Parallel use of a shared genomic island of speciation in clinal and mosaic hybrid zones between cryptic seahorse lineages

Florentine Riquet[1, 2], Cathy Liautard-Haag[1, 2], Lucy Woodall[3, 4], Carmen Bouza[5], Patrick Louisy[6, 7], Bojan Hamer[8], Francisco Otero-Ferrer[9], Philippe Aublanc[10], Vickie Béduneau[11], Olivier Briard[12], Tahani El Ayari[1, 2], Sandra Hochscheid[13], Khalid Belkhir[1, 2], Sophie Arnaud-Haond[1, 14], Pierre-Alexandre Gagnaire[1, 2], Nicolas Bierne[1, 2]

[1] Institut des Sciences de l'Evolution de Montpellier, Université Montpellier, Montpellier, France
[2] CNRS Institut des Sciences de l'Evolution, UMR5554 UM-CNRS-IRD-EPHE, Station Marine OREME, Sète, France
[3] Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK
[4] Natural History Museum, Cromwell Road, London SW7 5BD, UK
[5] Department of Genetics, Faculty of Veterinary Science, Universidade de Santiago de Compostela, Campus de Lugo, Lugo, Spain
[6] University of Nice Sophia Antipolis, ECOMERS Laboratory, Faculty of Sciences, Parc Valrose, Nice, France
[7] Association Peau-Bleue, 46 rue des Escais, Agde, France
[8] Center for Marine Research, Ruder Boskovic Institute, Giordano Paliaga 5, 52210 Rovinj, Croatia
[9] Grupo en Biodiversidad y Conservación, IU-ECOAQUA, Universidad de Las Palmas de Gran Canaria, Crta. Taliarte s/n, 35214 Telde, Spain
[10] Institut océanographique Paul Ricard, Ile des Embiez, Six-Fours-les-Plages, France
[11] Océarium du Croisic, Avenue de Saint Goustan, Le Croisic, France
[12] Aquarium de Biarritz, Biarritz Océan, Plateau de l'Atalaye, Biarritz, France
[13] Stazione Zoologica Anton Dohrn, Department Research Infrastructures for Marine Biological Resources, Aquarium Unit, Napoli, Italy
[14] Ifremer - MARine Biodiversity, Exploitation and Conservation, UMR 9190 IRD-IFREMER-UM-CNRS, Sète, France

Abstract

Diverging semi-isolated lineages either meet in narrow clinal hybrid zones, or have a mosaic distribution associated with environmental variation. Intrinsic reproductive isolation is often emphasized in the former and local adaptation in the latter, although both can contribute to isolation. Rarely these two patterns of spatial distribution are reported in the same study system, while this could provide fundamental information on the endless debate about the relative contribution of intrinsic reproductive isolation and local adaptation on the speciation process. Here we report that the low diversity long-snouted seahorse *Hippocampus guttulatus* is genetically subdivided into five cryptic semi-isolated lineages. Along the Atlantic coasts, northern and southern lineages meet and coexist with little hybridization in the southwest of France, forming a clinal hybrid zone. In the Mediterranean Sea, two lineages have a mosaic distribution associated with lagoon-like and marine habitats. A fifth lineage was identified in the Black Sea. Genetic homogeneity over large spatial scales within each lineage, together with among-locus

variance in differentiation levels between lineages and spatial patterns of introgression provide evidence that partial reproductive isolation is maintaining the divergence. Surprisingly, the North Atlantic and the Mediterranean lagoon lineages are genetically similar for a single chromosome-wide island showing parallel genetic differentiation among regions. Since Atlantic lineages distribution lacks association with habitat variation, genetic parallelism suggests that a shared genomic barrier contributes to reproductive isolation in contrasted contexts -*i.e.* spatial *vs*. ecological. We discuss how a genomic hotspot of parallel differentiation could have evolved and become associated either with space or with a patchy environment in a single study system.

Keywords: clinal hybrid zone, mosaic hybrid zone, reproductive isolation, local adaptation, ecological speciation, parallel evolution

Introduction

The spatial context of contact zones between partially isolated taxa and its relationship with environmental variation was long thought to offer great promises to unravel the nature and origin of species. Though each taxa may be genetically homogeneous over large distances, they meet in abrupt genetic discontinuities, called hybrid zones, in which partial reproductive isolation limits gene exchange (Barton and Hewitt 1985, Hewitt 1988). Hybrid zones are extremely informative for exploring the genetic basis of reproductive isolation (e.g. Teeter et al. 2008, Christe et al. 2016) and local adaptation (e.g. Jones et al. 2012, Larson et al. 2013, Soria-Carrasco et al. 2014), as well as identifying genomic regions involved either in increased genomic differentiation (reviewed in Ravinet et al. 2017) or adaptive introgression (Hedrick 2013). The hybrid zone literature usually contrasts two spatial patterns (Harrison 1993): (i) clinal hybrid zones, with parapatrically distributed parental forms on both sides of a cline-shaped genetic divide, and (ii) mosaic hybrid zones, when the environment consists of a mosaic of habitat patches to which taxa (ecotypes, host races, hybridizing species) are somehow specialized. Hybrid zones of both types have been recognized to be multifactorial and maintained both by exogenous and endogenous diverging mechanisms (*i.e.* local adaptation and intrinsic reproductive isolation, respectively; Barton and Hewitt 1985, Bierne et al. 2011). Nonetheless, clinal hybrid zones tend to be interpreted as being mainly maintained by intrinsic reproductive isolation evolved in allopatry before contact (the tension zone model, Barton and Hewitt 1985). Conversely, local adaptation is often emphasized to drive mosaic distributions which are interpreted as evidence that selection at local adaptation genes is acting to increase differentiation between habitats in a parallel fashion (Nosil and Feder 2012). This sketchy dichotomy is anchored by emblematic study systems for which decades of research allowed to support such interpretations. For instance, the hybrid zone between the mice *Mus musculus musculus* and *M. m. domesticus* in central Europe (Boursot et al. 1993) has been well demonstrated to be maintained by selection against hybrid genotypes (Britton-Davidian et al. 2005, Good et al. 2008) after secondary contact (Duvaux et al. 2011). Although the position of the zone was initially found associated with rainfall (Hunt and Selander 1973), local adaptation is not considered as contributing a lot to the isolation. At the other extreme, local adaptation at the same genetic variants has repeatedly allowed marine three-spined sticklebacks to evolve into a freshwater ecotype (Jones et al. 2012). Intrinsic selection is thought absent in the marine-freshwater sticklebacks system, although hybrids perform poorly in both environment and tend to reside in salinity ecotones (Vines et al. 2016). It would be misleading, however, to suggest the two alternative spatial contexts and relations to environmental variation may correspond to alternative routes to speciation (e.g. mutation-order *vs.* ecological speciation; Coyne and Orr 2004, Nosil and Feder 2012). The list is also long of hybrid zones maintained both by local adaptation and intrinsic reproductive isolation. *Bombina* toads (Szymura and Barton 1986), *Gryllus* crickets (Rand and Harrison 1989, Larson et al. 2014), or *Mytilus* mussels (Bierne et al. 2003), among others, are well-known examples of mosaic hybrid zones maintained by both exogenous and endogenous selection.

We here studied the population genetics of the long-snouted seahorse, *Hippocampus guttulatus,* across a large part of its distribution range. We developed an assay of 286 informative SNPs chosen among more than 2,500 SNPs from a population

transcriptomic study (Romiguier et al. 2014). *H. guttulatus* is thought to display poor dispersal abilities (e.g. site-fidelity, weak swimming performance, lack of dispersive stage) and inhabits fragmented habitats along its distribution range (from the English Channel through the Mediterranean and Black Seas; Lourie and Vincent 2004). In addition, small and patchy populations are often observed in this small coastal fish that is infrequently seen. Accordingly *H. guttulatus* has a very low genetic diversity (Romiguier et al. 2014). However, genetic differentiation proved to be very low over very large geographic distances (Woodall et al. 2015), or across the Cape Finisterre oceanographic barrier (López et al. 2015). Four well-differentiated genetic clusters, each distributed over extended regions, were delineated by abrupt genetic discontinuities corresponding to usual delimitations between vicariant marine lineages (Woodall et al. 2015) – between the Iberian Peninsula and the Bay of Biscay in the North Eastern Atlantic, between the Atlantic Ocean and the Mediterranean Sea, and between the Mediterranean and Black seas. Although the differentiation matches well with oceanographic barriers, this pattern is also in good accordance with the existence of reproductively isolated cryptic lineages, which boundaries were trapped by exogenous barriers (*i.e.* the coupling hypothesis; Bierne et al. 2011).

Our array of informative SNPs and a broader sampling in the Mediterranean Sea and contact zones, allowed us to describe five cryptic semi-isolated lineages: two paratrically distributed lineages in the Atlantic Ocean, two patchily distributed lineages associated with lagoon and marine environments in the Mediterranean Sea, and one in the Black Sea. Surprisingly, a shared genomic island of clustered outlier loci was involved both in the isolation between the two parapatric lineages in the Atlantic Ocean and between the marine and lagoon ecotypes in the Mediterranean Sea, with the North Atlantic lineage being related to the lagoon ecotype at this genomic island. However, the two Atlantic lineages inhabit both marine and lagoon habitats. The *H. guttulatus* complex becomes one of a few systems where a clinal and a mosaic hybrid zone are observed concomitantly, and a new counter-example that provides evidence of genetic parallelism in absence of ecological convergence.

Materials and Methods

Sampling and DNA extraction

H. guttulatus samples were collected alive from 25 sites (Fig. 1A) using a variety of methods (snorkeling, scuba diving, trawling nets, aquaria, donations). The dorsal fin of each individual was clipped using a non-lethal procedure (Woodall et al. 2012), before releasing back the individual into its natural habitat or aquaria, respectively. Each individualized sample was preserved in 96% ethanol for subsequent genetic analyses.

Whole genomic DNA was extracted following either Woodall et al. (2015), López et al. (2015) or using a standard CetylTrimethyl Ammonium Bromide (CTAB) Chloroform:Isoamyl alcohol (24:1) protocol (Doyle and Doyle 1987). For samples from Vigo (Spain), the whole genome was amplified using the GenomiPhi kit (GE HealthCare) according to the manufacturer's protocol. Quality and quantity of DNA extraction was checked on an agarose gel, and normalized to 35 $ng.\mu L^{-1}$ using Qubit Fluorometric Quantitation (Invitrogen). SNP-genotyping was performed (see below for SNP calling) using the GoldenGate Genotyping Assay with VeraCode technology at ADNid company (http://www.adnid.fr/index.html, Montpellier, France).

151
152 *Data mining for SNP markers*
153 A set of 12,613 contigs was examined to identify SNPs. This included one
154 mitochondrial contig (GenBank accession number: AF192664) and 12,612 contigs from
155 Romiguier et al. (2014). Romiguier et al. (2014) sequenced the transcriptomes of six *H.*
156 *guttulatus* from three locations (Le Croisic, Atlantic Ocean, France; Faro, Atlantic Ocean,
157 Portugal; Thau lagoon, Mediterranean Sea, France), and two *H. hippocampus* from two
158 locations (Sète, Mediterranean Sea, France; Bizerte, Mediterranean Sea, Tunisia) using a
159 combination of high-throughput sequencing methods. Based on these 12,613 contigs,
160 SNPs were identified using the same bioinformatic pipeline as described in
161 Bouchemousse et al. (2016). SNPs were called with Read2SNPs (Gayral et al. 2013) and
162 filtered out according to the following criteria to exclude: 1) SNPs showing more than
163 two alleles, 2) SNPs failing to be sequenced in at least one location, 3) SNPs present only
164 in one individual (*i.e.* singletons), 4) SNPs identified as paralogs using the paraclean
165 option of Read2SNPs, and 5) SNPs closer than 20 bp from a contig extremity or an exon
166 limit when blasted against the stickleback, cod and tilapia genomes. This resulted in
167 2,684 selected SNPs that were screened with the Illumina Assay Design Tool (ADT)
168 software to select 384 SNPs on the basis of their quality index (ADT score > 0.6).
169 To identify their chromosomal positions, the template sequences of the targeted
170 SNP were 1) blatted (Bhagwatt et al. 2012) against the genomes of the zebrafish (*Dano*
171 *rerio*, Howe et al. 2013), fugu (*Takifugu rubripes*, Kai et al. 2011), tetraodon (*Tetraodon*
172 *nigroviridis*, Jaillon et al. 2004), Nile tilapia (*Oreochromis niloticus*, Brawand et al.
173 2014), medaka (*Oryzias latipes*, Kasahara et al. 2007), stickleback (*Gasterosteus*
174 *aculeatus*, Jones et al. 2012), and European seabass (*Dicentrarchus labrax*, Tine et al.
175 2014), 2) blasted against the genome of the Gulf pipefish (*Syngnathus scovelli*, Small et
176 al. 2016) and 3) blasted against the scaffolds of the tiger tail seahorse (*Hippocampus*
177 *comes*, Lin et al. 2016), followed by blats and blasts of the scaffolds against the eight
178 well-assembled fish genomes.
179 All SNPs were oriented using *H. hippocampus* as an outgroup species to identify
180 the most parsimonious ancestral variant. The Joint Site-Frequency Spectrum obtained
181 from the original transcriptome-wide SNP dataset (Romiguier et al. 2014) was compared
182 to the JSFS obtained by the subset of 384 SNPs to investigate potential ascertainment
183 bias due to marker choice. The JSFS represents the bivariate distribution of allelic
184 frequencies in the two populations (Ewens 1972). It is a $(2n_1+1)$ x $(2n_2+1)$ dimension
185 matrix, with n_1 the number of individuals in population 1 and n_2 the number of
186 individuals in population 2, where each entry S(i,j) gives the number of SNPs for which
187 the derived allele was found i and j times in population 1 and 2, respectively. In order to
188 compare the JSFS obtained with the two datasets, the JSFS dimension was here projected
189 down to a 5x5 matrix, which was the dimension of the Rominguier et al. (2014) dataset.
190
191 *Analyses of Genetic Diversity and Genetic Structure*
192 The mean number of alleles per locus (N_{all}) and the allelic richness (A_r, *i.e.* the
193 expected number of alleles corrected for sampling size based on a rarefaction method)
194 were estimated using the function divBasic implemented in the R package diveRsity
195 (Keenan et al. 2013). Allelic frequencies, expected heterozygosity (H_e) and fixation index
196 (F_{IS}) were estimated using GENEPOP on the web (Raymond and Rousset 1995, Rousset

2008). The raw SNP data was visualized by a color diagram using the software INTROGRESS (Gompert and Buerkle 2010). Alleles derived from each of the parental populations were counted at the individual level and converted into a matrix of counts, used to visualize the multilocus genotype of each individual.

