraria flava*. By employing NGS sequencing, we were able to genotype thousands of genetic markers to clarify the processes and environmental factors underlying the population structure and diversity of a non-model marine organism. This approach seems to open new research opportunities to better understand marine molecular evolution in rocky shore organisms. The use of SNPs produced genetic structuring patterns divergent from those found in a previous work using allozymes, showing how the neutral assumptions of genetic markers can change the results and consequent interpretation of the data. The same was observed using both mitochondrial markers, a coding and a non-coding gene.

Even though the mtDNA-based analyses were clearly informative, the inferences from SNPs data revealed some significant inferences that were not evident from mtDNA. The demographic history investigation showed signs of range expansion for only two sampled locations. There was not evidences of subpopulations according to the established sites within transects. Because of the low amount of significant F_{ST}, significant heterozygote deficiency and high positive F_{IS} estimates in several localities, the findings might reflect a gene flow dynamics according to the metapopulation model. However, other phenomena, such as natural selection, Wahlund effects and small sample size, could produce similar patterns.

According to the environmental association analyses, environmental predictors related to temperature and precipitation explained almost the entire variation across the study area. Because of its relevance in intertidal zones, both predictors would be able to generate local adaptation on rocky shores. A set of putative adaptive genes associated with the cilia and flagella movement were found, which might have great biological relevance for the survival and performance on the marine environment. In gastropods, these structures can play several important functions for both adults and larvae, such as adhesion and mobility to the substrate, capture of food in suspension during the larval phase and the locomotion across the water column. Although these mechanisms are not well-known in *L. flava*, it is possible that these features have been

experiencing environmental adaptation and/or natural selection, given their biological importance for the gastropod growth and survival.

This is the first Brazilian study with a seascape genetic approach in a littorinid. This methodology provided results that can be useful for several types of comparative investigations, including population genetics, demographic histories and local adaptation, for non-model or economically important organisms. Furthermore, by incorporating other biological and abiotic features into posterior analyses, the data obtained here can serve as a basis for deeper studies about the connectivity of marine species, which should increase the understanding of the mechanisms underlying the population dynamics in marine systems.

Resumo

A genética de paisagens marinhas dedica-se a entender como o movimento dos organismos afeta a conectividade das populações. Para espécies com larvas planctotróficas, essa não é uma tarefa trivial, pois as trajetórias e a duração dos estágios larvais são dificilmente previsíveis. Muitos desses grupos frequentemente revelam pouca ou nenhuma diferenciação genética populacional. Entretanto, estudos recentes identificaram algumas espécies que apresentam uma forte estruturação genética em macro e micro escalas espaciais. O presente estudo buscou entender os processos demográficos e fatores ambientais que moldam a dinâmica populacional de um organismo não modelo. Para tanto, foi utilizada a técnica *Genotyping-by-Sequencing* (GBS) para obtenção de polimorfismos de nucleotídeo único (*Single-Nucleotide-Polymorphism*, SNPs), e dois genes mitocondriais (mtDNA) de *Littoraria flava*. As amostras foram coletadas em 11 localidades distribuídas ao longo da costa brasileira, onde em seis foram feitos transectos horizontais. A análise de história demográfica usando mtDNA sugeriu expansão demográfica nas populações Sabiaguaba e Alagoas (Tajima's D = -1.665 e -1.174, respectivamente, *p-value* < 0,05). Com base em 6.094 SNPs, foram encontrados três grupos genéticos distintos nas populações amostradas (K = 3). Além disso, uma estrutura genética fraca, porém significativa, foi detectada para ambos os marcadores (mtDNA F_{ST} = 0,01353 e SNPs F_{ST} = 0,07675, p <0,05). Não foram detectados sinais de subestruturação entre os pontos dos transectos, divergindo dos reusltados encontrados com alozimas em trabalhos prévios. A maioria das populações revelou deficiência de heterozigotos com altos valores de F_{IS}. Apesar desses resultados parecerem refletir um fluxo gênico de acordo com um modelo de metapopulação, outros fenômenos seriam capaz de produzir os mesmos padrões. A análise de genética de paisagens indicou que variáveis relacionadas a temperatura e precipitação continham quase toda a heterogeneidade ambiental. Alguns loci potencialmente sob seleção parecem ter papéis importantes na locomoção larval, órgãos sensoriais, mobilidade espermática e adesão epitelial ao substrato. Apesar do pouco conhecimento sobre esses mecanismos em *L. flava*, dada a relevância funcional, esses caracteres poderiam estar sob seleção e/ou adaptação ambiental.

Palavras-chaves: Fluxo gênico, *Littorinidae*, metapopulação, NGS, *seascape genetics*

Abstract

Seascape genetics has been dedicated to understanding how the movement of organisms affects populations connectivity. For species with planktonic larvae, this is not a trivial task, since trajectories and duration of larval stages are hardly predictable. Many of these groups often reveal little or no genetic differentiation among populations. However, recent studies have identified species presenting a strong genetic structure on both large and small spatial scales. This study aimed to understand the demographic processes and environmental factors shaping the population dynamics of a non-model organism. With this purpose, we used Genotyping-by-Sequencing (GBS) to obtain Single-Nucleotide-Polymorphisms (SNPs) markers and two mitochondrial genes (mtDNA) of *Littoraria flava*. The samples were collected from 11 locations distributed along the Brazilian coast, where in six horizontal transects were designed. The demographic history analysis using mtDNA suggested demographic expansion in the Sabiaguaba and Alagoas populations (Tajima's D = -1.665 and -1.174, respectively, p-value < 0.05). Based on 6,094 SNPs markers, three distinct clusters across the sampled populations (K = 3) were found. Additionally, a weak but significant genetic structure was detected for both sets of markers (mtDNA F_{ST} = 0.01353 and SNPs F_{ST} = 0.07675, p < 0.05). There were no signs of substructure among the sites within transects, diverging from previous results using allozymes. Most populations revealed heterozygote deficiency with high values of F_{IS}. Despite the results that might reflect a gene flow according to a metapopulation model, other phenomena could produce the same patterns. The seascape genetic analyses indicated that predictors related to temperature and precipitation contained almost the entire environmental heterogeneity. Some loci potentially under selection appear to be important functions on larval locomotion, sensory organs, sperm mobility and epithelial adhesion to the substrate. Despite the poor knowledge about these mechanisms in *L. flava*, considering their functional relevance, these traits could be under environmental selection and/or adaptation.

Key-words: Gene flow, *Littorinidae*, metapopulation, NGS, seascape genetics.

References

ACKERMAN, MW *et al.* **Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment**. Conservation Genetics, v. 14, n. 2, 483–498, 2013.

ADDISON, JA & HART, MW. **Spawning, copulation and inbreeding coefficients in marine invertebrates**. Biology Letters, v. 1, n. 4, 450–453, 22 2005.

ALBERTS, B *et al.* **Molecular biology or the cell**. Molecular biology or the cell, 2002.

ANDRADE, SCS; MAGALHAES, CA & SOLFERINI, VN. **Patterns of genetic variability in Brazilian Littorinids (Mollusca): a macrogeographic approach**. Journal of Zoological Systematics and Evolutionary Research, v. 41, n. 4, 249–255, 2003.

ANDRADE, SCS; MEDEIROS, HF & SOLFERINI, VN. **Homogeneity test of Hardy-weinberg deviations In Brazilian littorinids: Evidence for selection?** Journal of Molluscan Studies, v. 71, n. 2, 167–174, 2005.

ANDRADE, SCS; NORENBURG, JL & SOLFERINI, VN. **Worms without borders: genetic diversity patterns in four Brazilian *Ototyphlonemertes* species (Nemertea, Hoplonemertea)**. Marine Biology, v. 158, n. 9, 2109–2124, 2011.

ANDRADE, SCS & SOLFERINI, VN. **Fine-scale genetic structure overrides macro-scale structure in a marine snail: nonrandom recruitment, demographic events or selection?.** Biological Journal of the Linnean Society, v. 91, n. 1, 23–36, 2007.

ANNE, C. **Choosing the right molecular genetic markers for studying biodiversity: from molecular evolution to practical aspects**. Genetica, v. 127, n. 1–3, 101–120, 2006.

ARKETT, SA; MACKIE, GO & SINGLA, CL. **Neuronal Control of ciliary locomotion in a gastropod veliger (*Calliostoma*)**. The Biological Bulletin, v. 173, n. 3, 513–526, 1987.

AVISE, JC *et al.* **Intraspecific phylogeography: the mitochondrial dna bridge between population genetics and systematics**. n. 18, 36, 1987.

AVISE, JC. **Molecular markers, natural history and evolution**. Springer Science & Business Media, 2012.

AVISE, JC. **Phylogeography: retrospect and prospect**. Journal of Biogeography, v. 36, n. 1, 3–15, 2009.

BAKER, P. *et al.* **Range-wide population structure and history of the northern quahog (*Merceneria merceneria*) inferred from mitochondrial DNA sequence data**. ICES Journal of Marine Science, v. 65, n. 2, 155–163, 2008.

BARTON, NH. **Estimating linkage disequilibria**. Heredity, v. 106, n. 2, 205–206, 2011.

BARTON, NH *et al.* **Evolution.** Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2007.

BAYNE, BL; HAWKINS, AJS & NAVARRO, E. **Feeding and digestion in suspension-feeding bivalve molluscs: the relevance of physiological compensations**. American Zoologist, v. 28, n. 1, 147–159, 1988.

BEAUMONT, AR. **Geographic variation in allele frequencies at three loci in *Chlamys opercularis* from Norway to the Brittany coast**. Journal of the Marine Biological Association of the United Kingdom, v. 62, n. 2, 243–261, 1982.

BENESTAN, L *et al.* **RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*)**. Molecular Ecology, v. 24, n. 13, 3299–3315, 2015.

BENESTAN, L *et al.* **Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*)**. Molecular Ecology, v. 25, n. 20, 5073–5092, 2016.

BERGER, VJ & KHARAZOVA, AD. **Mechanisms of salinity adaptations in marine molluscs**. Oceanographic Literature Review, v. 7, n. 45, 1155-1156, 1998.

BOHONAK, AJ. **Dispersal, gene flow, and population structure**. The Quarterly Review of Biology, v. 74, n. 1, 21–45, 1999.

BOJAT, N & SAUDER, U & HAASE, M. **Functional anatomy of the sperm storage organs in Pulmonata: the simple spermatheca of *Bradybaena fruticum* (Gastropoda, Stylommatophora)**. Zoomorphology, v. 121, n. 4, 243–255, 2002.

BRAUBACH, OR *et al.* **Neural control of the velum in larvae of the gastropod, *Ilyanassa obsoleta***. Journal of Experimental Biology, v. 209, n. 23, 4676–4689, 2006.

BRITTON, JC. T**he relationship between position on shore and shell ornamentation in two size-dependent morphotypes of *Littorina striata*, with an estimate of evaporative water loss in these morphotypes and in *Melarhaphe neritoides***. Developments in Hydrobiology, 1995, Dordrecht. Anais. Dordrecht: Springer Netherlands, 129–142, 1995.

BUCKLAND-NICKS, JA & CHIA, FS. **Locomotion of the filiform sperm of *Littorina* (Gastropoda, Prosobranchia)**. Cell and Tissue Research, v. 219, n. 1, 1981.

BUCKLEY, BA; OWEN, M & HOFMANN, GE. **Adjusting the thermostat: the threshold induction temperature for the heat-shock response in intertidal mussels (genus *Mytilus*) changes as a function of thermal history**. Journal of Experimental Biology, v. 204, n. 20, 3571–3579, 2001.