Genetic structure among sampling sites was depicted using Principal Component Analysis (PCA) computed on the matrix of genotypes, and with an individual-based Bayesian clustering method. The aim of these two analyses was to identify and describe genetic clusters using different statistical approaches. The PCA were carried out using the R package ADEGENET 1.4–2 (Jombart 2008, Jombart et al. 2011). The individual-based Bayesian clustering analysis was performed with the software STRUCTURE 2.3.4 (Pritchard et al. 2000, Falush et al. 2003). For each value of K (ranging from 1 to 25), 30 replicate chains of 150,000 Monte Carlo Markov Chain (MCMC) iterations were run after discarding 50,000 burn-in iterations. An admixture model with correlated allele frequencies was applied with *a priori* information on samples origin. Note that this method makes the assumption of homogeneous admixture rate in the genome (neutral admixture) and therefore return a sort of weighted average admixture rate when introgression is heterogeneous across the genome. To determine individual ancestry proportions (Q-values) that best matched across all replicate runs, we used CLUMPP (Jakobsson and Rosenberg 2007) and visualized individuals' assignment under the R software.

PCAs and STRUCTURE were useful to identify different genetic clusters and to illustrate the lack of genetic structure among samples that belong to these clusters. Based on this, individuals were subsequently pooled into clusters with a minimum of seven individuals per cluster, and genetic structure was computed among clusters by calculating global and pairwise F_{ST} (Weir and Cockerham 1984) using GENEPOP on the web (Rousset 2008). Exact tests for population differentiation (10,000 dememorization steps, 500 batches and 5,000 iterations per batch) were carried out to test for differences in allele frequencies. Q-values, defined as the adjusted p-values using an optimized False Discovery Rate approach, were computed using the QVALUE package in the R software (Storey 2002) to correct for multiple testing.

Population trees were generated using the Neighbor Joining algorithm based on Nei's genetic distances (Nei 1972) in the software PHYLIP 6.695 (Felsentein 2005). Hossegor (site 6 in Fig. 1A) was removed from the analysis, as in this present site, individuals from both North and South genetic clusters were observed.

Evolutionary history of genetic clusters was also investigated under a model of divergence and admixture events using the population graph approach implemented in the TREEMIX software (Pickrell and Pritchard 2012). This software uses the covariance matrix of allele frequency between pairs of populations to infer both populations' splits and gene flow. A maximum likelihood population tree is first generated under the hypothesis of an absence of migration, and admixture events are sequentially added, improving (or not) the tree model. This statistical method shows the benefit of constructing population trees while testing for gene flow between diverged populations at the same time. Hossegor (site 6 in Fig. 1A) and Chalkida (site 24 in Fig. 1A) were removed from these analyses, the former showing individuals from both North and South genetic clusters in the same site, the latter generating noise, due to a small sample size (Pickrell and Pritchard 2012). Using the total data set, *i.e.* 286 loci, five migration events

were sequentially added to look for the best tree to fit the data and the number of migration events that reached an asymptotic likelihood was kept.

Outlier detection

We first measured the genome-wide distribution (measured by the Lewontin-Krakaeur statistic) of genetic differentiation and the neutral expectation to investigate any departure from neutrality. The Lewontin-Krakaeur statistic was measured for each locus with LK=(d-1)(K-1)$Fst/\overline{Fst}$, where d=2 is the number of populations and K=2 is number of alleles. We then fitted a Generalized Extreme Value Distribution using the Broyden-Fletcher-Goldfarb-Shanno algorithm to visualize highly differentiated loci between populations.

The use of several independent methods is recommended to improve accuracy of outlier loci detection (de Villemereuil et al. 2014). Each method is differently impacted by the genetic structure and/or demographic history of the study species, which is usually unknown, leading to frequent inconsistencies across methods (Gagnaire et al. 2015). To cope with these problems, outlier loci were detected using two different methods. First, BAYESCAN, which implements a method developed by Foll and Gaggiotti (2008), is a Bayesian method that uses a logistic regression model to estimate directly the posterior probability that a given locus is under selection. Briefly, locus population F_{ST} values are decomposed as a linear combination of a locus specific component α and a population specific component β. A locus is assumed to be under selection when α is necessary to explain the observed pattern of diversity. For each locus, a reversible jump MCMC explores two alternative models (with or without α) and estimates their relative posterior probabilities. We used default parameter values in our analyses to detect outliers among the clusters previously identified. The second approach used was the one developed by Duforet-Frebourg et al. (2014) and implemented in the R package PCAdapt (Luu et al. 2016). Briefly, it is based on a hierarchical model where population structure is first depicted using K factors. No *a priori* on the genetic structure (and thus, no clustering) is here required in advance. Loci that are atypically related to population structure, measured by the K factors, are identified as outliers. For each value of K (ranging from 1 to 10), 10 replicate chains of 150,000 Monte Carlo Markov Chain iterations were performed and we discarded the first 5,000 iterations as burn-in. PCA, Structure, and Treemix analyses were repeated on sub-datasets including 1) only outlier loci (*i.e.* 14 loci), and 2) only non-outlier loci (*i.e.* 272 loci). Identical parameters as described above were used.

We tested for genotype discordance among loci in an admixed sample by conducting a genomic cline analysis, using Barton's concordance method as described in Macholán et al. (2011). Basically, the method fits a quadratic function to the relationship between a single locus hybrid index (0 or 1 if homozygous, 0.5 if heterozygous) and the genome-wide hybrid index. The function parameters measure the deviation from x = y (*i.e.* the expectation of homogeneous genome-wide introgression) as a function of the expected heterozygosity. More than testing for the deviation from the diagonal, our aim here was instead to compare genomic clines (*i.e.* regression curves) between geographic samples. The reason was that discordance was observed in one population and not others.

Finally, we tested for gene ontology (GO) terms enrichment to determine if outlier loci displayed functional enrichment, compared to the full dataset, using Fisher's

Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg 1995) implemented in the software Blast2GO (Conesa and Götz 2008).

Results

SNPs characterization/calling

Out of the 384 SNPs, including five SNPs (one mitochondrial and four nuclear) that diagnostically distinguished *H. guttulatus* from *H. hippocampus*, a total of 318 SNPs successfully amplified. Four SNPs were monomorphic resulting in 314 SNPs examined for 292 *H. guttulatus*, among which 286 showed a minor allele frequency above 5%. Note that the markers diagnostic of *H. hippocampus* specimens were used to control the taxonomic status of each fish, but being monomorphic in *H. guttulatus*, were removed from the final dataset.

In order to evaluate the extent of ascertainment bias that may be induced by our procedure for selecting markers, we compared the Joint Site-Frequency Spectra (JSFS) obtained with the original dataset of Romiguier et al. (2014), which is assumed freed from ascertainment bias, and our 314-SNP dataset. JSFSs are shown in Supplementary Information 1 (Fig. SI1). As expected, the data of Romiguier et al. (2014) is principally composed of singletons that represented 35-40% of the SNPs (Fig. SI1A), while this proportion was reduced twofold in our 314-SNP dataset as desired (Fig. SI1B). We efficiently removed rare variants without biasing the frequency spectrum too much as the deficit of singletons was homogeneously compensated by every cell of the JSFSs: the comparison of the two JSFSs (Fig. SI1C) reveals that very few cells apart from singletons have an excess above 5%, suggesting limited ascertainment bias in our SNP panel, except with rare alleles as expected.

A strong genetic structure delineating five clusters

Estimates of mean number of alleles per locus (N_{all}), allelic richness (A_r), expected heterozygosity (H_e), and departure from Hardy-Weinberg equilibrium (F_{IS}), for each study site and each genetic cluster identified, are presented in Table SI1. The gene diversity was globally similar among populations and no significant departure from Hardy-Weinberg equilibrium (HWE) was observed.

The Principal Component Analysis (PCA) revealed a strong genetic structure separating five clusters (Fig. 2A). Along the first two axes (Fig. 2A, left panel), explaining 60.1% and 16.2% of the variance, the PCA showed clear differentiation between North Atlantic (sites 1-5, in red), South Atlantic (sites 7-12, in blue), Mediterranean Sea (sites 13-25, site 16 excluded, in green) and Thau lagoon (site 16, in pink). Hossegor individuals (site 6, in purple) clustered either with the North Atlantic (12 individuals) or with the South Atlantic groups (3 individuals). Mediterranean sites spread out along the third axis (9.9% of the variance explained, Fig. 2A, right panel), with Bizerte (site 20, in gold) being closer to the Thau lagoon on the one hand, and Varna (site 25, in turquoise) standing out from the Mediterranean cloud on the other hand. This pattern of genetic structure was also detected by the STRUCTURE analysis (Fig. 2B). Five clusters were identified, composed by: 1- the North Atlantic sites, 2- the South Atlantic sites, 3- the Mediterranean sites, 4- Thau lagoon and 5- Varna. A gradient of introgression is visible along the Atlantic coasts, with a decreasing proportion of south Atlantic cluster ancestry north to Hossegor (site 6), suggesting asymmetrical gene flow.

While most individuals from Hossegor shared the North Atlantic genetic background, three individuals were genetically assigned to the South Atlantic genetic background. Bizerte (site 20), and to a lesser extent Chalkida (site 24), proved to have mixed ancestries from Thau lagoon and Mediterranean clusters.

Representing the matrix of raw genotyping data (Fig. 2C) also clearly pointed out the delineations between the five detected clusters (highlighted by vertical dotted lines in Fig. 2C). The deepest node of Neighbor Joining tree (Fig. 1D) separated two groups, one composed by the North Atlantic Ocean sites only, and the second by South Atlantic and all Mediterranean Sea sites. Concerning the Mediterranean Sea group, Thau and Bizerte lagoons clustered together, and with Varna (site 25). Mediterranean group and Chalkida clustered together, also separated by one internal node to Varna.

No significant departure from panmixia was observed within each cluster (Table SI1). Further, genetic homogeneity was observed between sites within each cluster (Table SI2). In contrast, Fisher exact tests revealed significant differences in allelic frequencies among clusters *(p*-value < 0.001), with significant differentiation being observed for all pairwise comparisons (F_{ST} ranging from 0.025 between the Mediterranean cluster and Chalkida to 0.257 between the North Atlantic cluster and Varna).

Signature of selection, genetic parallelism and reticulated evolution

In all pairwise comparisons between clusters the genome-wide distribution of genetic differentiation (as measured by the LK statistic) departed from the neutral expectation. This is illustrated in Figure 3 for the comparison between the North Atlantic and South Atlantic clusters and holds for all other pairwise comparisons of clusters (data not shown).

Fourteen loci out of 286 (4.9%) were consistently identified to depart from neutrality in both BayeScan and PCAdapt analyses, using a 5% False Discovery Rate (FDR). These outlier loci were located on different *H. guttulatus* contigs, but six to eight of them were located on a unique chromosome when mapped to one of the seven well-assembled fish genomes -zebrafish aside-, of which scaffolds were placed (Fig. 4). Eight of these outlier loci were consistently found in this unique chromosome in, at least, four of the seven fish genomes. Furthermore, four of these outlier loci (SNPs 232, 312, 87 and 311) mapped to this chromosome for the seven fish genomes (colored in orange and pink in Fig. 4). SNP 29 was also consistently located on this chromosome, except in the Gulf pipefish, for which it mapped to an unplaced scaffold. However, there was little consistency in their position or order along the chromosome among species (Fig. 4). Interestingly, two non-outlier loci also mapped to this unique chromosome in, at least, two of the seven fish genomes (Fig. 4). When blasted against the tiger tail seahorse scaffolds, the fourteen outlier loci were located on eleven scaffolds. Two SNPs (SNPs 29 and 286) mapped to *H. comes* scaffold 8. Similarly, three SNPs (SNPs 87, 212 and 311) mapped to scaffold 154. Out of the eleven scaffolds involving the outliers, three scaffolds (scaffolds 8, 339, 397) consistently mapped to the unique chromosome identified here above, when blatted and basted against the seven well-assembled fish genomes, though with some exceptions. *H. comes* scaffold 8 was not identified in this chromosome only for the Nile tilapia, while scaffolds 339 and 397 were not identified in the medaka and zebrafish and the Nile tilapia and zebrafish, respectively. *H. comes* blasts and blats suggested that SNPs 87, 311, 212, 312 and 232 also mapped to a unique chromosome in

381 this species. Similarly, SNP 29, the most differentiated locus (Fig.5C), mapped to the
382 unique chromosome in the European seabass and the stickleback, though to an unplaced
383 scaffold in the other well-assembled fish genomes, suggesting that this SNP (and SNP
384 286 blasted on the same *H. comes* scaffold) may also be located on this chromosome. A
385 similar pattern holds for SNP 34. Note that four outliers were also found on the same
386 chromosome in the genome of the zebrafish, which is the most divergent species used as
387 an outgroup (Fig. 4). The analysis of gene ontology terms for outlier loci did not reveal
388 any significant functional enrichment.
389 Interestingly, the visualization of raw allele frequency data at individual SNPs
390 (Fig. 2C) nicely revealed how the pattern observed at outlier loci departed from the one
391 observed at other loci, not only in the level of differentiation but also in reinforcing the
392 previously observed groupings of populations. Outliers were classified into five
393 categories based on their allele frequency patterns (Fig. 5): those distinguishing 1- North
394 Atlantic sites, Thau lagoon, Bizerte and, in a lesser extent Chalkida and Varna, from
395 South Atlantic and Mediterranean sites (Fig. 5A) or 2- only from South Atlantic (Fig.
396 5B); 3- those separating Thau lagoon (and in a lesser extent Bizerte, Chalkida and Varna)
397 from all other sites (Fig. 5C), 4- those differentiating North Atlantic sites from all other
398 sites, Thau lagoon included (Fig. 5E); and 5- the SNPs distinguishing Mediterranean sites
399 *vs*. Atlantic sites (Fig. 5F). Note that Hossegor and Chalkida individuals were both split
400 into 2 groups according to their North or South Atlantic / Thau lagoon or Mediterranean
401 genetic background, respectively. The first and most abundant category of outliers
402 pointed out the convergence of allele frequency patterns (*i.e.* genetic parallelism)
403 between North Atlantic and Thau lagoon individuals. Interestingly, most markers of this
404 category were found to map to a single chromosome in our cross-mapping approach
405 using different fish genomes, along with the two markers highly differentiated between
406 lagoon and sea ecotypes (Fig. 4 and 5). Contrasting with a consistent pattern in allelic
407 frequencies in each of the five outlier categories, the non-outlier loci that mapped to a
408 unique chromosome did not show any consistent pattern (Fig. 5D). Note that SNP 242
409 (Fig. 5B) that is an outlier of differentiation between the Atlantic Ocean and the
410 Mediterranean Sea mapped also to this unique chromosome but only in the Gulf pipefish
411 (Fig. 4).
412 In order to visualize parallelism in the extent of genetic differentiation in the
413 whole SNP dataset, allele frequency differences (AFDs) between Mediterranean lagoon
414 and sea locations (x-axis in Fig. 5G) were plotted against AFDs between North and South
415 Atlantic clusters (y-axis in Fig. 5G). As expected, outliers showing genetic parallelism
416 appear in the top right of Figure 5G, showing both elevated genetic differentiation
417 between Mediterranean lagoon and marine sites (x-axis) and between north and south
418 Atlantic sites (y-axis). The third category of outliers (Fig. 5C), also observed along the
419 same chromosome when compared against the other fishes (Fig. 4), revealed a higher
420 differentiation only between habitats in the Mediterranean Sea. The fifth category (Fig.
421 5E) revealed a higher genetic differentiation only between Atlantic lineages.
422 Figure 6 shows the result of the genomic cline analysis obtained with SNP 29, one
423 of the outliers that clustered on the same chromosome, and SNP 130, an outlier between
424 the northern and the southern lineage but that maps to another chromosome and which is
425 not differentiated in the Mediterranean Sea (no genetic parallelism). Similar results were
426 obtained with other loci from the same categories. We found that the regressions were

427 different at SNP 29, or other outliers that clustered on the same chromosome, when we
428 included the Hossegor sample in the analysis or not (Fig. 6), while regressions were
429 always close to the diagonal with SNP 130, or other outliers that maps to other
430 chromosomes. This analysis demonstrates the discordance between the clustered outlier
431 loci and the rest of the genome in the Hossegor sample, in which north Atlantic-like
432 seahorses have the south Atlantic allele at clustered outlier.
433 The PCA performed with the fourteen outliers differed from the PCA based on
434 286 SNPs. Bizerte, Varna, the North Atlantic sites and Chalkida (Greece) were found
435 closer to lagoon along the first axis (59.5% of the variance explained; Fig. SI2A).
436 STRUCTURE analysis identified three clusters (data not shown), composed by 1- the
437 North Atlantic sites, 2- the South Atlantic and Mediterranean sites, and 3- Thau lagoon,
438 Bizerte and, on a lesser extent Varna that have mixed ancestries from both North Atlantic
439 and Thau lagoon while Chalkida revealed ancestries of Thau lagoon and the
440 Mediterranean clusters. Genetic affinities between the North Atlantic sites and Thau
441 lagoon also emerged in the Neighbor Joining tree (Fig. 1C): Thau and Bizerte lagoons
442 branched together, and with the North Atlantic populations to form a group composed of
443 Mediterranean lagoons and North Atlantic populations.
444 In contrast, the PCA based on 272 neutral SNPs showed a very similar pattern to
445 the PCA obtained with 286 SNPs along the three first axes (10.2%, 6.0% and 2.3% of the
446 variance explained; Fig. SI2B). STRUCTURE analysis identified four clusters (data not
447 shown), with Varna clustering with the Mediterranean sites. Two groups of populations
448 emerged based on 272 neutral SNPs in the Neighbor Joining tree (Fig. 1B), one
449 composed by the Atlantic Ocean samples, the second by all Mediterranean Sea samples.
450 Neutral SNPs seemed reflecting geography, although Isolation By Distance was not
451 significant (Mantel test p-value = 0.23).
452 Finally, the population tree inferred using TREEMIX without accounting for
453 migration (Fig. 2D, left panel) was highly consistent with all above analyses, with
454 Atlantic clusters branching together on one hand, and Thau and Bizerte lagoons
455 branching together on the other hand. Interestingly, three admixture events significantly
456 improved the model as compared to a situation without migration (p-value < 0.001; Fig.
457 2D, right panel). This population tree indicated significant gene flow among four *H.*
458 *guttulatus* clusters, between the north and the south genetic clusters in the Atlantic coasts,
459 in agreement with the gradient of introgression observed in Figure 2B, between marine
460 and lagoon samples in the Mediterranean Sea, and finally between the North Atlantic and
461 Mediterranean lagoon samples. We suggest that this latter observation is most probably
462 mainly driven by parallel outlier loci (see discussion below).
463
464 **Discussion**
465 Genetic analyses of the long-snouted seahorse revealed that this species is
466 subdivided into several cryptic semi-isolated lineages. Strong genetic subdivisions
467 delineate genetic clusters that otherwise showed a remarkable genetic homogeneity over
468 large areas, despite limited dispersal abilities and seemingly small population sizes (but
469 see Curtis and Vincent 2006 in Ria Formosa, Portugal). The spatial organization of the
470 different genetic backgrounds proved to be an unusual combination of geographic
471 subdivision in the Atlantic Ocean, and genetic structure related to the sea-lagoon
472 ecological contrast in the Mediterranean Sea. Surprisingly, these two subdivisions partly

473 relied on the same genetic architecture. This was supported by the finding of genetic
474 parallelism at some markers showing extreme levels of differentiation between habitats in
475 the Mediterranean Sea, being also amongst the most differentiated between geographic
476 lineages in the Atlantic Ocean. Intriguingly, these loci showing convergent allele
477 frequency patterns all mapped to a single chromosome in seven well-assembled fish
478 genomes. Our results thus suggest the existence of a shared evolutionary history of
479 speciation between Atlantic parapatric lineages and Mediterranean ecotypes. Namely, the
480 Mediterranean lagoon ecotype was anciently related to the north-Atlantic lineage,
481 whereas the south-Atlantic lineage shared ancestral variation with the Mediterranean
482 marine ecotype. The underlying reproductive isolation mechanisms may thus involve a
483 combination of intrinsic and extrinsic genetic barriers, where relative contributions may
484 differ between the two contexts. We propose that the genomic island (*i.e.* the part of the
485 genome gathering outlier loci in a unique chromosome) shared in the two alternative
486 spatial contexts is mainly involved in intrinsic reproductive isolation and have coupled
487 with adaptation to a mosaic environment in the Mediterranean Sea whereas it would have
488 been trapped by an extrinsic barrier to dispersal or have coupled with an alternative
489 environmental heterogeneity in the Atlantic Ocean.
490
491 *1- Efficient genetic analysis in a species with low polymorphism*
492 The long-snouted seahorse has a very low genetic diversity, ranking among the
493 animal species with the lowest synonymous nucleotide diversity (πs=0.0016, Romiguier
494 et al. 2014). For comparison, its level of genetic diversity is similar to that observed in
495 less diverse primates, including humans (Perry et al. 2012), or in the giant Galapagos
496 tortoise (Loire et al. 2013). Although RADseq (Baird et al. 2008), RNAseq (Gayral et al.
497 2013) or targeted capture (Jones and Good 2016) provide a tool of choice in non-model
498 organisms for genetic analyses, this reduced genome representation typically retrieves
499 only a few thousands of SNPs in low-diversity species (~4,000 SNPs in *H. guttulatus*
500 with RADseq, unpublished data). Most of these markers are rare variants providing
501 relatively little information. The strategy implemented here relied on a selection of SNPs
502 discovered from six transcriptomes (Romiguier et al. 2014), ending up with a dataset of
503 286 informative SNPs. The allele frequency distributions of the newly developed SNPs
504 were not significantly different from those derived from the original SNP dataset,
505 suggesting a limited ascertainment bias with rare alleles being underrepresented.
506 Selecting SNPs characterized from a preliminary population transcriptomic survey
507 therefore proved to be a straightforward strategy for genome-wide investigation of the
508 spatial distribution of well-characterized genetic clusters in this protected species. Since
509 our genotyping method only requires a small amount of DNA, it allows the use of non-
510 lethal fin-clipping sampling techniques (Woodall et al. 2012). Moreover, the approach
511 generated no missing data compared to classical genotyping-by-sequencing approaches.
512
513 *2- Genome scans in hybrid zones*
514 Our SNP panel allowed us to define five strongly differentiated panmictic genetic
515 clusters, two in the Atlantic Ocean, two in the Mediterranean Sea, and one in the Black
516 Sea, which cannot be morphologically distinguished with reliability so far (*i.e.* cryptic
517 lineages). The average genetic differentiation between clusters and its associated variance
518 were strong ($0.1 \pm 0.02 < F_{ST} < 0.25 \pm 0.04$). In this context, identifying outlier loci is a

complex task with problems of false positives on the one hand (Lotterhos and Whitlock 2014, 2015). Indeed, the neutral model can easily underestimate the variance of differentiation. On the other hand, it may result in misinterpretations, because many loci of various sorts interact together to generate a semi-permeable barrier to gene flow (Bierne et al. 2011, 2013). Although we used recent outlier tests designed to account for complex demography and spatial population structure, we are likely to deviate from an implicit but basic assumption of outlier test approaches - that most loci are evolving neutrally without being affected by the indirect effect of selection through linkage (Bierne et al. 2013). However, genomic data increasingly support the old prediction that genetic barriers to gene flow can affect a substantial proportion of the genome (Harrison and Larson 2016). When explicitly tested with inferential methods that account for among-locus variation in effective migration rates, semi-permeable barriers to gene flow prove to be widespread (Roux et al. 2016). It is also increasingly acknowledged that signal of local adaptation should not be easily captured without a broad sampling of the genome (Hoban et al. 2016), suggesting that the outliers easily identified with a limited number of loci could reflect the existence of a cryptic genetic barrier to gene flow involving many selected loci (Bierne et al. 2011). However, the most differentiated loci can still provide useful information, if we simply consider them as the best available markers of differentiated genetic backgrounds.