BUCKLIN, A. **Population genetic variation of *Calanus finmarchicus* in Icelandic waters: preliminary evidence of genetic differences between Atlantic and Arctic populations**. ICES Journal of Marine Science, v. 57, n. 6, 1592–1604, 2000.

BURFORD, MO *et al.* **Local adaptation of a marine invertebrate with a high dispersal potential: evidence from a reciprocal transplant experiment of the eastern oyster *Crassostrea virginica***. Marine Ecology Progress Series, v. 505, 161–175, 2014.

BUSTAMANTE, RH; BRANCH, GM & EEKHOUT, S. **Maintenance of an exceptional intertidal grazer biomass in south africa: subsidy by subtidal kelps**. Ecology, v. 76, n. 7, 2314–2329, 1995.

BUSTAMANTE, RH; BRANCH, GM & EEKHOUT, S. **The influences of physical factors on the distribution and zonation patterns of south african rocky-shore communities**. African Journal of Marine Science, v. 18, 1997.

CARINI, G & HUGHES, JM. **Subdivided population structure and phylogeography of an endangered freshwater snail, *Notopala sublineata* (Conrad, 1850) (Gastropoda: Viviparidae), in Western Queensland, Australia**. Biological Journal of the Linnean Society, v. 88, n. 1, 1–16, 2006.

CARLSON, CS *et al.* **Direct detection of null alleles in SNP genotyping data**. Human Molecular Genetics, v. 15, n. 12, 1931–1937, 2006.

CHAN, LM; BROWN, JL & YODER, AD. **Integrating statistical genetic and geospatial methods brings new power to phylogeography**. Molecular Phylogenetics and Evolution, v. 59, n. 2, 523–537, 2011.

CHAPARRO, OR; THOMPSON, RJ & PEREDA, SV. **Feeding mechanisms in the gastropod *Crepidula fecunda***. Marine Ecology Progress Series, v. 234, 171–181, 2002.

CHOI, YH *et al.* **Modelling *Deroceras reticulatum* (Gastropoda) population dynamics based on daily temperature and rainfall**. Agriculture, Ecosystems & Environment, v. 103, n. 3, 519–525, 2004.

CLEMENT, M *et al.* **TCS: estimating gene genealogies**. In: Parallel and Distributed Processing Symposium, International. 0184-0184, 2002.

CONOVER, O *et al.* **Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation**. Journal of Fish Biology, v. 69, n. sc, 21–47, 2006.

COSTANTINI, F; FAUVELOT, C & ABBIATI, M. **Fine-scale genetic structuring in *Corallium rubrum*: evidence of inbreeding and limited effective larval dispersal**. Marine Ecology Progress Series, v. 340, 109–119, 2007.

COWDEN, C; YOUNG, CM & CHIA, FS. **Differential predation on marine invertebrate larvae by two benthic predators**. Marine Ecology Progress Series, v. 14, 145–149, 1984.

COWEN, RK. **Connectivity of marine populations: open or closed?** Science, v. 287, n. 5454, 857–859, 2000.

COWEN, RK. **Scaling of connectivity in marine populations**. Science, v. 311, n. 5760, 522–527, 2006.

CRANDALL, ED *et al.* **Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods**. Molecular Ecology, v. 17, n. 2, 611–626, 2007.

CROOKS, L *et al.* **Identification of null alleles and deletions from snp genotypes for an intercross between domestic and wild chickens**. v. 3, n. 8, 1253–1260, 2013.

CROW, JF & KIMURA, M. **An introduction to population genetics theory.** An introduction to population genetics theory, 1970.

DAHLHOFF, EP; BUCKLEY, BA. & MENGE, BA. **Physiology of the rocky intertidal predator *Nucella Ostrina* along an environmental stress gradient**. Ecology, v. 82, n. 10, 2816–2829, 2001.

DAVENPORT, J & DAVENPORT, JL. **Effects of shore height, wave exposure and geographical distance on thermal niche width of intertidal fauna**. Marine Ecology Progress Series, v. 292, 41–50, 2005.

DAVID, P; DELAY, B & JARNE, P. **Heterozygosity and growth in the marine bivalve *Spisula ovalis*: testing alternative hypotheses**. Genetical Research, v. 70, n. 3, 215–223, 1997.

DE DONATO, M *et al.* **Genotyping-by-Sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using Next-Generation Sequencing**. PLoS ONE, v. 8, n. 5, e62137, 2013.

DE MEDEIROS, B & FARRELL, BD. **Whole-genome amplification in double-digest RADseq results in adequate libraries but fewer sequenced loci**. PeerJ, v. 6, e5089, 2018.

DIAS, GM; DUARTE, LFL & SOLFERINI, VN. **Low genetic differentiation between isolated populations of the colonial ascidian *Symplegma rubra* Monniot, C. 1972**. Marine Biology, v. 148, n. 4, 807–815, 2006.

DIBATTISTA, JD *et al.* **Seascape genomics reveals fine-scale patterns of dispersal for a reef fish along the ecologically divergent coast of Northwestern Australia**. Molecular Ecology, v. 26, n. 22, 6206–6223, 2017.

DOYLE, JJ & DOYLE, JL. **A rapid DNA isolation procedure for small quantities of fresh leaf tissue**, 1987.

EARL, A & VONHOLDT, BM. **STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method**. Conservation Genetics Resources, v. 4, n. 2, 359–361, 2012.

EATON, DA. **PyRAD: assembly of de novo RADseq loci for phylogenetic analyses**. Bioinformatics, v. 30, n. 13, 1844–1849, 2014.

EKBLOM, R & GALINDO, J. **Applications of next generation sequencing in molecular ecology of non-model organisms**. Heredity, v. 107, n. 1, 1–15, 2011.

ELSHIRE, R *et al.* **A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species**. PLoS ONE, v. 6, n. 5, e19379, 2011.

EXCOFFIER, L & LISCHER, H. **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows**. Molecular Ecology Resources, v. 10, n. 3, 564–567, 2010.

FOLL, M & GAGGIOTTI, O. **A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective**. Genetics, v. 180, n. 2, 977–993, 2008.

FOLMER, O *et al.* **DNA primers for amplification of mitochondrial cytochrome coxidase subunit I from diverse metazoan invertebrates**. 3, 294–299, 1994.

FORD, MJ. **Applications of selective neutrality tests to molecular ecology**. Molecular Ecology, v. 11, n. 8, 1245–1262, 2002.

FRAENKEL, G. **The heat resistance of intertidal snails at Bimini, Bahamas; Ocean Springs, Mississippi; and Woods Hole, Massachusetts**. Physiological Zoology, v. 41, n. 1, 1–13, 1968.

FRANÇOIS, O *et al.* **Controlling false discoveries in genome scans for selection**. Molecular Ecology, v. 25, n. 2, 454–469, 2016.

FRICHOT, E *et al.* **Testing for associations between loci and environmental gradients using Latent Factor Mixed Models**. Molecular Biology and Evolution, v. 30, n. 7, 1687–1699, 2013.

FRICHOT, E & FRANÇOIS, O. **LEA : An R package for landscape and ecological association studies**. Methods in Ecology and Evolution, v. 6, n. 8, 925–929, 2015.

FUCHS, H; SOLOW, A & MULLINEAUX, L. **Larval responses to turbulence and temperature in a tidal inlet: Habitat selection by dispersing gastropods?** Journal of Marine Research, v. 68, n. 1, 153–188, 2010.

FUNK, WC *et al.* **Harnessing genomics for delineating conservation units**. Trends in Ecology & Evolution, v. 27, n. 9, 489–496, 2012.

GAFFNEY, PM. **Enzyme heterozygosity, growth rate, and viability in *Mytilus edulis*: another look**. Evolution, v. 44, n. 1, 204–210, 1990.

GALINDO, HM *et al.* **Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal: Seascape genetics and clines**. Molecular Ecology, v. 19, n. 17, 3692–3707, 2010.

GERLACH, J. **Short-term climate change and the extinction of the snail *Rhachistia aldabrae* (Gastropoda: Pulmonata)**. Biology Letters, v. 3, n. 5, 581–585, 2007.

GIBBONS, IR. **Cilia and flagella of eukaryotes.** The Journal of Cell Biology, v. 91, n. 3, 107s–124s, 1 1981.

GILPIN, M. **The genetic effective size of a metapopulation**. Biological Journal of the Linnean Society, v. 42, n. 1–2, 165–175, 1991.

GOLDSON, AJ; HUGHES, RN & GLIDDON, CJ. **Population genetic consequences of larval dispersal mode and hydrography: a case study with bryozoans**. Marine Biology, v. 138, n. 5, 1037–1042, 2001.

GRIMM, V; REISE, K & STRASSER, M. **Marine metapopulations: a useful concept?** Helgoland Marine Research, v. 56, n. 4, 222–228, 2003.

GROSBERG, R & CUNNINGHAM, CW. **Genetic Structure in the Sea**. 25, 2001.

HANSKI, I. **Habitat Connectivity, habitat continuity, and metapopulations in dynamic landscapes**. Oikos, v. 87, n. 2, 209, 1999.

HARLEY, CDG & HELMUTH, BST. **Local and regional-scale effects of wave exposure, thermal stress, and absolute versus effective shore level on patterns of intertidal zonation**. Limnology and Oceanography, v. 48, n. 4, 1498–1508, 2003.

HARPENDING, HC *et al.* **The genetic structure of ancient human populations**. Current Anthropology, v. 34, n. 4, 483–496, 1993.

HART, MW & MARKO, PB. **It's about time: divergence, demography, and the evolution of developmental modes in marine invertebrates**. Integrative and Comparative Biology, v. 50, n. 4, 643–661, 2010.

HARTL, DL. & CLARK, AG. **Principles of population genetics**. 4. ed. Sinauer Associates, 2006.

HEDGECOCK, D. **Does variance in reproductive success limit effective population sizes of marine organisms**. Genetics and evolution of aquatic organisms, v. 122, 122–134, 1994.

HEDGECOCK, D; LI, G & NELSON, K. **The central stock of northern anchovy (*Engraulis mordax*) is not a randomly mating population**. v. 35, 16, 1994.

HEIPEL, DA; BISHOP, JD & BRAND, AR. **Mitochondrial DNA variation among open-sea and enclosed populations of the scallop *Pecten maximus* in western Britain**. Journal of the Marine Biological Association of the United Kingdom, v. 79, n. 4, 687–695, 1999.

HEJNOL, A & PFANNENSTIEL, HD. **Myosin and actin are necessary for polar lobe formation and resorption in *Ilyanassa obsoleta* embryos**. Development Genes and Evolution, v. 208, n. 4, 229–233, 1998.

HELLBERG, ME. **Gene flow and isolation among populations of marine Animals**. Annual Review of Ecology, Evolution, and Systematics, v. 40, n. 1, 291–310, 2009.

HELMUTH, BST *et al.* **Climate change and latitudinal patterns of intertidal thermal stress**. Science, v. 298, n. 5595, 1015–1017, 2002.

HELMUTH, BST & HOFMANN, GE. **Microhabitats, thermal heterogeneity, and patterns of physiological stress in the rocky intertidal zone**. The Biological Bulletin, v. 201, n. 3, 374–384, 2001.

HELYAR, S *et al.* **Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges: analytical approaches**. Molecular Ecology Resources, v. 11, 123–136, 2011.

HEREFORD, J. **A quantitative survey of local adaptation and fitness trade-offs**. The American Naturalist, v. 173, n. 5, 579–588, 2009.