For the seahorse as for many other species however, the genetic structure displayed by outlier loci strongly departs from the average genetic structure, and shows a signature of genetic parallelism (Fig. 2A –C, and Fig. SI2). Therefore, the signal of outlier loci (especially the A category in Fig. 5A - B) is not just consistent with what is expected for the most differentiated loci of a neutral distribution. Instead, they display a genetic structure that is not compatible with the sample covariance matrix. Besides, when compared to well-assembled fish genomes, these outlier loci mapped to a single chromosome in all species (Fig. 4). Although interchromosomal rearrangements are relatively uncommon in teleosts (Naruse et al. 2004, Small et al. 2016), the order was not so well conserved among species. Nonetheless this result suggests the existence of a large genomic island of differentiation as already reported in other fishes (e.g. in sticklebacks: Jones et al. 2012, in cod: Hemmer-Hansen et al. 2013, in seabass: Tine et al. 2014), and possibly involving a chromosomal inversion or another form of recombination suppression.

3- Parapatrically distributed semi-isolated cryptic lineages in the Atlantic Ocean

In the Atlantic Ocean we confirmed the existence of two well-differentiated genetic clusters previously identified by Woodall et al. (2015) with five microsatellites and two mitochondrial genes. Our broader sampling of the genome allowed us to capture genome regions with very strong genetic differentiation. The most differentiated SNP is nearly fixed between the two lineages (SNP 130 in Fig. 5, F_{ST} = 0.9), and many other SNPs demonstrate an abrupt genetic shift between the two clusters (Fig. 2C and 5). Another crucial new observation was the co-existence of the two lineages in the same lagoon in South West of France (Hossegor lagoon, sample 6 in Fig. 1), with an absence of early generation hybrids in the sample but some indirect evidence of hybridization and introgression (Fig. 2B). Such a zone of co-existence, with two genetic backgrounds in sympatry in the absence of genetic differentiation at a large spatial scale within each

565 lineage, suggests strong reproductive isolation (Jiggins and Mallet 2000). This may either
566 be due to pre-zygotic isolation or strong selection against hybrids at early life stages.
567 Nonetheless, and although early generation hybrids were not detected in our sample, the
568 decreasing proportion of south lineage ancestry from Hossegor to the English Channel
569 provides evidence for asymmetric introgression (Fig. 2B). The signal was genome-wide
570 (Fig. 2C) and was also detected by the TREEMIX analysis.
571 However, an interesting pattern was observed at the genomic island outliers.
572 North Atlantic-type individuals sampled at Hossegor hold a different genetic signal at the
573 genomic island loci. The frequency of the southern lineage allele was similar between
574 North Atlantic-type individuals from Hossegor and South Atlantic samples (Hossegor
575 pale pink dot (site 6A) in Fig. 5, see also Fig. 2C). Figure 6 provides a statistical test of
576 this discordance and evidence for hybridization, as it requires recombination between
577 parental backgrounds to produce North Atlantic-type individuals with the southern allele
578 at the genomic island. Genomic cline analysis is usually performed to test whether the
579 quadratic regression of a locus hybrid index to the genome-wide hybrid index deviates
580 from the diagonal, which is the expectation when introgression occurs at the same rate for
581 all loci (Macholán et al. 2011). Here we included the geographic information to contrast
582 the results obtained when the Hossegor sample is included or not in the analysis, as our
583 objective was to infer the existence of a local discordance, in the Hossegor lagoon only.
584 Figures 6 shows that North-Atlantic seahorses, with a high genome-wide hybrid index,
585 are South Atlantic-type at the genomic island in Hossegor (low hybrid index at SNP 29),
586 but are North Atlantic-type elsewhere (high hybrid index at SNP 130). This result is also
587 an indirect evidence for epistatic or coupling interactions among loci implied in
588 reproductive isolation. Endogenous post-zygotic selection against hybrids does not result
589 in a stable polymorphism in a single isolated population. Also known as bistable variants,
590 the maintenance of underdominant or epistatically interacting genetic incompatibilities
591 requires a migration-selection balance in a spatially subdivided population (Barton and
592 Turelli 2011), or frequency-dependent selection (Barton and de Cara 2009). We can
593 hypothesize that the system has fixed one state of a bistable variant in the Hossegor
594 lagoon. One may alternatively suggest that the southern allele confers a selective
595 advantage in the Hossegor lagoon environment whatever the genetic background,
596 however this hypothesis does not seem coherent with the fact that the northern allele is
597 always in higher frequency in lagoon-like habitats elsewhere.
598 Overall, the Atlantic contact zone possesses all the characteristics of standard
599 clinal hybrid zones that follow the tension zone model (Barton and Hewitt 1985); *i.e.* a
600 secondary contact zone maintained by a balance between migration and selection against
601 hybrids due to intrinsic reproductive isolation. Exogenous selection is nonetheless likely
602 to contribute, as the environments are quite obviously different between the two
603 distribution ranges. However, the two Atlantic lineages both inhabit indifferently lagoon
604 and sea habitats (Fig. 1A) suggesting that this ecological contrast plays little role in their
605 genetic. In any case, only strong intrinsic reproductive isolation can guarantee a genome-
606 wide barrier to gene flow explaining the co-existence of the two lineages in sympatry in
607 the Hossegor lagoon.
608
609 *4- Sea and lagoon ecotypes in the Mediterranean Sea*

610 Contrasting with the Atlantic hybrid zone, the two cryptic lineages identified in
611 the Mediterranean Sea were associated with lagoon/sea habitat variation. Our broader
612 genomic and spatial sampling revealed two crucial observations: (i) although we only
613 obtained samples from two lagoons (Thau in France and Bizerte in Tunisia, Fig. 1A),
614 this was sufficient to reveal an association with the environment that was previously
615 misinterpreted as possible spatial differentiation, because the Thau lagoon was the only
616 sample from Western Mediterranean basin in Woodall et al. (2015), and (ii) we observed
617 fixed differences between lagoon and marine samples that are only few kilometers apart
618 in the Gulf of Lion (Fig. 2 and 5). While the marine lineage was surprisingly
619 homogeneous over the whole Mediterranean Sea, from Greece to Spain (F_{ST} = 0.0078
620 n.s.), lagoon-like samples showed different genetic compositions. The Thau lagoon
621 sample showed a strong and genome-wide genetic differentiation from sea samples (Fig.
622 2). In addition, a single but important seahorse was sampled on the seaside of the Thau
623 lagoon (site 15 in Fig. 1A), and seven others sampled on the seaside of another lagoon of
624 the region (Leucate, site 14 in Fig. 1A). They all proved to belong to the marine genetic
625 cluster, without any sign of introgression of the lagoon background. This result suggests
626 strong reproductive isolation as well as strong habitat specialization of the two lineages
627 around French Mediterranean lagoons. However, the sample from the Bizerte lagoon
628 (Tunisia, sample 20 in Fig. 1), which is ecologically similar to the Thau lagoon (Sakka
629 Hlaili et al. 2008), is genetically similar to marine Mediterranean samples at most loci
630 and only share the genetic composition of the Thau lagoon at the genomic island and
631 other sea-lagoon outlier loci (Fig. 2, 5). In addition, a subsample of the Chalkida
632 population (Greece, sample 24B in Fig. 5) was composed of five individuals
633 heterozygous at the genomic island. The environmental condition of the Chalkida
634 location is hypothesized to be more lagoon-like, and this sample only provides evidence
635 that the genomic island polymorphism and the mosaic spatial structure extends to the
636 eastern basin without really providing further clues about the role of the environment.
637 Our results in the Mediterranean Sea (*i.e.* lagoon/maritime system) resemble those
638 obtained in the emblematic three-spined sticklebacks marine/freshwater system (Jones et
639 al. 2012), or more recently in marine migratory and freshwater resident lampreys
640 (Rougemont et al. 2017) or coastal-marine ecotypes in European anchovies (Le Moan et
641 al. 2016). When a shared divergent genomic island is observed –*i.e.* genetic parallelism,
642 as here in long-snouted seahorse and in the examples cited above, there could be three
643 possible interpretations: (i) parallel gene reuse from a shared ancestral polymorphism
644 present in the marine supposedly ancestral population (Jones et al. 2012), (ii) the spread
645 of a locally adapted allele (*i.e.* the 'transporter hypothesis'; Schluter and Conte 2009) or
646 (iii) secondary contact followed by spatial reassortment of the divergent lineages and
647 extensive introgression swamping such that only selected loci and their chromosomal
648 neighborhood retain the history of adaptation (Bierne et al. 2013). The three scenarios
649 are difficult to discriminate as they converge toward a similar pattern (Johannesson et al.
650 2010, Bierne et al. 2013, Welch and Jiggins 2014). In addition, the debate can also be
651 understood as a semantic dispute between a phylogenetic view of populations defined by
652 the genome-wide majority tree and the genetic view of lineages defined by the tree of
653 adaptive/speciation genes (Mallet et al. 2016). The Bizerte lagoon population can either
654 be interpreted as a marine background (defined by the majority tree) introgressed by the
655 lagoon allele at the genomic island, or as a lagoon background (defined by

656 adaptive/speciation genes) massively introgressed by neutral marine alleles.
657 Incorporating heterogeneous migration rates in demographic inference methods allowed
658 Le Moan et al. (2016), Rougeux et al. (2016) and Rougemont et al. (2017) to identify the
659 signal of a secondary contact history carried by islands of differentiation in lampreys,
660 white fishes and anchovies. Unfortunately our 286-SNPs dataset does not allow
661 performing such historical demographic reconstruction, although the TREEMIX analysis
662 reveals that episodes of secondary admixtures strongly improve the fit to the sample
663 covariance matrix. Anyhow, demographic reconstruction does not completely refute the
664 'transporter hypothesis' which stipulates freshwater alleles spread from lagoon to lagoon
665 (or river to river) and is a scenario that produces a very similar genomic pattern of
666 differentiation to the one produced by a standard secondary contact (Bierne et al. 2013,
667 Rougemont et al. 2016). In the case of the seahorse system, however, we made the new
668 observation that genetic parallelism is observed despite an apparent absence of a
669 common selective pressure, which offers complementary arguments to the debate.
670
671 *5- Genetic parallelism in two different spatial/ecological contexts*
672 Now that we have discussed the cryptic differentiation observed in the Atlantic
673 Ocean, on the one hand, and the Mediterranean Sea, on the other hand, we address
674 potential links stemming from a shared history between Atlantic and Mediterranean
675 lineages, underlying genetic parallelism. Clearly, the most astonishing result of our
676 genetic analysis was genetic parallelism between the Mediterranean lagoon ecotype and
677 the north Atlantic lineage at a large genomic island. Parallel evolution usually implies
678 distinct ecological characteristics (e.g. Butlin et al. 2014). In the present study, we found
679 that the genomic island was associated with the sea-lagoon ecological contrast in the
680 Mediterranean Sea, while there was no such genetic differentiation between lagoon and
681 sea samples in the Atlantic Ocean. In addition we found evidence from a genomic cline
682 analysis that the genomic island could be involved in intrinsic reproductive isolation in
683 the Hossegor lagoon hybrid zone. Shared ancestral polymorphism sieved by adaptation in
684 a patchy environment (Bierne et al. 2013) and incipient speciation (Guerrero and Hahn
685 2017 BioRxiv) would better explain our data. The association with habitat in the
686 Mediterranean Sea and with space in the Atlantic Ocean could be explained either by (i)
687 the secondary evolution of locally adapted genes within the genomic island in the
688 Mediterranean Sea, benefiting from the barrier to gene flow imposed by intrinsic
689 selection (divergence hitchhiking; Via 2012, Yeaman 2013) or (ii) alternatively the island
690 could have coupled with local adaptation polymorphism localized elsewhere in the
691 genome in the Mediterranean Sea (Bierne et al. 2011) while it would have been trapped
692 by a barrier to dispersal in the Atlantic Ocean (Barton 1979, Barton and Hewitt 1985).
693 Without further data and the true genomic position of loci in the seahorse genome, it is
694 difficult to disentangle the two hypotheses.
695 Two outlier loci (SNP 87 and SNP 311, Fig. 5C) localized in the same
696 chromosome as convergent SNPs (Fig. 4) are differentiated between the marine and
697 lagoon ecotypes in the Mediterranean Sea, but are not differentiated between the northern
698 and southern lineages in the Atlantic Ocean. At first sight, they could be interpreted as
699 evidence for secondary local sweeps in Mediterranean lagoons. However, as the position
700 is not well inferred from available genomes and the barrier to gene flow is less effective
701 in the Atlantic, they could also be interpreted as being localized in the island "shoulders"

(*i.e.* loci in the vicinity of the regions harboring local adaptation and/or reproductive isolation loci; Gagnaire et al. 2015, Le Moan et al. 2016) in which a stronger introgression rate has erased the differentiation faster in the Atlantic than in the Mediterranean. According to the gene order inferred from the closest species (Gulf pipefish and Tiger tail seahorse), these two loci could potentially be in an island shoulder (Fig. 4C). Importantly, whatever the explanation -divergence hitchhiking or coupling- it requires invoking interaction between intrinsic and ecological selection and not ecological selection alone (Bierne et al. 2011, Kulmuni and Westram 2017). Furthermore, intrinsic selection has most probably evolved first in this system as no genetic parallelism was observed in outliers discriminating lagoon to marine ecotypes. This contradicts the predominant view that ecological selection is necessarily the initial catalyzer of the chain of accumulation of barriers in ecological speciation.