HOELZEL, AR *et al.* **Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal: Impact of elephant seal bottleneck**. Journal of Evolutionary Biology, v. 15, n. 4, 567–575, 2002.

HOFMANN, GE. **Ecologically relevant variation in induction and function of heat shock proteins in marine organisms**. American Zoologist, v. 39, n. 6, 889–900, 1999.

HOLDEREGGER, R *et al.* **Land ahead: using genome scans to identify molecular markers of adaptive relevance**. Plant Ecology & Diversity, v. 1, n. 2, 273–283, 2008.

HOWARD, D & MARTIN, G. **Fine structure of the eyes of the interstitial gastropod *Fartulum orcutti* (Gastropoda, Prosobranchia)**. Zoomorphology, v. 104, n. 4, 197–203, 1984.

HURTADO, LA. *et al.* **Geographical subdivision, demographic history and gene flow in two sympatric species of intertidal snails, *Nerita scabricosta* and *Nerita funiculata*, from the tropical eastern Pacific**. Marine Biology, v. 151, n. 5, 1863–1873, 2007.

INABA, K. **Molecular architecture of the sperm flagella: molecules for motility and signaling**. Zoological science, v. 20, n. 9, 1043–1056, 2003.

JANSON, K. **Allozyme and shell variation in two marine snails (Littorina, Prosobranchia) with different dispersal abilities**. Biological Journal of the Linnean Society, v. 30, n. 3, 245–256, 1987.

JOHANNESSON, K *et al.* **Island isolation and habitat heterogeneity correlate with DNA variation in a marine snail (*Littorina saxatilis*): island isolation and habitat heterogeneity**. Biological Journal of the Linnean Society, v. 82, n. 3, 377–384, 2004.

JOHANNESSON, K; JOHANNESSON, B & LUNDGREN, U. **Strong natural selection causes microscale allozyme variation in a marine snail.** Proceedings of the National Academy of Sciences, v. 92, n. 7, 2602–2606, 1995.

JOHANNESSON, K & TATARENKOV, A. **Allozyme variation in a snail (*Littorina Saxatilis*) - deconfounding the effects of microhabitat and gene flow**. Evolution, v. 51, n. 2, 402–409, 1997.

JOHNSON, MS & BLACK, R. **Pattern beneath the chaos: The effect of recruitment on genetic patchiness in an intertidal limpet**. Evolution, v. 38, n. 6, 1371–1383, 1984.

JOMBART, T. **adegenet: a R package for the multivariate analysis of genetic markers**. Bioinformatics, v. 24, n. 11, 1403–1405, 2008.

JOMBART, T; DEVILLARD, S & BALLOUX, F. **Discriminant analysis of principal components: a new method for the analysis of genetically structured populations**. BMC Genetics, v. 11, n. 1, 94, 2010.

JOSÉ, J & SOLFERINI, VN. **Population genetics of *Collisella subrugosa* (Patellogastropoda: Acmaeidae): evidence of two scales of population structure**. Genetica, v. 130, n. 1, 73–82, 2007.

JOYEUX, JC *et al.* **Biogeography of tropical reef fishes: the South Atlantic puzzle: Biogeography of tropical reef fishes**. Journal of Biogeography, v. 28, n. 7, 831–841, 2008.

KANNO, Y; VOKOUN, J & LETCHER, B. **Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks**. Molecular Ecology, v. 20, n. 18, 3711–3729, 2011.

KARL, SA *et al.* **Common misconceptions in molecular ecology: echoes of the modern synthesis: eight misconceptions in molecular ecology**. Molecular Ecology, v. 21, n. 17, 4171–4189, 2012.

KAWECKI, TJ & EBERT, D. **Conceptual issues in local adaptation**. Ecology Letters, v. 7, n. 12, 1225–1241, 2004.

KING, PP & BRODERIP, WJ. **Description of Cirrhipeda, Conchifera and Mollusca, in a collection formed by the officers of H.M.S. Adventure and *Beagle* employed between the years 1826 and 1830 in surveying the southern coasts of South America, including the straits of Magulhaens and the coast of Tierra del Fuego**. Zoological Journal, v. 5, n. 19, 332–349, 1832.

KNOWLES, LL. **Statistical Phylogeography**. Annual Review of Ecology, Evolution, and Systematics, v. 40, n. 1, 593–612, 2009.

KNUTSEN, H. *et al.* **Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod**. Molecular Ecology, v. 12, n. 2, 385–394, 2003.

KRAMARENKO, SS & SNEGIN, EA. **Genetic structure of the continuous and ephemeral populations of the land snail *Brephulopsis cylindrica* (Gastropoda; Pulmonata; Enidae)**. Russian Journal of Genetics: Applied Research, v. 5, n. 5, 469–478, 2015.

KRITZER, JP. & SALE, PF. **Marine metapopulations**. Elsevier, 2010.

KUMAR, S; BRANCH, G & TAMURA, K. **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets**. Molecular Biology and Evolution, v. 33, n. 7, 1870–1874, 2016.

KYLE, CJ & BOULDING, EG. **Comparative population genetic structure of marine gastropods (*Littorina spp.*) with and without pelagic larval dispersal**. Marine Biology, v. 137, n. 5–6, 835–845, 2000.

LAL, M *et al.* **Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera***. Marine Genomics, v. 25, 57–68, 2016.

LATCH, E *et al.* **Relative performance of Bayesian clustering software for inferringpopulation substructure and individual assignment at low levels of population differentiation**. Conservation Genetics, v. 7, n. 2, 295–302, 2006.

LAUNEY, S. **Geographic structure in the European flat oyster (*Ostrea edulis* L.) as revealed by microsatellite polymorphism**. Journal of Heredity, v. 93, n. 5, 331–351, 2002.

LAZOSKI, C *et al.* **Cryptic speciation in a high gene flow scenario in the oviparous marine sponge *Chondrosia reniformis***. Marine Biology, v. 139, n. 3, 421–429, 2001.

LEE, HJ & BOULDING, EG. **Mitochondrial DNA variation in space and time in the northeastern Pacific gastropod, *Littorina keenae*.** Molecular Ecology, v. 16, n. 15, 3084–3103, 2007.

LEIGH, J & BRYANT, D. **popart : full-feature software for haplotype network construction**. Methods in Ecology and Evolution, v. 6, n. 9, 1110–1116, 2015.

LESOWAY, MP; ABOUHEIF, E & COLLIN, R. **Comparative transcriptomics of alternative developmental phenotypes in a marine gastropod: transcriptomics of nutritive embryos**. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, v. 326, n. 3, 151–167, 2016.

LEWONTIN, RC & KRAKAUER, J. **Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms**. Genetics, v. 74, n. 1, 175–195, 1973.

LIBRADO, P & ROZAS, J. **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data**. Bioinformatics, v. 25, n. 11, 1451–1452, 2009.

LIGGINS, L; TREML, EA & RIGINOS, C. **Seascape genomics: contextualizing adaptive and neutral genomic variation in the ocean environment**. In: Oleksiak M., Rajora O. (eds) Population Genomics: Marine Organisms. Population Genomics. Springer, Cham, 171–218, 2019.

LIMA, GM & PECHENIK, JA. **The influence of temperature on growth rate and length of larval life of the gastropod, *Crepidula plana* Say**. Journal of Experimental Marine Biology and Ecology, v. 90, n. 1, 55–71, 1985.

LIMBORG, MT *et al.* **Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*): environmental adaptation in Atlantic Hering**. Molecular Ecology, v. 21, n. 15, 3686–3703, 2012.

LISCHER, H & EXCOFFIER, L. **PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs**. Bioinformatics, v. 28, n. 2, 298–299, 2012.

LIU, H *et al.* **Population structuring and historical demography of a common clam worm *Perinereris aibuhitensis* near the coasts of Shandong Peninsula**. Biochemical Systematics and Ecology, v. 44, 70–78, 2012.

LUIKART, G *et al.* **The power and promise of population genomics: from genotyping to genome typing**. Nature Reviews Genetics, v. 4, n. 12, 981–994, 2003.

MAGOULAS, A *et al.* **Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*)**. Molecular Phylogenetics and Evolution, v. 39, n. 3, 734–746, 2006.

MANEL, S *et al.* **Landscape genetics: combining landscape ecology and population genetics**. Trends in Ecology & Evolution, v. 18, n. 4, 189–197, 2003.

MARKO, PB *et al.* **The 'Expansion-Contraction' model of Pleistocene biogeography: rocky shores suffer a sea change?** Molecular Ecology, v. 19, n. 1, 146–169, 2010.

MARKO, PB & HART, MW. **The complex analytical landscape of gene flow inference**. Trends in Ecology & Evolution, v. 26, n. 9, 448–456, 2011.

MCMAHON, RF. **Thermal tolerance, evaporative water loss, air-water oxygen consumption and zonation of intertidal prosobranchs: a new synthesis.** Developments in Hydrobiology, 1990, Dordrecht. Anais. Dordrecht: Springer Netherlands, 241–260, 1990.

MENGE, BA *et al.* **Mussel zonation in New Zealand: an integrative eco-physiological approach**. Marine Ecology Progress Series, v. 345, 129–140, 2007.

MENGE, BA. **Top-down and bottom-up community regulation in marine rocky intertidal habitats**. Journal of Experimental Marine Biology and Ecology, v. 250, n. 1–2, 257–289, 2000.

MILANO, I *et al.* **Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius)***. Molecular Ecology, v. 23, n. 1, 118–135, 2014.

MINTON, D & GOCHFELD, DJ. **Is life on a tropical shore really so hard?: the role of abiotic factors in structuring a supralittoral molluscan assemblage**. Journal of Shellfish Research, v. 20, n. 1, 477–483, 2001.

MIRANDA, AA; ALMEIDA, ACS & VIEIRA, LM. **Non-native marine bryozoans (Bryozoa: Gymnolaemata) in Brazilian waters: Assessment, dispersal and impacts**. Marine Pollution Bulletin, v. 130, 184–191, 2018.

MOREIRA, AA; TOMÁS, ARG & HILSDORF, AW. **Evidence for genetic differentiation of *Octopus vulgaris* (Mollusca, Cephalopoda) fishery populations from the southern coast of Brazil as revealed by microsatellites**. Journal of Experimental Marine Biology and Ecology, v. 407, n. 1, 34–40, 2011.

MOURA, AE *et al.* **Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift**. Molecular Ecology, v. 23, n. 21, 5179–5192, 2014.

MURAEVA, OA *et al.* **Mechanisms of adaption to salinity stress in marine gastropods *Littorina saxatilis*: a proteomic analysis**. Cell and Tissue Biology, v. 10, n. 2, 160–169, 2016.

MUSTONEN, V & LÄSSIG, M. **From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation**. Trends in Genetics, v. 25, n. 3, 111–119, 2009.

NAVARRETE, SA *et al.* **Scales of benthic-pelagic coupling and the intensity of species interactions: From recruitment limitation to top-down control**. Proceedings of the National Academy of Sciences, v. 102, n. 50, 18046–18051, 2005.

NEI, M. **Molecular evolutionary genetics**. Columbia university press, 1987.

NÓBREGA, R; SOLÉ-CAVA, AM & RUSSO, CA. **High genetic homogeneity of an intertidal marine invertebrate along 8000 km of the Atlantic coast of the Americas**. Journal of Experimental Marine Biology and Ecology, v. 303, n. 2, 173–181, 2004.

NOSIL, P; FUNK, DJ & ORTIZ-BARRIENTOS, D. **Divergent selection and heterogeneous genomic divergence**. Molecular Ecology, v. 18, n. 3, 375–402, 2009.

PAGE, LR. **Comparative structure of the larval apical sensory organ in gastropods and hypotheses about function and developmental evolution**. Invertebrate Reproduction & Development, v. 41, n. 1–3, 193–200, 2002.

PAIVA, PC *et al.* **Comparative phylogeography of two coastal species of *Perinereis* Kinberg, 1865 (Annelida, Polychaeta) in the South Atlantic**. Marine Biodiversity, v. 49, n. 3, 1537–1551, 2019.

PALUMBI, SR; MARTIN, A & ROMANO, S. **The simple fool's guide to PCR**. The simple fool's guide to PCR, v. 2, 28, 1991.

PALUMBI, SR. **Genetic divergence, reproductive isolation, and marine speciation**. Annual Review of Ecology and Systematics, v. 25, n. 1, 547–572, 1994.

PALUMBI, SR. **Population genetics, demographic connectivity, and the design of marine reserves**. Ecological Applications, v. 13, n. sp1, 146–158, 2003.

PRITCHARD, JK; STEPHENS, M & DONNELLY, P. **Inference of population structure using multilocus genotype data**. 15, 2000.

PRZESLAWSKI, R. **A review of the effects of environmental stress on embryonic development within intertidal gastropod egg masses**. Molluscan Research, v. 24, n. 1, 43, 2004.

PRZESLAWSKI, R. **Combined effects of solar radiation and desiccation on the mortality and development of encapsulated embryos of rocky shore gastropods**. Marine Ecology Progress Series, v. 298, 169–177, 2005.

PURCELL, S *et al.* **PLINK: A tool for whole-genome association and population-based linkage analyses**. The American Journal of Human Genetics, v. 81, n. 3, 559–575, 2007.

R CORE TEAM. **R development core team**. RA Lang Environ Stat Comput, v. 55, 275–286, 2013.

RAMÍREZ-SORIANO, A *et al.* **Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination**. Genetics, v. 179, n. 1, 555–567, 2008.

RAMOS-ONSINS, SE & ROZAS, J. **Statistical properties of new neutrality tests against population growth**. Molecular Biology and Evolution, v. 19, n. 12, 2092–2100, 2002.

RAYMOND, M *et al.* **Heterozygote deficiency in the mussel *Mytilus edulis* species complex revisited**. Marine Ecology Progress Series, v. 156, 225–237, 1997.

REID, DG. **The genus *Littoraria* Griffith & Pidgeon, 1834 (Gastropoda: Littorinidae) in the Tropical Eastern Pacific. El género Littoraria Griffith & Pidgeon, 1834 (Gastropoda: Littorinidae) en el Pacífico oriental tropical.** The Veliger., v. 42, n. 1, 21–53, 1999.

REID, DG. **Littorinid molluscs of mangrove forests in the Indo-Pacific region**. 1986.

REINKE, EE. **Report upon the dimorphic spermatozoa of marine prosobranchs**. Carnegie Institution of Washington Year Book, v. 10, 133–136, 1911.

RELLSTAB, C *et al.* **A practical guide to environmental association analysis in landscape genomics**. Molecular Ecology, v. 24, n. 17, 4348–4370, 2015.

RIASCOS, JM *et al.* **Thriving and declining: climate variability shaping life-history and population persistence of *Mesodesma donacium* in the Humboldt Upwelling System**. Marine Ecology Progress Series, v. 385, 151–163, 2009.

ROBERTS, DA; HOFMANN, GE & SOMERO, GN. **Heat-shock protein expression in *Mytilus californianus* : acclimatization (seasonal and tidal-height comparisons) and acclimation effects**. The Biological Bulletin, v. 192, n. 2, 309–320, 1997.

ROBLES, CD; ALVARADO, MA & DESHARNAIS, RA. **The shifting balance of littoral predator-prey interaction in regimes of hydrodynamic stress**. Oecologia, v. 128, n. 1, 142–152, 2001.

ROCHA, LA *et al.* **Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae)**. Molecular Ecology, v. 11, n. 2, 243–251, 2002.

ROCHA, LA. **Patterns of distribution and processes of speciation in Brazilian reef fishes**. Journal of Biogeography, v. 30, n. 8, 1161–1171, 2003.

ROCHA, R; KREMER, L & FEHLAUER-ALE, K. **Lack of COI variation for *Clavelina oblonga* (Tunicata, Ascidiacea) in Brazil: Evidence for its human-mediated transportation?** Aquatic Invasions, v. 7, n. 3, 419–424, 2012.

ROGERS, AR & HARPENDING. **Population growth makes waves in the distribution of pairwise genetic differences.** Molecular Biology and Evolution, 1992.

ROLÁN-ALVAREZ, E. **Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone**. Journal of Molluscan Studies, v. 73, n. 1, 1–10, 2007.

ROMERO, MR; KELSTRUP, HC & STRATHMANN, RR. **Capture of particles by direct interception by cilia during feeding of a gastropod veliger**. The Biological Bulletin, v. 218, n. 2, 145–159, 2010.

ROSENBERG, N *et al.* **Clines, clusters, and the effect of study design on the inference of human population structure**. PLoS Genetics, v. 1, n. 6, e70, 2005.

RYNEARSON, T & ARMBRUST, V. **Genetic differentiation among populations of the planktonic marine diatom *Ditylum Brightwellii* (bacillariophyceae): genetically distinct diatom populations**. Journal of Phycology, v. 40, n. 1, 34–43, 2004.

SANDOVAL-CASTILLO, J *et al.* **Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient**. Molecular Ecology, v. 27, n. 7, 1603–1620, 2018.

SCHELTEMA, RS. **The relationship of temperature to the larval development of *Nassarius obsoletus* (Gastropoda)**. The Biological Bulletin, v. 132, n. 2, 253–265, 1967.

SCHLÖTTERER, C. **The evolution of molecular markers — just a matter of fashion?** Nature Reviews Genetics, v. 5, n. 1, 63–69, 2004.

SCHMIDT, PS. & RAND, DM. **Adaptive maintenance of genetic polymorphism in an intertidal barnacle: habitat and life-stage-specific survivorship of Mpi genotypes**. Evolution, v. 55, n. 7, 1336–1344, 2001.

SEEB, J *et al.* **Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms**. Molecular Ecology Resources, v. 11, n. s1, 1–8, 2011.

SEIXAS, VC; PAIVA, PC & RUSSO, CAM. **Comparative population genetics and demographic history of two polychaete species suggest that coastal lagoon populations evolve under alternate regimes of gene flow**. Marine Biology, v. 165, n. 11, 179, 2018.

SELKOE, KA *et al.* **A decade of seascape genetics: contributions to basic and applied marine connectivity**. Marine Ecology Progress Series, v. 554, 1–19, 2016.

SELKOE, KA; HENZLER, CM & GAINES, SD. **Seascape genetics and the spatial ecology of marine populations**. Fish and Fisheries, v. 9, n. 4, 363–377, 2008.

SHIBA, K; SHIBATA, D & INABA, K. **Autonomous changes in the swimming direction of sperm in the gastropod *Strombus luhuanus***. Journal of Experimental Biology, v. 217, n. 6, 986–996, 2014.

SHIMA, JF & SWEARER, SE. **Spatially variable larval histories may shape recruitment rates of a temperate reef fish**. Marine Ecology Progress Series, v. 394, 223–229, 2009.

SHIMA, JF & SWEARER, SE. **The legacy of dispersal: larval experience shapes persistence later in the life of a reef fish**. Journal of Animal Ecology, v. 79, n. 6, 1308–1314, 2010.

SIEGEL, DA *et al.* **The stochastic nature of larval connectivity among nearshore marine populations**. Proceedings of the National Academy of Sciences, v. 105, n. 26, 8974–8979, 2008.

SLATKIN, M & HUDSON, RR. **Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations.** Genetics, v. 129, n. 2, 555–562, 1991.

SMEDBOL, R *et al.* **Myths and moderation in marine "metapopulations"?** Fish and Fisheries, v. 3, n. 1, 20–35, 2002.

SNYDER, MJ; GIRVETZ, E & MULDER, EP. **Induction of marine mollusc stress proteins by chemical or physical stress**. Archives of Environmental Contamination and Toxicology, v. 41, n. 1, 22–29, 2001.

SOKOLOVA, IM. & BOULDING, EG. **A neutral DNA marker suggests that parallel physiological adaptations to open shore and salt marsh habitats have evolved more than once within two different species of gastropods**. Marine Biology, v. 145, n. 1, 2004.

STAPLEY, J *et al.* **Adaptation genomics: the next generation**. Trends in Ecology & Evolution, v. 25, n. 12, 705–712, 2010.

STILLMAN, JH & SOMERO, GN. **A comparative analysis of the ipper thermal tolerance limits of Eastern Pacific porcelain crabs, genus *Petrolisthes* : Influences of latitude, vertical zonation, acclimation, and phylogeny**. Physiological and Biochemical Zoology, v. 73, n. 2, 200–208, 2000.

STORFER, A; PATTON, A & FRAIK, A. **Navigating the interface between landscape genetics and landscape genomics**. Frontiers in Genetics, v. 9, 68, 2018.

STORZ, JF. **Invited review: Using genome scans of DNA polymorphism to infer adaptive population divergence**. Molecular Ecology, v. 14, n. 3, 671–688, 2005.

SUNNUCKS, P. **Efficient genetic markers for population biology**. Trends in Ecology & Evolution, v. 15, n. 5, 199–203, 2000.

TAJIMA, F. **Evolutionary relationship of DNA sequences in finite populations**. 24, 1983.

TAJIMA, F. **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** Genetics, v. 123, n. 3, 585–595, 1989.

TATARENKOV, A & JOHANNESSON, K. **Micro and macrogeographic allozyme variation in *Littorina fabalis*; do sheltered and exposed forms hybridize?** Biological Journal of the Linnean Society, v. 67, n. 2, 199–212, 1999.

TAYLOR, MS. **genetic evidence for local retention of pelagic larvae in a Caribbean reef fish**. Science, v. 299, n. 5603, 107–109, 2003.

THIVAKARAN, GA & KASINATHAN, R. **Salinity, temperature and desiccation tolerance of intertidal gastropods *Littorina quadricentus* and *Nodilittorina pyramidalis***. 4, 1990.

TOMANEK, L & SOMERO, GN. **Heat-shock proteins, thermotolerance and biogeography**. The Journal of Experimental Biology, v. 202, 2925–2936, 1999.

TOMANEK, L & SOMERO, GN. **Time course and magnitude of synthesis of heat-shock proteins in congeneric marine snails (Genus *Tegula*) from different tidal heights**. Physiological and Biochemical Zoology, v. 73, n. 2, 249–256, 2000.

TONAR, Z & MARKOŠ, A. **Microscopy and morphometry of integument of the foot of pulmonate gastropods *Arion rufus* and *Helix pomatia***. Acta Veterinaria Brno, v. 73, n. 1, 3–8, 2004.

URREGO-BLANCO, J & SHENG, J. **Formation and distribution of sea ice in the Gulf of St. Lawrence: A process-oriented study using a coupled ocean-ice model**. Journal of Geophysical Research: Oceans, v. 119, n. 10, 7099–7122, 2014.

UTHICKE, S & BENZIE, JA. **Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific**. Molecular Ecology, v. 12, n. 10, 2635–2648, 2003.