Acknowledgments

We are grateful to Lucas Beranger, Michel Cantou, Philippe Lenfant, Pablo Liger, CPIE Bassin de Thau, Patrick Lelong, Francesco Di Liello and Stéphane Auffret for their help in providing *H. guttulatus* fin-clippings and to Fabienne Moreau for the BeadXpress experiment. Many thanks to Nicolas Duforet-Frebourg for computational advices. This work was funded by a Languedoc-Roussillon Region "Chercheur(se)s d'avenir" grant to NB (Connect7 project), by a LabEx CeMEB postdoctoral fellowship to FR and by Chocolaterie Guylian and a Natural Environment Research Council Industrial Case studentship (NER/S/C/2005/13461) to LW.

Literature

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLOS ONE 3:e3376

Barton, N. 1979. Heredity - Abstract of article: The dynamics of hybrid zones. Heredity 43:341–359.

Barton, N. H., and M. A. R. de Cara. 2009. The evolution of strong reproductive isolation. Evolution 63:1171–1190.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of Hybrid Zones. Annu. Rev. Ecol. Syst. 16:113–148.

Barton, N. H., and M. Turelli. 2011. Spatial Waves of Advance with Bistable Dynamics: Cytoplasmic and Genetic Analogues of Allee Effects. Am. Nat. 178:E48–E75.

Bhagwat, M., L. Young, and R. R. Robison. 2012. Using BLAT to Find Sequence Similarity in Closely Related Genomes. Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al 0 10:Unit10.8

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. 57:289–300.

Bierne, N., P. Borsa, C. Daguin, D. Jollivet, F. Viard, F. Bonhomme, and P. David. 2003. Introgression patterns in the mosaic hybrid zone between Mytilus edulis and M. galloprovincialis. Mol. Ecol. 12:447–461.

Bierne, N., D. Roze, and J. J. Welch. 2013. Pervasive selection or is it…? why are FST outliers sometimes so frequent? Mol. Ecol. 22:2061–2064.

Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. Mol. Ecol. 20:2044–2072.

Bouchemousse, S., C. Liautard-Haag, N. Bierne, and F. Viard. 2016. Distinguishing contemporary hybridization from past introgression with postgenomic ancestry-informative SNPs in strongly differentiated Ciona species. Mol. Ecol. 25:5527–5542.

Boursot, P., J.-C. Auffray, J. Britton-Davidian, and F. Bonhomme. 1993. The Evolution of House Mice. Annu. Rev. Ecol. Syst. 24:119–152.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, J. Turner-Maier, J. Johnson, R. Alcazar, H. J. Noh, P. Russell, B. Aken, J. Alföldi, C. Amemiya, N. Azzouzi, J.-F. Baroiller, F. Barloy-Hubler, A. Berlin, R. Bloomquist, K. L. Carleton, M. A. Conte, H. D'Cotta, O. Eshel, L. Gaffney, F. Galibert, H. F. Gante, S. Gnerre, L. Greuter, R. Guyon, N. S. Haddad, W. Haerty, R. M. Harris, H. A. Hofmann, T. Hourlier, G. Hulata, D. B. Jaffe, M. Lara, A. P. Lee, I. MacCallum, S. Mwaiko, M. Nikaido, H. Nishihara, C. Ozouf-Costaz, D. J. Penman, D. Przybylski, M. Rakotomanga, S. C. P. Renn, F. J. Ribeiro, M. Ron, W. Salzburger, L. Sanchez-Pulido, M. E. Santos, S. Searle, T. Sharpe, R. Swofford, F. J. Tan, L. Williams, S. Young, S. Yin, N. Okada, T. D. Kocher, E. A. Miska, E. S. Lander, B. Venkatesh, R. D. Fernald, A. Meyer, C. P. Ponting, J. T. Streelman, K. Lindblad-Toh, O. Seehausen, and F. Di Palma. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.

Britton-Davidian, J., F. Fel-Clair, J. Lopez, P. Alibert, and P. Boursot. 2005. Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. Biol. J. Linn. Soc. 84:379–393.

Butlin, R. K., M. Saura, G. Charrier, B. Jackson, C. André, A. Caballero, J. A. Coyne, J. Galindo, J. W. Grahame, J. Hollander, P. Kemppainen, M. Martínez-Fernández, M. Panova, H. Quesada, K. Johannesson, and E. Rolán-Alvarez. 2014. Parallel Evolution of Local Adaptation and Reproductive Isolation in the Face of Gene Flow. Evolution 68:935–949.

Christe, C., K. N. Stölting, L. Bresadola, B. Fussi, B. Heinze, D. Wegmann, and C. Lexer. 2016. Selection against recombinant hybrids maintains reproductive isolation in hybridizing Populus species despite F1 fertility and recurrent gene flow. Mol. Ecol. 25:2482-2498.

Conesa, A., and S. Götz. 2008. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int. J. Plant Genomics 2008.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer.

Curtis, J. M. R., and A. C. J. Vincent. 2006. Life history of an unusual marine fish: survival, growth and movement patterns of *Hippocampus guttulatus* Cuvier 1829. J. Fish Biol. 68:707–733.

de Villemereuil, P., É. Frichot, É. Bazin, O. François, and O. E. Gaggiotti. 2014. Genome scan methods against more complex models: when and how much should we trust them? Mol. Ecol. 23:2006–2019.

Doyle, J., and J. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11–15.

Duforet-Frebourg, N., E. Bazin, and M. G. B. Blum. 2014. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. Mol. Biol. Evol. 31:2483–2495.

Duvaux, L., K. Belkhir, M. Boulesteix, and P. Boursot. 2011. Isolation and gene flow: inferring the speciation history of European house mice. Mol. Ecol. 20:5248–5264.

Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3:87–112.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics 164:1567–1587.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Foll, M., and O. Gaggiotti. 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics 180:977–993.

Gagnaire, P.-A., T. Broquet, D. Aurelle, F. Viard, A. Souissi, F. Bonhomme, S. Arnaud-Haond, and N. Bierne. 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. Evol. Appl. 8:769:786.

Gayral, P., J. Melo-Ferreira, S. Glémin, N. Bierne, M. Carneiro, B. Nabholz, J. M. Lourenco, P. C. Alves, M. Ballenghien, N. Faivre, K. Belkhir, V. Cahais, E. Loire, A. Bernard, and N. Galtier. 2013. Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. PLOS Genet. 9:e1003457.

Gompert, Z., and C. Alex Buerkle. 2010. introgress: a software package for mapping components of isolation in hybrids. Mol. Ecol. Resour. 10:378–384.

Good, J. M., M. A. Handel, and M. W. Nachman. 2008. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. Evol. Int. J. Org. Evol. 62:50–65.

Guerrero, R., and M. W. Hahn. 2017. Speciation as a sieve for ancestral polymorphism. bioRxiv 155176.

Harrison, R. G. 1993. Hybrid Zones and the Evolutionary Process. Oxford University Press.

Harrison, R. G., and E. L. Larson. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Mol. Ecol. 25:2454:2466.

Hedrick, P. W. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22:4606–4618.

Hemmer-Hansen, J., E. E. Nielsen, N. O. Therkildsen, M. I. Taylor, R. Ogden, A. J. Geffen, D. Bekkevold, S. Helyar, C. Pampoulie, T. Johansen, FishPopTrace Consortium, and G. R. Carvalho. 2013. A genomic island linked to ecotype divergence in Atlantic cod. Mol. Ecol. 22:2653–2667.

Hewitt, G. M. 1988. Hybrid zones-natural laboratories for evolutionary studies. Trends Ecol. Evol. 3:158–167.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, L. K. Reed, A. Storfer, and M. C. Whitlock. 2016. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am. Nat. 188:379–397.

Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C. Churcher, C. Scott, J. C. Barrett, R. Koch, G.-J. Rauch, S. White, W. Chow, B. Kilian, L. T. Quintais, J. A. Guerra-Assunção, Y. Zhou, Y. Gu, J. Yen, J.-H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd, E. Kenyon, S.

886 Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D.
887 Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R.
888 Clark, C. Riddle, D. Elliot, D. Eliott, G. Threadgold, G. Harden, D. Ware, S. Begum, B.
889 Mortimore, B. Mortimer, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M.
890 Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P.
891 Howden, N. Barker, C. Lloyd, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H.
892 Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K.
893 Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C.
894 Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C.
895 Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A.
896 Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A.
897 Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. D. Cooper, J. Cooper,
898 A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley,
899 J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A.
900 Berger, C. M. Dooley, Z. Ersan-Ürün, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn,
901 J. Konantz, M. Konantz, M. Oberländer, S. Rudolph-Geiger, M. Teucke, C. Lanz, G.
902 Raddatz, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, S. C. Schuster,
903 N. P. Carter, J. Harrow, Z. Ning, J. Herrero, S. M. J. Searle, A. Enright, R. Geisler, R. H. A.
904 Plasterk, C. Lee, M. Westerfield, P. J. de Jong, L. I. Zon, J. H. Postlethwait, C. Nüsslein-
905 Volhard, T. J. P. Hubbard, H. Roest Crollius, J. Rogers, and D. L. Stemple. 2013. The
906 zebrafish reference genome sequence and its relationship to the human genome.
907 Nature 496:498–503.
908
909 Hunt, W. G., and Selander, R. K. 1973. Biochemical genetics of hybridisation in
910 European house mice. Heredity, 31, 1133
911
912 Jaillon, O., J.-M. Aury, F. Brunet, J.-L. Petit, N. Stange-Thomann, E. Mauceli, L.
913 Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot, S. Nicaud, D. Jaffe, S. Fisher, G.
914 Lutfalla, C. Dossat, B. Segurens, C. Dasilva, M. Salanoubat, M. Levy, N. Boudet, S.
915 Castellano, V. Anthouard, C. Jubin, V. Castelli, M. Katinka, B. Vacherie, C. Biémont, Z.
916 Skalli, L. Cattolico, J. Poulain, V. de Berardinis, C. Cruaud, S. Duprat, P. Brottier, J.-P.
917 Coutanceau, J. Gouzy, G. Parra, G. Lardier, C. Chapple, K. J. McKernan, P. McEwan, S.
918 Bosak, M. Kellis, J.-N. Volff, R. Guigó, M. C. Zody, J. Mesirov, K. Lindblad-Toh, B.
919 Birren, C. Nusbaum, D. Kahn, M. Robinson-Rechavi, V. Laudet, V. Schachter, F.
920 Quétier, W. Saurin, C. Scarpelli, P. Wincker, E. S. Lander, J. Weissenbach, and H. Roest
921 Crollius. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals
922 the early vertebrate proto-karyotype. Nature 431:946–957.
923
924 Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and
925 permutation program for dealing with label switching and multimodality in analysis
926 of population structure. Bioinforma. Oxf. Engl. 23:1801–1806.
927
928 Jiggins, C. D., and J. Mallet. 2000. Bimodal hybrid zones and speciation. Trends Ecol.
929 Evol. 15:250–255.
930

Johannesson, K., M. Panova, P. Kemppainen, C. André, E. Rolán-Alvarez, and R. K. Butlin. 2010. Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. Philos. Trans. R. Soc. B Biol. Sci. 365:1735–1747.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jombart, T., R. M. Eggo, P. J. Dodd, and F. Balloux. 2011. Reconstructing disease outbreaks from genetic data: a graph approach. Heredity 106:383–390.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Jones, M. R., and J. M. Good. 2016. Targeted capture in evolutionary and ecological genomics. Mol. Ecol. 25:185–202.

Kai, W., K. Kikuchi, S. Tohari, A. K. Chew, A. Tay, A. Fujiwara, S. Hosoya, H. Suetake, K. Naruse, S. Brenner, Y. Suzuki, and B. Venkatesh. 2011. Integration of the Genetic Map and Genome Assembly of Fugu Facilitates Insights into Distinct Features of Genome Evolution in Teleosts and Mammals. Genome Biol. Evol. 3:424–442.