VAN WYNGAARDEN, M *et al.* **Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus,* using RADseq-derived SNPs**. Evolutionary Applications, v. 10, n. 1, 102–117, 2017.

VIEIRA, LM *et al.* **Bryozoan biodiversity in Saint Peter and Saint Paul Archipelago, Brazil**. 10, 2012.

WAHLUND, S. **Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet**. Hereditas, v. 11, n. 1, 65–106, 1928.

WAPLES, RS. **Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species**. Journal of Heredity, v. 89, n. 5, 438–450, 1998.

WATABE, S *et al.* **Phosphorylation and immunological cross-reactivity of paramyosin: A comparative study**. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry, v. 96, n. 1, 81–88, 1990.

WEIR, BS & COCKERHAM, CC. **Estimating *F-Statistics* for the analysis of population structure**. Evolution, v. 38, n. 6, 1358–1370, 1984.

WEISS, IM *et al.* **The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain**. FEBS Letters, v. 580, n. 7, 1846–1852, 2006.

WETHEY, DS. **Geographic limits and local zonation: the barnacles *Semibalanus* (*Balanus*) and *Chthamalus* in new england**. The Biological Bulletin, v. 165, n. 1, 330–341, 1983.

WHITAKER, K. **Non-random mating and population genetic subdivision of two broadcasting corals at Ningaloo Reef, Western Australia**. Marine Biology, v. 144, n. 3, 593–603, 2004.

WHITE, C *et al.* **Ocean currents help explain population genetic structure**. Proceedings of the Royal Society B: Biological Sciences, v. 277, n. 1688, 1685–1694, 2010.

WILLIAMS, GA & MORRITT, D. **Habitat partitioning and thermal tolerance in a tropical limpet, *Cellana grata***. Marine Ecology Progress Series, v. 124, 89–103, 1995.

WIT, PD & PALUMBI, SR. **Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation**. Molecular Ecology, v. 22, n. 11, 2884–2897, 2013.

YAMADA, SB. **Are direct developers more locally adapted than planktonic developers?** Marine Biology, v. 103, n. 3, 403–411, 1989.

YEAMAN, S & WHITLOCK, MC. **The genetic architecture of adaptation under migration-selection balance: the genetic architecture of local adaptation**. Evolution, v. 65, n. 7, 1897–1911, 2011.

ZHBANNIKOV, IY *et al.* SeqyClean: A **Pipeline for high-throughput sequence data preprocessing**. In: the 8th ACM international Conference, 2017, Boston, Massachusetts, USA. Anais.. Boston, Massachusetts, USA: ACM Press, 407–416, 2017.

ZOUROS, E & FOLTZ, DW. **Minimal selection requirements for the correlation between heterozygosity and growth, and for the deficiency of heterozygotes, in oyster population**. Developmental Genetics, v. 4, n. 4, 393–405, 1983.

Supplementary Material

Table 1S. List of environmental variables submitted to PCA. The variables are ordered according to its source.

BioClim
Average annual temperature
Diurnal range temperature
Isothermality
Seasonal temperature
Maximum temperature of the warmest month
Minimum temperature of the coldest month
Temperature by annual interval
Average temperature of the wettest quarter
Average temperature of the driest quarter
Average temperature of the warmest quarter
Average temperature of the coldest quarter
Annual precipitation
Precipitation of wettest month
Precipitation of driest month
Seasonal precipitation
Quarter with highest precipitation
Precipitation of the driest quarter
Precipitation of warmest quarter
Precipitation of coldest quarter
Bio-Oracle
Chlorophyll A
pH
Salinity (mean)
Temperature Sea surface
Current velocity (average and variation)
Salinity of the sea surface
Dissolved oxygen
Collected Data
Salinity of the sea surface
Sea surface temperature
Associated fauna
Associated flora

Table 2S. Reads of *Littoraria flava* in each lane before and after the Seqyclean pipeline. The species identification used the barcode sequences attributed for each individual.

	Total reads	**Retained reads (%)**
Lane1	42,647,989	21,807,771 (51.13%)
Lane2	126,211,485	113,626,000 (90.02%)
Lane6	22,597,941	21,807,771 (95.50%)
Lane14	153,618,649	70,139,428 (45.66%)
Total	322,479,123	227,389,910 (70.51%)

Total reads – reads from Littoraria flava within each lane; Retained reads – reads from Littoraria flava retained after the Seqyclean pipeline application.

Table 3S. Individual reads number of the remaining 85 *Littoraria flava* samples after the filtering steps of Seqyclean and iPyrad. Abbreviations as in Table 1.

Sample	Number of reads	Sample	Number of reads	Sample	Number of reads
ACF4.2	522,628	DF64.2	1,938,143	RBF6	322,911
ACF4.3	349,474	DF64.3	3,567,039	RBF7	2,010,401
ACF4.4	338,329	DF64.4	6,499,214	SBF10	3,219,396
ACF64.1	858,509	GAF0.1	3,030,427	SBF11	3,732,789
ACF64.2	1,001,575	GAF0.3	1,225,823	SBF12	1,469,650
ACF64.3	345,993	GAF0.4	356,486	SBF13	3,013,842
ACF8.1	347,190	GAF2.2	203,705	SBF14	2,565,488
ACF8.3	784,551	GAF2.3	1,752,462	SBF3	131,2011
ALF1	3,897,391	GAF2.4	3,090,870	SBF5	213,2877
ALF2	10,356,692	GAF32.1	1,621,796	SBF6	2,266,450
ALF3	8,331,833	GAF32.2	663,560	SBF7	2,998,274
ALF4	7,186,532	PGF0.1	1,024,485	SBF8	3,570,973
ALF5	3,569,055	PGF0.3	2,434,688	SBF9	3,271,778
ALF6	4,639,229	PGF0.4	2,992,776	SJF0.2	2,980,129
ARF16.2	4,568,280	PGF32.2	5,260,332	SJF0.3	3,376,567
ARF16.3	2,759,362	PGF32.3	1,576,822	SJF0.4	3,355,621
ARF4.1	3,587,733	PGF32.4	4,626,778	SJF32.1	2,352,544
ARF4.2	3,350,379	PGF8.1	3,528,917	SJF32.2	5,179,336
ARF4.3	2,631,063	PGF8.2	1,591,923	SJF32.3	3,454,503
ARF4.4	3,772,072	PGF8.3	2,533,452	SJF8.1	5,932,428
ARF64.2	1,810,775	PGF8.4	1,626,483	SJF8.2	2,866,329
ARF64.3	4,292,607	PIF1	1,155,773	SJF8.3	5,930,375
DF16.1	4,979,697	PIF2	3,067,520	SJF8.4	1,909,197
DF16.2	3,177,452	RBF1	1,713,998	STF10	1,190,707
DF16.3	3,784,539	RBF2	516,749	STF2	1,376,267
DF4.1	4,166,576	RBF3	977,974	STF3	990,306
DF4.3	5,241,570	RBF4	693,838	STF4	536,949
DF64.1	4,475,209	RBF5	5,191,876	STF8	439,814
				STF9	281,040

Table 4S. Haplotypes' distribution across the sampled locations. Abbreviation as in Table 1.

	SBF	ALF	ACF	GAF	SJF	PGF	DF	ARF	STF	RBF	PIF	NE	SE	S
Hap_1	0	0	1	0	0	0	0	0	0	0	0	0	1	0
Hap_2	0	0	1	0	0	0	0	0	0	0	0	0	1	0
Hap_3	1	0	0	0	1	1	0	0	0	0	0	1	2	0
Hap_4	2	2	0	2	0	0	0	3	2	1	1	4	5	4
Hap_5	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Hap_6	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Hap_7	0	0	0	0	2	1	4	2	1	3	0	0	9	4
Hap_8	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Hap_9	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Hap_10	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Hap_11	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Hap_12	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Hap_13	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Hap_14	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Hap_15	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Hap_16	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Hap_17	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Hap_18	0	0	0	0	0	1	0	0	0	0	0	0	1	0
Hap_19	0	0	0	1	0	0	0	0	0	0	0	0	1	0
Hap_20	0	0	0	0	0	1	0	0	0	0	0	0	1	0
Hap_21	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Hap_22	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Hap_23	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Hap_24	1	0	0	0	0	0	0	0	0	0	0	1	0	0
Hap_25	1	0	0	0	0	0	0	0	0	0	0	1	0	0
Hap_26	1	0	0	0	0	0	0	0	0	0	0	1	0	0
Hap_27	1	0	0	0	0	0	0	0	0	0	0	1	0	0
Hap_28	0	0	0	0	1	0	0	0	0	0	0	0	1	0
Hap_29	0	0	0	0	1	0	0	0	0	0	0	0	1	0
Hap_30	0	0	0	0	1	0	0	0	0	0	0	0	1	0
Hap_31	0	0	0	0	1	0	0	0	0	0	0	0	1	0
Hap_32	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Hap_33	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Hap_34	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Hap_35	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Unique (%)	71,4	60	100	33,3	71,4	100	55,6	50	71,4	40	100	75	50	42,9

Unique (%) - percentage of unique haplotypes.

Table 5S. Pairwise F_{ST} among sampled locations along the Brazilian coast. The index values were calculated for (A-B) mtDNA and (C-D) SNPs. The comparisons were performed according to localities and regions. The F_{ST} indexes are in the lower matrix, and the p-values are in the upper matrix. Significant indexes are in bold. Abbreviations as in Table 1.

A)

	SBF	ALF	ACF	GAF	SJF	PGF	DF	ARF	STF	RBF	PIF
SBF		0.951	0.051	0.971	0.361	0.465	0.105	0.655	0.752	0.451	0.835
ALF	-0.072		0.051	1.000	0.040	0.172	0.030	0.423	0.172	0.523	0.563
ACF	0.295	0.484		0.105	0.245	0.526	0.033	0.015	0.055	0.050	0.343
GAF	-0.133	-0.121	0.628		0.133	0.199	0.185	0.535	0.437	0.304	0.403
SJF	0.025	**0.1650***	0.084	0.118		0.924	0.085	0.024	0.330	0.101	0.384
PGF	-0.009	0.119	-0.028	0.100	-0.099		0.036	0.066	0.197	0.192	0.519
DF	0.072	**0.1873***	**0.5180***	0.154	0.128	**0.1914***		0.209	0.777	0.769	0.940
ARF	-0.028	0.004	**0.4805***	-0.047	**0.1485***	0.149	0.029		0.718	0.817	0.843
STF	-0.033	0.068	0.435	0.033	0.040	0.078	-0.055	-0.035		0.837	0.917
RBF	-0.001	0.076	0.501	0.093	0.147	0.135	-0.045	-0.062	-0.059		0.708
PIF	-0.162	-0.012	0.500	0.089	-0.011	0.014	-0.212	-0.186	-0.238	-0.255	

B)

	NE	SE	S
NE		0.141	0.070
SE	0.029		0.274
S	0.059	0.009	

C)

	SBF	ALF	ACF	GAF	SJF	PGF	DF	ARF	STF	RBF	PIF
SBF		0.505	0.212	0.139	0.811	0.033	0.068	0.301	0.170	0.019	0.272
ALF	0.006		0.966	0.352	0.388	0.446	0.226	0.348	0.018	0.072	0.606
ACF	0.008	-0.007		0.277	0.094	0.362	0.420	0.085	0.001	0.008	0.207
GAF	0.010	0.011	0.009		0.635	0.021	0.272	0.371	0.047	0.043	0.528
SJF	0.001	0.009	0.011	0.003		0.160	0.205	0.559	0.325	0.138	0.522
PGF	**0.0123***	0.008	0.007	**0.0165***	0.010		0.049	0.003	0.001	0.046	0.166
DF	0.010	0.011	0.005	0.009	0.008	**0.0119***		0.287	0.176	0.225	0.693
ARF	0.009	0.013	0.015	0.011	0.006	**0.0221****	0.009		0.246	0.202	0.742
STF	0.013	**0.0319***	**0.0332****	**0.0222***	0.012	**0.0324****	0.015	0.018		0.031	0.647
RBF	**0.0201***	0.022	**0.0217****	**0.0214***	0.014	**0.0175***	0.012	0.017	**0.0311***		0.739
PIF	0.023	0.014	0.021	0.015	0.015	0.032	0.004	0.013	0.016	0.018	

D)

	NE	SE	S
NE		0.291	0.007
SE	0.003		0.001
S	**0.0127****	**0.0071****	

** Statistically significant values ($p < 0.05$)*
*** Statistically significant values ($p < 0.1$)*

Table 6S. Pairwise F_{ST} calculated for individuals within transects. The F_{ST} indexes are in the lower matrix and the p-values are in the upper matrix. There are not significant p-values ($p < 0.5$). The number after the locality code (detailed in Table 1) represents the distance from the first sampled point.