Kasahara, M., K. Naruse, S. Sasaki, Y. Nakatani, W. Qu, B. Ahsan, T. Yamada, Y. Nagayasu, K. Doi, Y. Kasai, T. Jindo, D. Kobayashi, A. Shimada, A. Toyoda, Y. Kuroki, A. Fujiyama, T. Sasaki, A. Shimizu, S. Asakawa, N. Shimizu, S. Hashimoto, J. Yang, Y. Lee, K. Matsushima, S. Sugano, M. Sakaizumi, T. Narita, K. Ohishi, S. Haga, F. Ohta, H. Nomoto, K. Nogata, T. Morishita, T. Endo, T. Shin-I, H. Takeda, S. Morishita, and Y. Kohara. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719.

Keenan, K., P. McGinnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol. Evol. 4:782–788.

Kulmuni, J., and A. M. Westram. 2017. Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. Mol. Ecol. 26:3093-3103.

Larson, E. L., J. A. Andrés, S. M. Bogdanowicz, and R. G. Harrison. 2013. Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. Evol. Int. J. Org. Evol. 67:3653–3661.

977
978 Larson, E. L., T. A. White, C. L. Ross, and R. G. Harrison. 2014. Gene flow and the
979 maintenance of species boundaries. Mol. Ecol. 23:1668–1678.
980
981 Le Moan, A., P.-A. Gagnaire, and F. Bonhomme. 2016. Parallel genetic divergence
982 among coastal-marine ecotype pairs of European anchovy explained by differential
983 introgression after secondary contact. Mol. Ecol. 25:3187-3202.
984
985 Loire, E., Y. Chiari, A. Bernard, V. Cahais, J. Romiguier, B. Nabholz, J. M. Lourenço, and
986 N. Galtier. 2013. Population genomics of the endangered giant Galápagos tortoise.
987 Genome Biol. 14:R136.
988
989 López, A., M. Vera, M. Planas, and C. Bouza. 2015. Conservation Genetics of
990 Threatened Hippocampus guttulatus in Vulnerable Habitats in NW Spain: Temporal
991 and Spatial Stability of Wild Populations with Flexible Polygamous Mating System in
992 Captivity. PLoS ONE 10:e0117538.
993
994 Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and
995 neutral parameterization on the performance of FST outlier tests. Mol. Ecol.
996 23:2178–2192.
997
998 Lotterhos, K. E., and M. C. Whitlock. 2015. The relative power of genome scans to
999 detect local adaptation depends on sampling design and statistical method. Mol.
1000 Ecol. 24:1031–1046.
1001
1002 Lourie, S. A., and A. C. J. Vincent. 2004. Using Biogeography to Help Set Priorities in
1003 Marine Conservation. Conserv. Biol. 18:1004–1020.
1004
1005 Luu, K., E. Bazin, and M. G. B. Blum. 2016. pcadapt: an R package to perform genome
1006 scans for selection based on principal component analysis. bioRxiv 056135.
1007
1008 Macholán, M., S. J. E. Baird, P. Dufková, P. Munclinger, B. V. Bímová, and J. Piálek.
1009 2011. Assessing multilocus introgression patterns: a case study on the mouse X
1010 chromosome in central Europe. Evol. Int. J. Org. Evol. 65:1428–1446.
1011
1012 Mallet, J., N. Besansky, and M. W. Hahn. 2016. How reticulated are species? Bioessays
1013 38:140–149.
1014
1015 Naruse, K., M. Tanaka, K. Mita, A. Shima, J. Postlethwait, and H. Mitani. 2004. A
1016 Medaka Gene Map: The Trace of Ancestral Vertebrate Proto-Chromosomes Revealed
1017 by Comparative Gene Mapping. Genome Res. 14:820–828.
1018
1019 Nei, M. 1972. Genetic Distance between Populations. Am. Nat. 106:283–292.
1020
1021 Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and
1022 consequences. Phil Trans R Soc B 367:332–342.

1023
1024 Perry, G. H., P. Melsted, J. C. Marioni, Y. Wang, R. Bainer, J. K. Pickrell, K. Michelini, S.
1025 Zehr, A. D. Yoder, M. Stephens, J. K. Pritchard, and Y. Gilad. 2012. Comparative RNA
1026 sequencing reveals substantial genetic variation in endangered primates. Genome
1027 Res. 22:602–610.
1028
1029 Pickrell, J. K., and J. K. Pritchard. 2012. Inference of Population Splits and Mixtures
1030 from Genome-Wide Allele Frequency Data. PLoS Genet 8:e1002967.
1031
1032 Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of Population
1033 Structure Using Multilocus Genotype Data. Genetics 155:945–959.
1034
1035 R Development Core Team (2011) R: A language and environment for statistical
1036 computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-
1037 07-0, http://www.R-project.org/.
1038
1039 Rand, D. M., and R. G. Harrison. 1989. Ecological Genetics of a Mosaic Hybrid Zone:
1040 Mitochondrial, Nuclear, and Reproductive Differentiation of Crickets by Soil Type.
1041 Evolution 43:432–449.
1042
1043 Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F.,
1044 Mehlig, B., and A.M. Westram (accepted in Journal of Evolutionary Biology).
1045 Interpreting the genomic landscape of speciation: a road map for finding barriers to
1046 gene flow
1047
1048 Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population Genetics
1049 Software for Exact Tests and Ecumenicism. J. Hered. 86:248–249.
1050
1051 Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R.
1052 Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G.
1053 Tsagkogeorga, A. a.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, and N.
1054 Galtier. 2014. Comparative population genomics in animals uncovers the
1055 determinants of genetic diversity. Nature 515:261–263.
1056
1057 Rougemont, Q., C. Roux, S. Neuenschwander, J. Goudet, S. Launey, and G. Evanno.
1058 2016. Reconstructing the demographic history of divergence between European
1059 river and brook lampreys using approximate Bayesian computations. PeerJ 4.
1060
1061 Rougeux, C., L. Bernatchez, and P.-A. Gagnaire. 2016. Modeling the Multiple Facets of
1062 Speciation-with-Gene-Flow Towards Inferring the Divergence History of Lake
1063 Whitefish Species Pairs (coregonus Clupeaformis). bioRxiv 068932.
1064
1065 Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop
1066 software for Windows and Linux. Mol. Ecol. Resour. 8:103–106.
1067

1068 Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding
1069 Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence.
1070 PLOS Biol. 14:e2000234.
1071
1072 Sakka Hlaili, A., B. Grami, N. Niquil, M. Gosselin, D. Hamel, M. Troussellier, and H.
1073 Hadj Mabrouk. 2008. The planktonic food web of the Bizerte lagoon (south-western
1074 Mediterranean) during summer: I. Spatial distribution under different
1075 anthropogenic pressures. Estuar. Coast. Shelf Sci. 78:61–77.
1076
1077 Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl.
1078 Acad. Sci. U. S. A. 106 Suppl 1:9955–9962.
1079
1080 Small, C. M., S. Bassham, J. Catchen, A. Amores, A. M. Fuiten, R. S. Brown, A. G. Jones,
1081 and W. A. Cresko. 2016. The genome of the Gulf pipefish enables understanding of
1082 evolutionary innovations. Genome Biol. 17:258.
1083
1084 Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S.
1085 Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P.
1086 Nosil. 2014. Stick insect genomes reveal natural selection's role in parallel
1087 speciation. Science 344:738–742.
1088
1089 Storey, J. D. 2002. A direct approach to false discovery rates. J. R. Stat. Soc. Ser. B Stat.
1090 Methodol. 64:479–498.
1091
1092 Szymura, J. M., and N. H. Barton. 1986. Genetic Analysis of a Hybrid Zone Between
1093 the Fire-Bellied Toads, Bombina bombina and B. variegata, Near Cracow in Southern
1094 Poland. Evolution 40:1141–1159.
1095
1096 Teeter, K. C., B. A. Payseur, L. W. Harris, M. A. Bakewell, L. M. Thibodeau, J. E. O'Brien,
1097 J. G. Krenz, M. A. Sans-Fuentes, M. W. Nachman, and P. K. Tucker. 2008. Genome-
1098 wide patterns of gene flow across a house mouse hybrid zone. Genome Res. 18:67–
1099 76.
1100
1101 Tine, M., H. Kuhl, P.-A. Gagnaire, B. Louro, E. Desmarais, R. S. T. Martins, J. Hecht, F.
1102 Knaust, K. Belkhir, S. Klages, R. Dieterich, K. Stueber, F. Piferrer, B. Guinand, N.
1103 Bierne, F. A. M. Volckaert, L. Bargelloni, D. M. Power, F. Bonhomme, A. V. M. Canario,
1104 and R. Reinhardt. 2014. European sea bass genome and its variation provide
1105 insights into adaptation to euryhalinity and speciation. Nat. Commun. 5.
1106
1107 Tyner, C., G. P. Barber, J. Casper, H. Clawson, M. Diekhans, C. Eisenhart, C. M. Fischer,
1108 D. Gibson, J. N. Gonzalez, L. Guruvadoo, M. Haeussler, S. Heitner, A. S. Hinrichs, D.
1109 Karolchik, B. T. Lee, C. M. Lee, P. Nejad, B. J. Raney, K. R. Rosenbloom, M. L. Speir, C.
1110 Villarreal, J. Vivian, A. S. Zweig, D. Haussler, R. M. Kuhn, and W. J. Kent. 2017. The
1111 UCSC Genome Browser database: 2017 update. Nucleic Acids Res. 45:D626–D634.
1112

1113 Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during
1114 ecological speciation-with-gene-flow. Philos. Trans. R. Soc. Lond. B Biol. Sci.
1115 367:451–460.
1116
1117 Vines, T., A. Dalziel, A. Albert, T. Veen, P. Schulte, and D. Schluter. 2016. Cline
1118 coupling and uncoupling in a stickleback hybrid zone. Evolution 70:1023-38.
1119
1120 Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the Analysis of
1121 Population Structure. Evolution 38:1358–1370.
1122
1123 Welch, J. J., and C. D. Jiggins. 2014. Standing and flowing: the complex origins of
1124 adaptive variation. Mol. Ecol. 23:3935–3937.
1125
1126 Woodall, L. C., R. Jones, B. Zimmerman, S. Guillaume, T. Stubbington, P. Shaw, and H.
1127 J. Koldewey. 2012. Partial fin-clipping as an effective tool for tissue sampling
1128 seahorses, Hippocampus spp. J. Mar. Biol. Assoc. U. K. 92:1427–1432.
1129
1130 Woodall, L. C., H. J. Koldewey, J. T. Boehm, and P. W. Shaw. 2015. Past and present
1131 drivers of population structure in a small coastal fish, the European long snouted
1132 seahorse Hippocampus guttulatus. Conserv. Genet. 1–15.
1133
1134 Yeaman, S. 2013. Genomic rearrangements and the evolution of clusters of locally
1135 adaptive loci. Proc. Natl. Acad. Sci. 110:E1743–E1751.
1136
1137 **Data archival location**
1138 SNPs data have been deposited at DRYAD. DOI: *to be completed after the manuscript is*
1139 *accepted for publication*
1140
1141

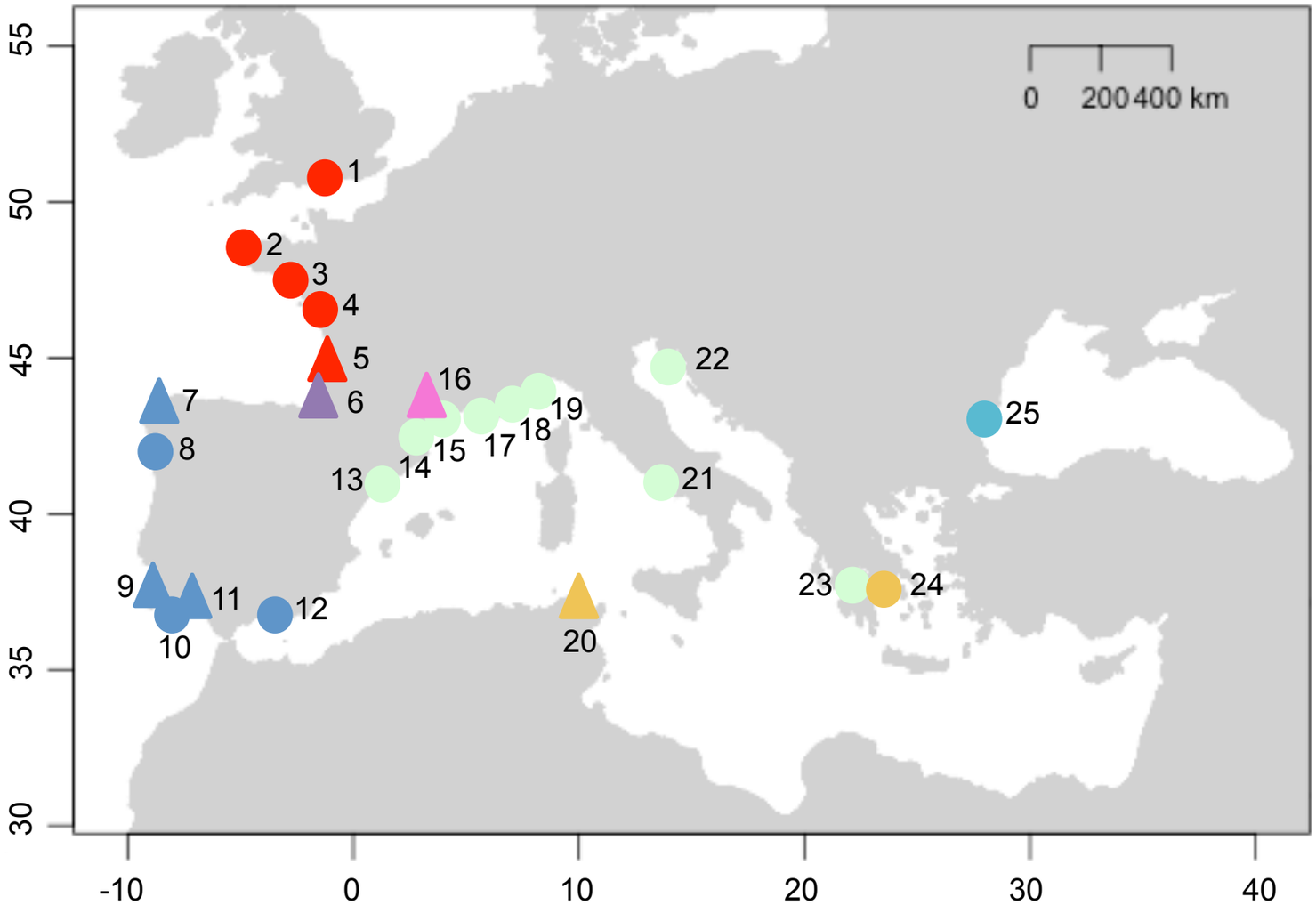
1142 **Figures**
1143
1144 **Figure 1** Sampling locations of *Hippocampus guttulatus* (A-) and Neighbor Joining trees
1145 (B-D-). Each study site is labeled as follow: 1- Poole, United Kingdom, 2- Brest, France,
1146 3- Le Croisic, France, 4- Ré Island, France, 5- Arcachon, France, 6- Hossegor, France, 7-
1147 Coruna, Spain, 8- Vigo, Spain, 9- Portimao, Portugal, 10- Faro (maritime site), Portugal,
1148 11- Faro (lagoon site), Portugal, 12- Malaga, Spain, 13- Tossa de Mar, Spain, 14-
1149 Leucate, France, 15- Sète (maritime site), France, 16- Thau lagoon, France, 17- La Ciotat,
1150 France, 18- Le Brusc, France, 19- Cavalaire-sur-Mer, France, 20- Bizerte lagoon,
1151 Tunisia, 21- Naples, Italy, 22- Rovinj, Croatia, 23- Kalamaki, Greece, 24- Chalkida,
1152 Greece, and 25- Varna, Bulgaria. Red, violet, blue, green, pink, yellow and turquoise dots
1153 stand for North Atlantic cluster, Hossegor, South Atlantic cluster, Mediterranean cluster,
1154 Thau lagoon, Bizerte lagoon and Chalkida, and Varna, respectively. B-D- Neighbor
1155 Joining trees of *H. guttulatus* based on 272 non-outlier SNPs (B-), 14 outliers (C-) and all
1156 markers (D-).
1157
1158 **Figure 2** Genetic population structure analyses based on 286 SNP markers.
1159 A-: Principal Component Analyses depicting axis 1 (explaining 60.1% of the variance)
1160 and axis 2 (explaining 16.2% of the variance; left panel) and axes 1 and 3 (explaining
1161 9.9% of the variance; right panel). Samples are labeled and colored according to Fig. 1.
1162 Each label shows the barycenter of each study site.
1163 B- Individual Bayesian ancestry proportions determined using STRUCTURE with K=5
1164 clusters identified. Dotted black lines separate each study site, which is labeled following
1165 Fig. 1. The five clusters identified are distinguished by different colors: red, blue, green,
1166 pink and turquoise. Each individual is depicted as a vertical bar with colors distinguishing
1167 its ancestries to the five clusters.
1168 C- The raw genotype matrix for a subset of 105 most differentiated SNP markers and all
1169 individuals. Individuals are ordered based on their locations (x-axis). Each rectangle
1170 denotes an individual's genotype at a given locus: dark green (homozygote for the allele
1171 most frequent in the North Atlantic cluster), green (heterozygotes) and light green
1172 (homozygote for the minor allele in the North Atlantic cluster). The names of the outlier
1173 SNPs are mentioned at the right of the matrix. Outliers' names were highlighted
1174 according to the category defined in Fig. 5: the first category of outliers (*i.e.* convergent
1175 SNPs) were represented in orange (A-) and blue dots (B-), the 3rd category in pink (C-),
1176 the 5th category in red (E-), and in green dots the last category of outliers (F-).
1177 D- Population trees inferred by TREEMIX without (left panel) or with 3 migration events
1178 (right panel). Admixture arrows are colored according to the migration weight. The
1179 model including three admixture events significantly improved the fit as compared to a
1180 situation without migration (p-value < 0.001).
1181
1182 **Figure 3** Empirical distributions of the original Lewontin-Krakaeur statistic (grey bars)
1183 for each locus and fitted a theoretical generalized extreme value distribution using BFGS
1184 (red line) with K = 2 alleles per locus and two demes (North and South Atlantic clusters).
1185
1186 **Figure 4** Phylogenomic tree (A-), outlier cross-mapping (B-) and F_{ST} along the unique
1187 chromosome (C-). The phylogenomic tree (A-) reconstructed from Tine et al. (2014) Lin

et al. (2016) and Small et al. (2016) is illustrated with regards to the mapping location of *H. guttulatus* outlier SNPs on eight well-assembled fish genomes (*Syngnathus scovelli*, *Dicentrarchus labrax*, *Gasterosteus aculeatus*, *Oryzias latipes*, *Oreochromis niloticus*, *Takifugu rubripes*, *Tetraodon nigroviridis* and *Danio rerio*) and unplaced genomic scaffolds (*Hippocampus comes*), showing a single matching chromosome for each target species. As *H. comes* scaffolds are unplaced on the genome, *H. guttulatus* SNPs were first blasted on *H. comes* scaffolds, then each scaffold was mapped to the other well-assembled fish genomes. Three scaffolds (339, 397 and 8) mapped to a single chromosome for each species and a putative chromosome was here reconstructed. The order of the scaffolds and SNPs according to the blasts and blats was conserved. Outlier SNPs displaying parallel differentiation between Atlantic lineages and Mediterranean ecotypes (the A and B categories in Fig. 5) are colored in orange and blue, respectively, while outliers highly differentiated between Mediterranean ecotypes (the C category) are colored in red.

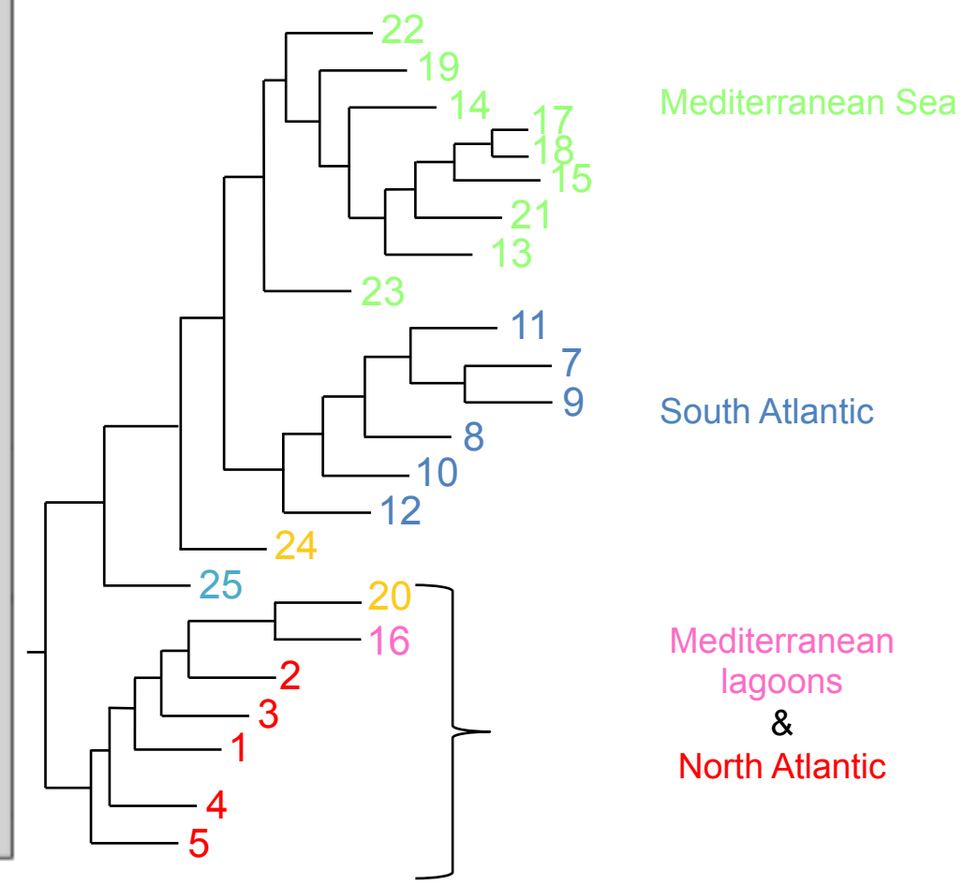
Figure 5 *H. guttulatus* allele frequencies for the 14 outliers identified, along with two neutral SNPs mapped to the genomic island. Each study site is labeled and colored according to Fig. 1, with Hossegor separated in 6A and 6B and Chalkida in 24A and 24B- and colored according to Fig. 1. SNP allele frequencies are categorized in: outliers distinguishing higher frequencies in North Atlantic samples, Thau and Bizerte lagoons and, in a lesser extent Chalkida and Varna (A-) plus Mediterranean samples (B-); Thau lagoon, Chalkida, Varna (C-); North Atlantic samples (E-); and Mediterranean samples to the other sites (F-), and neutral SNPs but mapped to the genomic island (D-). Allelic frequency differences (AFD, G-) between lagoon and sea ecotypes (x-axis) were plotted against AFD of North *vs*. South Atlantic lineages (y-axis) for all SNPs. Outliers were highlighted according to their category: the first two categories of outliers (*i.e.* convergent SNPs) were represented in orange (A-) and blue dots (B-), the 3rd category in pink (C-), the 5th category in red (E-), and in green dots the last category of outliers (F).

Figure 6 Genomic cline plots for two illustrative markers showing extreme level of differentiation (SNPs 29 and 130). Genomic clines were conducted on North and South Atlantic lineages following polynomial regressions on all Atlantic individuals without Hossegor (in yellow), on all Atlantic individuals (in grey) and on all Atlantic individuals, without North Atlantic (in blue). Dotted lines represent the 95% confidence intervals. Circles indicate the raw genotype data (ancestral homozygotes on the top line –the "1" genotype, heterozygotes in the middle –the "0.5" genotype, and derived homozygotes on the bottom line –the "0" genotype).