Anchieta - ES

	ACF4	ACF64	ACF8
ACF4		0.500	0.379
ACF64	0.021		0.803
ACF8	0.025	0.007	

Praia de Gamboa - ES

	GAF0	GAF2	GAF32
GAF0		0.597	0.402
GAF2	0.033		0.798
GAF32	0.045	0.029	

Praia Dura - SP

	DF16	DF4	DF64
DF16		0.699	0.319
DF4	0.026		0.471
DF64	0.018	0.027	

Araçá - SP

	ARF16	ARF4	ARF64
ARF16		0.132	0.332
ARF4	0.041		0.066
ARF64	0.036	0.048	

Barra de São João - RJ

	SJF0	SJF32	SJF8
SJF0		0.894	0.573
SJF32	0.005		0.629
SJF8	0.015	0.019	

Praia da Gorda - RJ

	PGF0	PG32	PGF8
PGF0		0.999	0.453
PGF32	0.018		0.598
PGF8	0.020	0.015	

Table 7S. Blast results of sequences containing SNPs identified on BayeScan, DAPC and LFMM. The blast was performed with the *Littoraria flava* transcriptome previously annotated.

SNPs	Locus	Size	Origin	Sequence Description
SNP_45	239	59	c	INO80 complex subunit C
SNP_80	347	80	b	RNA-binding protein Raly
SNP_107	654	90	b	TBC1 domain family member 2B
SNP_148	767	87	c, e, f	Inositol polyphosphate multikinase
SNP_231	1127	87	f	DNA helicase MCM9
SNP_318	1644	74	c	Zinc finger protein basonuclin-2
SNP_374	1963	75	e, f	Quinone oxidoreductase
SNP_417	2316	74	d	RNA-directed DNA polymerase from mobile element jockey
SNP_444	2404	92	c	Polyadenylate-binding protein-interacting protein 1
SNP_486	2617	*62*	a, c	Selenoprotein Pa
SNP_584	2857	91	f	Syndetin
SNP_615, SNP_609	3026	48	c, f	Slit homolog 2 protein
SNP_700, SNP_801	3781	56	b, c, e	Endo-1,4-beta-xylanase 2
SNP_805	6221	89	b	Transcription factor Adf-1
SNP_819	6647	66	b	Alpha-amylase
SNP_840, SNP_828, SNP_836	7063	97	a, b, f	Caprin-2
SNP_845	7394	82	b	Nuclear factor 1 X-type
SNP_995	10201	89	b	Vitamin D3 receptor
SNP_1049	10543	65	a, b	Protein NipSnap homolog 3A N6-acetyllysine N6-acetyllysine
SNP_1062, SNP_1065	10967	83	a	Versican core protein
SNP_1074	11603	66	b	Leukotriene A-4 hydrolase
SNP_1102	11716	94	b	DDB1- and CUL4-associated factor 5
SNP_1168	12846	77	a	Kinetochore-associated protein 1
SNP_1180	12902	59	a, b, c, f	Rho GTPase-activating protein 20
SNP_1246	13791	73	f	Neutrophil collagenase
SNP_1305	14216	86	c	Cyclin-I
SNP_1340	14333	86	c	Myoneurin
SNP_1542	14852	87	e	Transcription initiation factor TFIID subunit 4
SNP_1595	15100	89	b, f	protein phosphatase 2 regulatory subunit B"gamma (ppp2r3c), mRNA
SNP_1798	15617	62	b	1-deoxyxylulose-5-phosphate synthase YajO

SNP_2025	16277	86	f	Vacuolar protein sorting-associated protein 13D
SNP_2127, SNP_2120	16658	86	c, f	Oxalate:formate antiporter
SNP_2453, SNP_2459	17493	89	b, c, e, f	BAC clone RP11-556I14 from 4, complete sequence
SNP_2904	18317	63	f	Tripartite motif-containing protein 2
SNP_3054	18475	86	f	Perlucin
SNP_3192	18697	85	a, c, f	Glycoprotein 3-alpha-L-fucosyltransferase A
SNP_3289	18851	*60*	a	Inner centromere protein
SNP_3343	18980	86	c	Leucine-rich repeat and death domain-containing protein 1
SNP_3403	19125	89	b	Transcription initiation factor TFIID subunit 3
SNP_3412	19173	87	f	Probable helicase with zinc finger domain
SNP_3433	19244	87	f	Zinc finger protein 518B
SNP_3534	19523	86	c	PAX3- and PAX7-binding protein 1
SNP_3929	20739	86	f	Vacuolar protein sorting-associated protein 13B
SNP_3933	20751	86	c	Dynein heavy chain 5, axonemal
SNP_4006, SNP_4013	21008	74	c, f	Tripartite motif-containing protein 3
SNP_4101, SNP_4108, SNP_4103	21634	73	a, c, e	Tudor domain-containing protein 5
SNP_4223	22496	52	e	Protein unc-80 homolog
SNP_4432	25051	82	b	Synergin gamma
SNP_4560	27564	82	b	SET and MYND domain-containing protein 4
SNP_4594	28146	109	b	Neprilysin-1
SNP_4604, SNP_4602, SNP_4606	28511	48	c, d, f	Protein bicaudal C homolog 1-B
SNP_4616	28683	71	a	Ficolin-2
SNP_5010	30172	87	f	ATP-binding cassette sub-family A member 2
SNP_5227	30746	*89*	a	Probable 3',5'-cyclic phosphodiesterase pde-5
SNP_5676	31943	86	d	Innexin unc-9
SNP_5886	32705	86	f	PPM-type phosphatase domain-containing protein
SNP_6016	33027	89	b	Fasciclin-2
SNP_6233	33700	89	b, f	Tereporin-Ca1
SNP_6263	33781	87	c	Estrogen-related receptor gamma
SNP_6336	33926	86	e	Neuropeptide FF receptor 1
SNP_6369	33984	89	b	Galaxin
SNP_6455	34240	*89*	a	Twitchin
SNP_6488	34305	87	f	Rho GTPase-activating protein 26
SNP_6646	34588	89	b	Probable RNA-directed DNA polymerase from transposon BS
SNP_6841	35012	86	b, c, e, f	Fibrinogen C domain-containing protein 1-A

SNP_6923	35300	89	b	Protein mono-ADP-ribosyltransferase PARP10
SNP_7018	35528	86	f	Sarcolemmal membrane-associated protein
SNP_7027	35530	86	c	Toll-like receptor 2 type-2
SNP_7183	36447	*88*	a	Cleavage and polyadenylation specificity factor subunit 1
SNP_7295, SNP_7293	36605	86	a, f	Collagen alpha-1(XII) chain
SNP_7398	36865	86	d	Triple functional domain protein
SNP_7460, SNP_7451	37040	86	d, e	Unconventional myosin-VI
SNP_7475	37229	48	e, f	Complement C1q-like protein 2
SNP_7630	38457	87	f	SUMO-specific isopeptidase USPL1
SNP_7846	39710	83	f	Beta-1,4-galactosyltransferase 7
SNP_8004	41207	82	b	Vitamin D3 receptor B
SNP_8030	41333	89	b	Cadherin-87A
SNP_8060	41474	89	b	genome assembly, chromosome: 9
SNP_8061	41498	89	b	Tumor susceptibility gene 101 protein
SNP_8072	41538	89	b	Cullin-5
SNP_8081	41659	89	b	Carbohydrate deacetylase
SNP_8098	41851	89	b	Structural maintenance of chromosomes protein 4
SNP_8100	41882	73	b	Girdin
SNP_8110	41980	89	b	Rho-related BTB domain-containing protein 1
SNP_8122	42017	57	b	Divergent protein kinase domain 1A
SNP_8159	42228	89	b	Myeloid differentiation primary response protein MyD88
SNP_8175	42256	89	b	E3 ubiquitin-protein ligase TRIM56
SNP_8207	42836	89	b	Ataxin-7-like protein 1
SNP_8213	42879	88	b	Androglobin
SNP_8227	42920	89	b	Cyclin-dependent kinase 5 activator 1
SNP_8233	42987	57	b	Thyroxine 5-deiodinase
SNP_8239	43007	89	b	Ubiquitin carboxyl-terminal hydrolase 8
SNP_8265	43295	103	b	Glutamate receptor 4
SNP_8278	43372	89	b	Fibropellin-1
SNP_8303	43416	89	b	Leucine-rich repeat and death domain-containing protein 1
SNP_8305	43430	89	b	Ataxin-7-like protein 1
SNP_8334	43571	88	b	Autotransporter adhesin BpaC
SNP_8397	44253	89	b	Mucin-like protein
SNP_8457	44921	88	b, f	Vacuolar protein sorting-associated protein 13D
SNP_8473, SNP_8462	44975	113	a, b, c, f	Tripartite motif-containing protein 3
SNP_8507	45164	92	b	RAC serine/threonine-protein kinase
SNP_8550	45399	89	b	Spectrin beta chain

SNP_8553	45463	89	b	Trafficking kinesin-binding protein 1
SNP_8567	45574	89	b	Trypsin-2
SNP_8620	45947	89	b	Poly(U)-binding-splicing factor half pint
SNP_8636, SNP_8639	46059	89	b, c, d, e, f	Ribonuclease 3
SNP_8640	46068	88	b	Cytospin-A
SNP_8655	46124	89	b	Myeloid differentiation primary response protein MyD88
SNP_8688	46181	89	b	Neuronal acetylcholine receptor subunit alpha-7
SNP_8694	46254	88	b	Mitogen-activated protein kinase 15
SNP_8711	46508	55	b	Centrosomal protein of 135 kDa
SNP_8757	46814	89	b	Nuclear factor interleukin-3-regulated protein
SNP_8767	47185	88	b	Dopamine receptor 1
SNP_8796	47525	90	b	Sulfotransferase family cytosolic 1B member 1
SNP_8813	47674	89	b	Zinc finger protein 778
SNP_8825, SNP_8826	47773	88	b, c	BTB/POZ domain-containing protein KCTD3
SNP_8890	48102	89	b	Host cell factor 1
SNP_8912	48174	88	b	Wings apart-like protein homolog
SNP_8932	48423	88	b	Sentrin-specific protease 7
SNP_8966	48521	89	b	Amine sulfotransferase
SNP_9112	49630	109	b	Probable E3 ubiquitin-protein ligase HECTD4
SNP_9344	50838	115	b	Tereporin-Ca1
SNP_9362	51118	*90*	a, b	Eukaryotic initiation factor 4A-III
SNP_9512	52337	70	b	Transmembrane protein 180
SNP_9872	56193	53	d	Yolk ferritin
SNP_9950	57030	98	b	Transcriptional repressor CTCFL
SNP_10079	58011	88	a	Dynein heavy chain 7
SNP_10123, SNP_10125	58991	69	c, e, f	Tripartite motif-containing protein 2
SNP_10171	59194	80	b	Cilia- and flagella-associated protein 299
SNP_10249	59667	87	c	General transcription and DNA repair factor IIH helicase subunit XPB
SNP_10308	59966	55	c	Transmembrane protein 183
SNP_10362, SNP_10365	60146	92	a, c, f	SAGA-associated factor 29
SNP_10536	60886	89	b	NAD-dependent protein deacetylase sirtuin-2
SNP_10554	61039	67	b	Glutamate-rich protein 6
SNP_10605, SNP_10603	61414	64	b, d, f	Versican core protein
SNP_10723	62093	88	b	Ankyrin repeat domain-containing protein 24
SNP_10725	62132	89	b	Glycine receptor subunit alpha-2

SNP_10749	62482	82	b	Ankyrin-3
SNP_10757	62498	72	b	Protein jagged-1b
SNP_10764	62567	88	b	Cytochrome P450 2U1
SNP_10797	62752	88	b	Ubiquitin carboxyl-terminal hydrolase 38
SNP_10873, SNP_10874, SNP_10868	62977	88	b, d, f	Zinc finger protein 79
SNP_10934	63211	89	b	Zinc finger protein 79
SNP_10953	63323	*88*	a	Nuclear receptor subfamily 2 group F member 6
SNP_10986	63409	56	a, e, f	Dynein heavy chain 12, axonemal
SNP_11000	63410	88	b	Zinc finger protein 541
SNP_11001	63592	91	b	Putative uncharacterized protein FLJ45035
SNP_11012, SNP_11019	63651	89	b, f	Retinoic acid-induced protein 1
SNP_11028, SNP_11023	63680	86	a, b, d, e	NLR family CARD domain-containing protein 3
SNP_11032	63686	88	b	Xylosyltransferase oxt
SNP_11060	63723	88	b	Dicer-like protein 1
SNP_11082	63828	89	b	NFX1-type zinc finger-containing protein 1
SNP_11262	64791	87	f	WD repeat and SOCS box-containing protein 1
SNP_11470	66429	53	c, f	Ataxin-3
SNP_11530	66718	88	a	Unhealthy ribosome biogenesis protein 2 homolog
SNP_11685	67971	87	c, e, f	Protein NEDD1
SNP_11734	68365	87	f	Zinc finger protein 90
SNP_11739	68376	87	d	Craniofacial development protein 2
SNP_11807, SNP_11810	68855	89	b, c	1-phosphatidylinositol 3-phosphate 5-kinase
SNP_12055	70162	87	d	snRNA-activating protein complex subunit 1
SNP_12072	70297	87	f	TBC1 domain family member 1
SNP_12205, SNP_12213	70907	68	c, f	Ribonuclease P protein subunit p29
SNP_12571	73680	89	b	Leucine zipper putative tumor suppressor 2
SNP_12797	75255	*78*	a, b	Aryl hydrocarbon receptor nuclear translocator homolog
SNP_12866	75551	83	e	Cullin-5
SNP_13171	77558	*90*	a	Cyclin-C
SNP_13218	77809	92	f	Oxysterol-binding protein-related protein 8
SNP_13255	78005	86	b	Histone-lysine N-methyltransferase ASH1L
SNP_13285	78111	53	c, f	Nucleoprotein TPR
SNP_13802	82767	68	c, f	Zinc finger protein 16
SNP_13890	83797	57	b	Thioredoxin domain-containing protein 2

SNP_13937, SNP_13942	84394	59	b, f	G patch domain and ankyrin repeat-containing protein 1 homolog
SNP_14074	87830	117	e, f	Fibropellin-1
SNP_14288	90831	78	c	Protein phosphatase 1L
SNP_14303	90958	69	d	BMP-binding endothelial regulator protein
SNP_14315, SNP_14316	91098	88	b, c	Pericentriolar material 1 protein
SNP_14334	91125	120	c, e	Retinal Mueller cells isomerohydrolase
SNP_14346	91191	68	a	Von Willebrand factor type A, EGF and pentraxin domain
SNP_14431	92017	74	c	Cilia- and flagella-associated protein 58
SNP_14942, SNP_14943	100495	53	c, d, f	Nucleoprotein TPR
SNP_15021	101164	71	f	Cyclic nucleotide-binding domain-containing protein 2
SNP_15122	104234	82	a	Organic cation transporter protein
SNP_15131	104259	94	f	Protein phosphatase 1 regulatory subunit 21
SNP_15148	104388	77	b	E3 ubiquitin-protein ligase NRDP1
SNP_15204	105794	79	f	Kelch-like protein 5
SNP_15544	111055	*89*	a, b	Cleavage and polyadenylation specificity factor subunit 1
SNP_15649	111507	87	f	Unconventional myosin-Va
SNP_15695	111996	86	c	Differentially expressed in FDCP 8 homolog
SNP_15727	112120	91	c, e, f	Src substrate cortactin
SNP_15831	112699	104	f	Sodium channel and clathrin linker 1
SNP_15928	114342	87	d	Kinesin-like protein KIN-14I
SNP_16036, SNP_16039	115081	88	b, f	Proteasome activator complex subunit 4
SNP_16210	116021	106	b	Glycoprotein 3-alpha-L-fucosyltransferase A
SNP_16241, SNP_16242	116493	53	d, f	NLR family CARD domain-containing protein 3
SNP_16281	116627	65	a, c	Calcineurin-binding protein cabin-1
SNP_16303	116894	84	b, c	Neuronal PAS domain-containing protein 4
SNP_16451	117685	84	f	Rho guanine nucleotide exchange factor 33
SNP_16458	117734	86	c	Malignant fibrous histiocytoma-amplified sequence 1 homolog
SNP_16574	118564	86	c	Protocadherin-12
SNP_16644	118958	86	a, c, d, f	Pyridoxal-dependent decarboxylase domain-containing protein 1
SNP_16694	119750	82	b	60 kDa SS-A/Ro ribonucleoprotein
SNP_16729	120214	94	b	LIM homeobox transcription factor 1-beta

SNP_16916, SNP_16923	121390	100	c, d	Cilia- and flagella-associated protein 20
SNP_17019	122054	86	f	Baculoviral IAP repeat-containing protein 3
SNP_17225, SNP_17227	123803	86	c, e	E3 ubiquitin-protein ligase rnf213-alpha
SNP_17281	124332	86	c	Multidrug resistance-associated protein 7
SNP_17312	125105	66	c	Ubiquitin-protein ligase E3B
SNP_17408, SNP_17406	125731	54	d, f	Cysteine protease ATG4D
SNP_17560, SNP_17562, SNP_17565	126785	89	a, b, c, f	Zinc finger protein 345
SNP_17619	127207	87	c	Tyrosine-protein kinase Src42A
SNP_17711	128318	86	c	Protein phosphatase 1 regulatory subunit 7
SNP_17737	128337	86	c, f	Protein NO VEIN
SNP_17745, SNP_17748, SNP_17742	128341	115	a, c, d	Dynein beta chain, ciliary
SNP_17853	128906	89	b	Synaptotagmin-7
SNP_17964	129881	*64*	a	Glutamate receptor 1
SNP_18051	130561	74	f	Gamma-1-syntrophin
SNP_18134	131189	86	f	Sacsin
SNP_18245	131886	*66*	a	Brefeldin A-inhibited guanine nucleotide-exchange protein 1
SNP_18259	131977	75	a	Protein transport protein Sec31A
SNP_18281	132347	75	b	Glutathione S-transferase P 1
SNP_18314	132835	55	c	Serine/threonine-protein phosphatase PP1
SNP_18543	134652	62	f	Dynein heavy chain 2, axonemal
SNP_18572	134662	111	e	Delta-like protein 1
SNP_18656	135006	84	a, d, f	Ral GTPase-activating protein subunit beta
SNP_18951	135772	64	c	Sterile alpha motif domain-containing protein 9

Size - locus size; Origin – Program or predictor associated with the SNP identification.

a – BayeScan;

b – DAPC;

c – LFMM (Temperature Driest Quarter);

d – LFMM (Diurnal Range Temperature);

e – LFMM (Precipitation Wettest Month);

f – LFMM (Precipitation Coldest Quarter).

Fig. 1S. Matrix condenser of 2,249 loci (columns) and 85 samples of *Littoraria flava* (rows). The locus presence (black lines) or absence (white lines) are represented for each individual. Abbreviations as in Table 1.

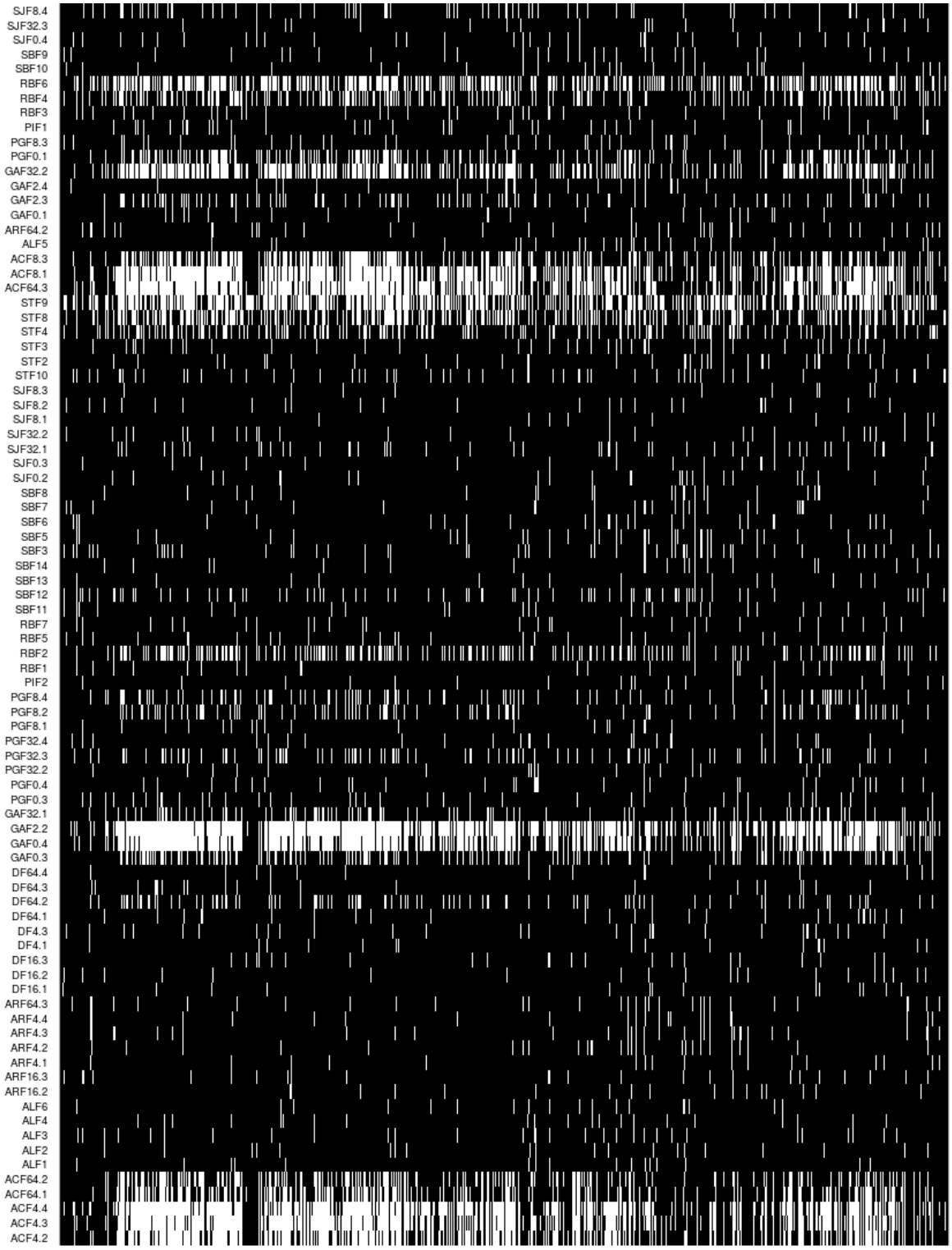


Fig. 2S. Mantel test performed with mtDNA and SNPs markers for macro geographic scale. Each graph contains the geographic distance in the x-axis and the genetic distance (F_{ST}) in the y-axis. The regression line is in blue, and the values on the side of each graph indicates the observed correlation and its p-value. The titles specifies the genetic marker and the hierarchical level used in the analysis.

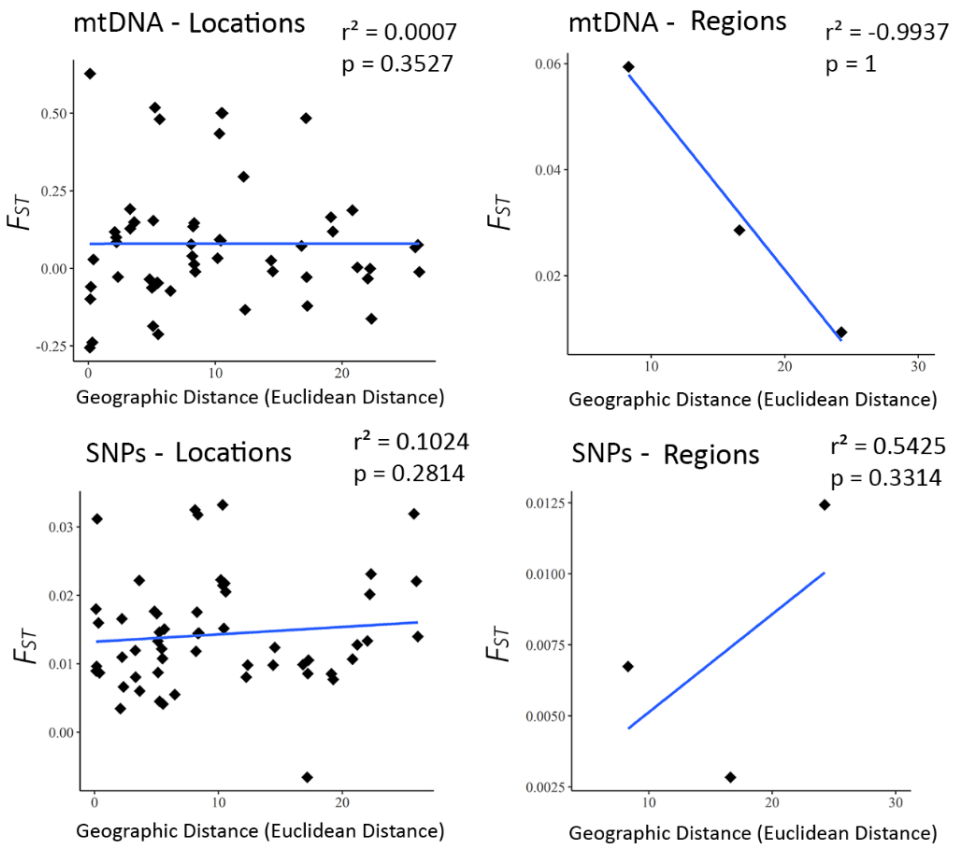


Fig. 3S. Representation of SNPs contribution to the genetic structure found among the sampled populations of *Littoraria flava* (K = 3). The SNPs with weight higher than 0.1% are above the gray line.

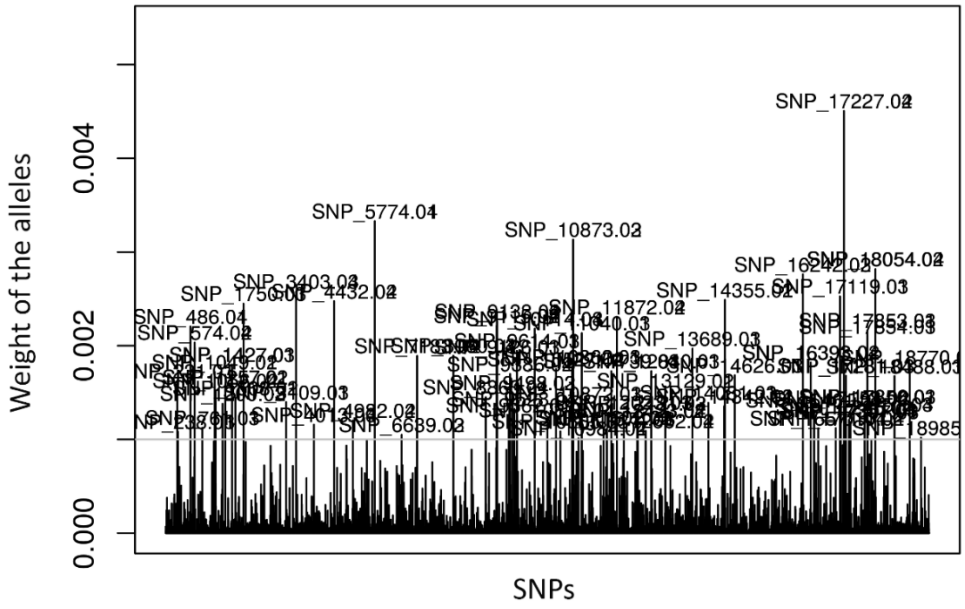


Fig. 4S. Mantel test performed with SNPs markers for micro geographic scale (within transects). Each graph contains the geographic distance in the x-axis and the genetic distance (F_{ST}) in the y-axis. The regression line is in blue, and the values on the side of each graph indicates the observed correlation and its p-value. The titles specifies the locality. Abbreviations as in Table 1.

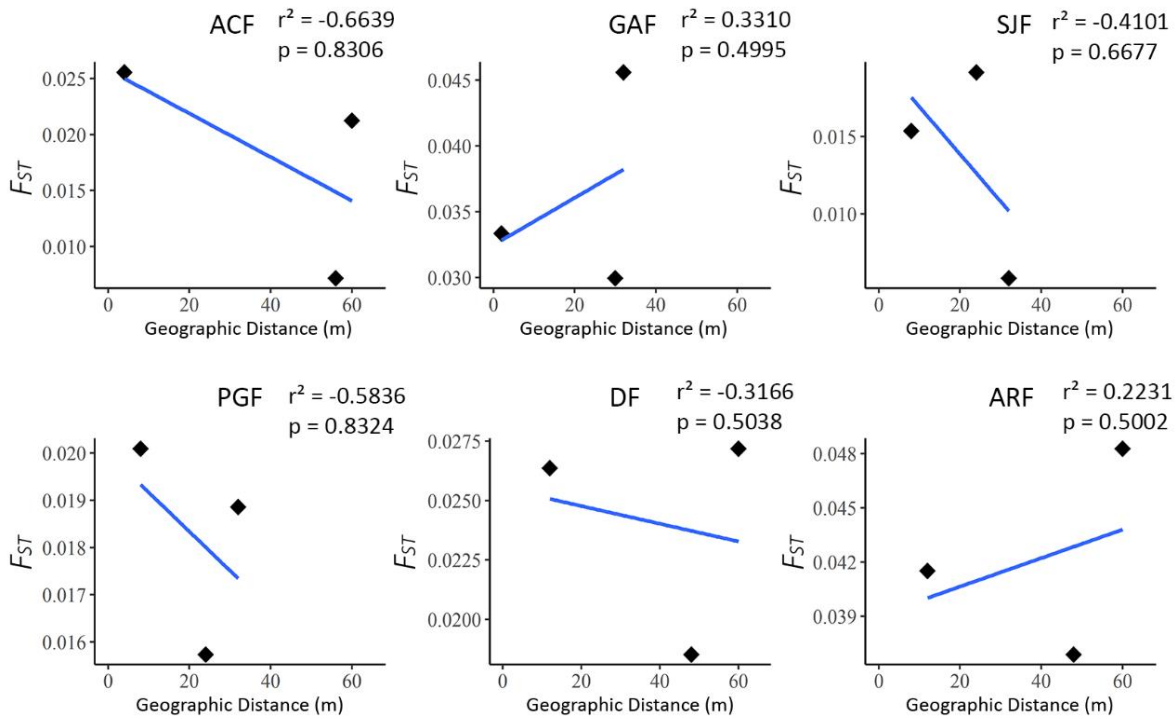


Fig. 5S. PCA statistics over the climatic variables. (A) The scree plot of eigenvalues showing the variation within each principal component. (B) The cumulative explained variance plot indicating the four first components cumulatively accounting ~92% of the total variance.

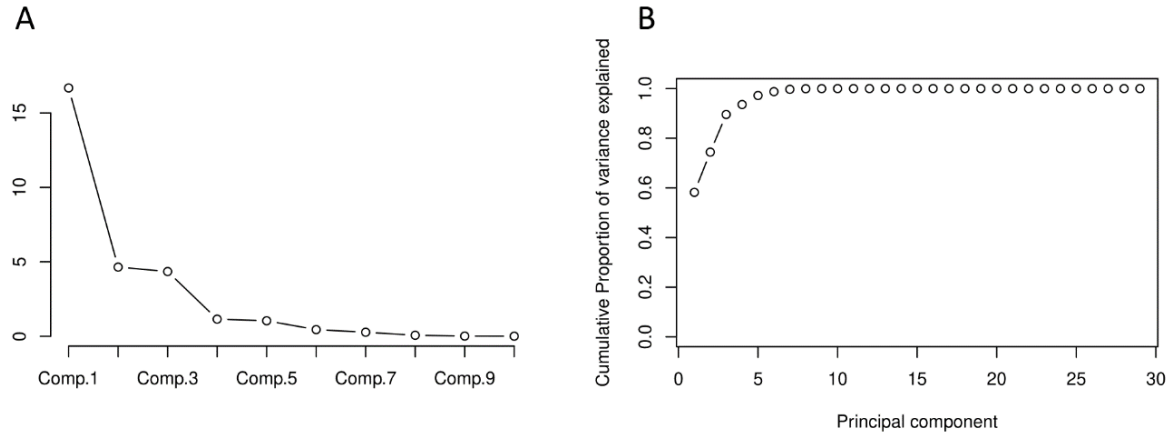


Fig. 6S. Bayesian test for selection on individual SNPs in BayeScan. Each dot represents a single SNP with their respective q-value on the x-axis and F_{ST} on the y-axis. The gray vertical line indicates the significance level (0.05). SNPs to the left of the vertical line represent outliers (q < 0.05).