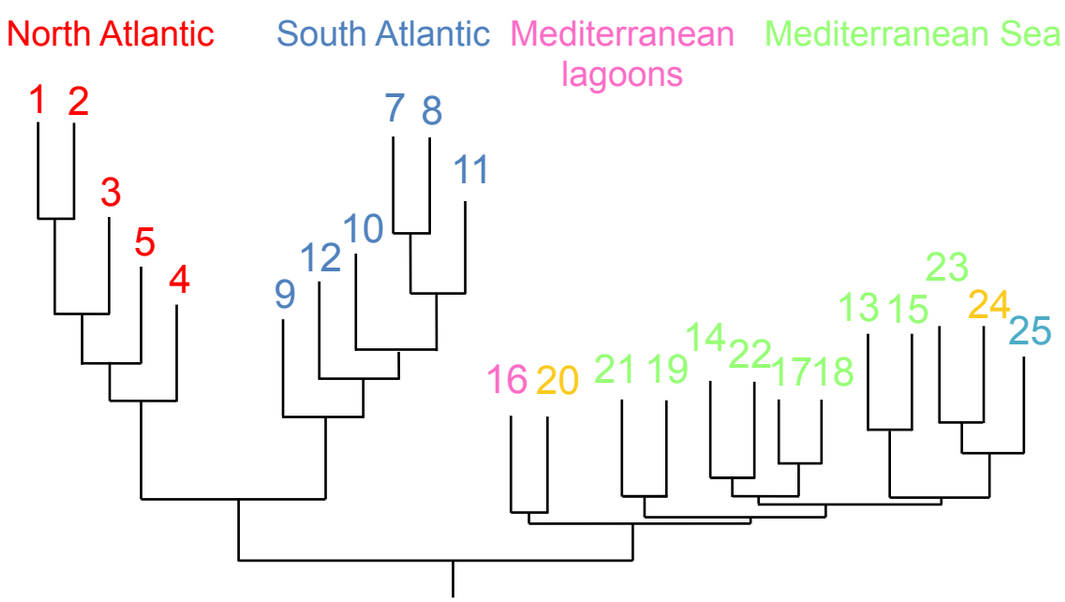
A-



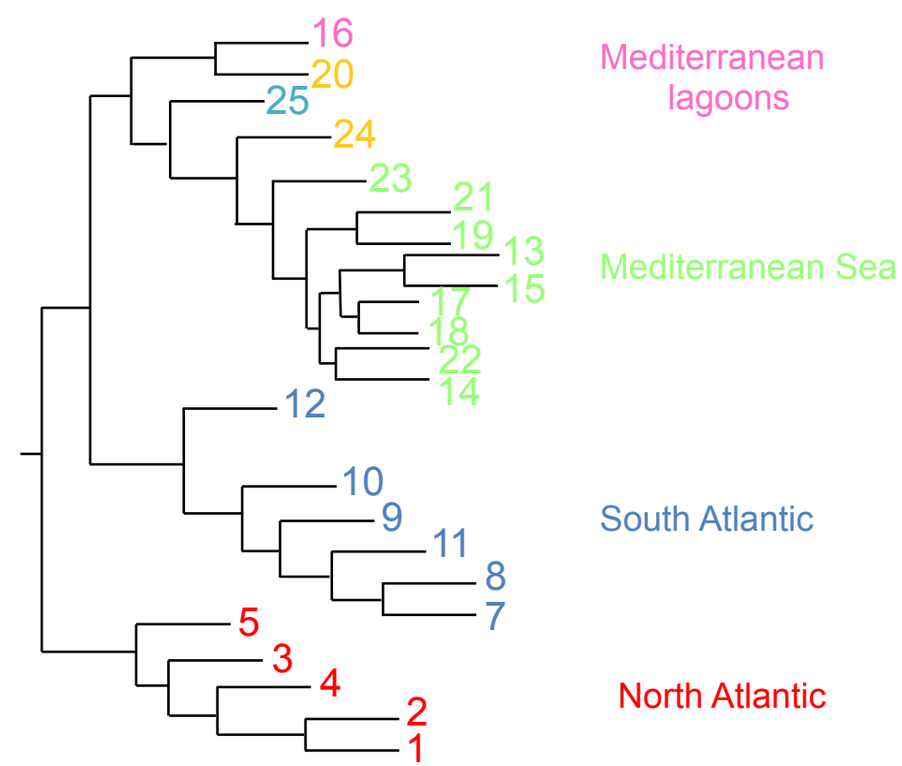
C- Neighbor-joining tree based on 14 outliers

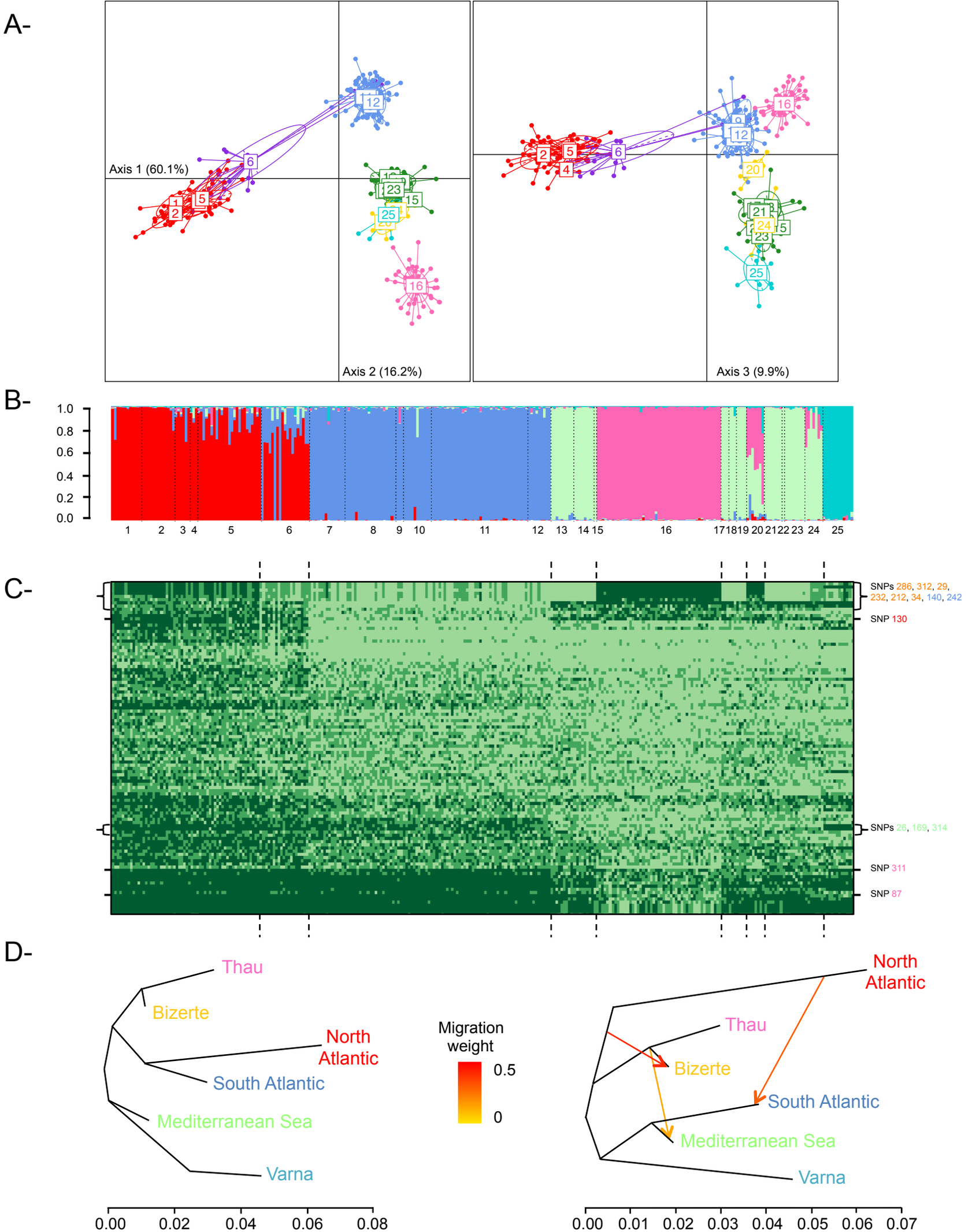


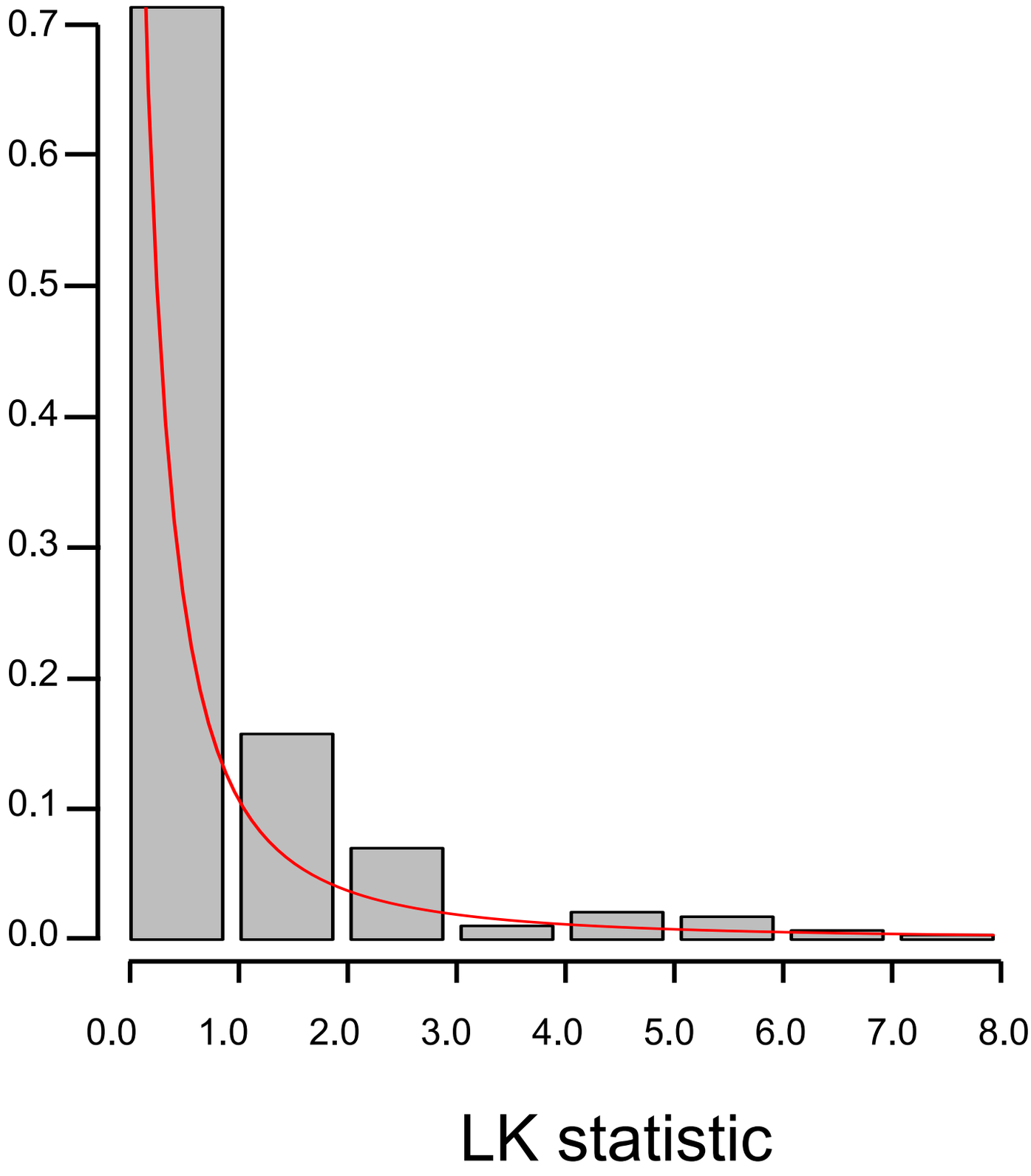
B- Neighbor-joining tree based on neutral markers

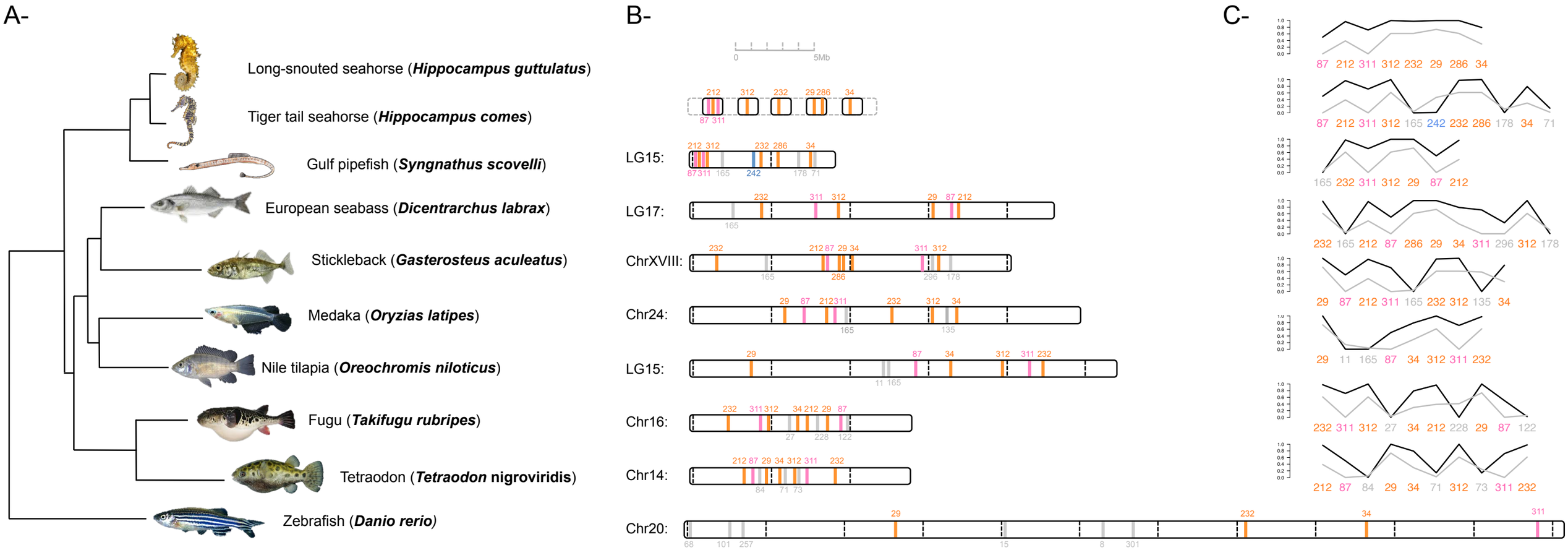


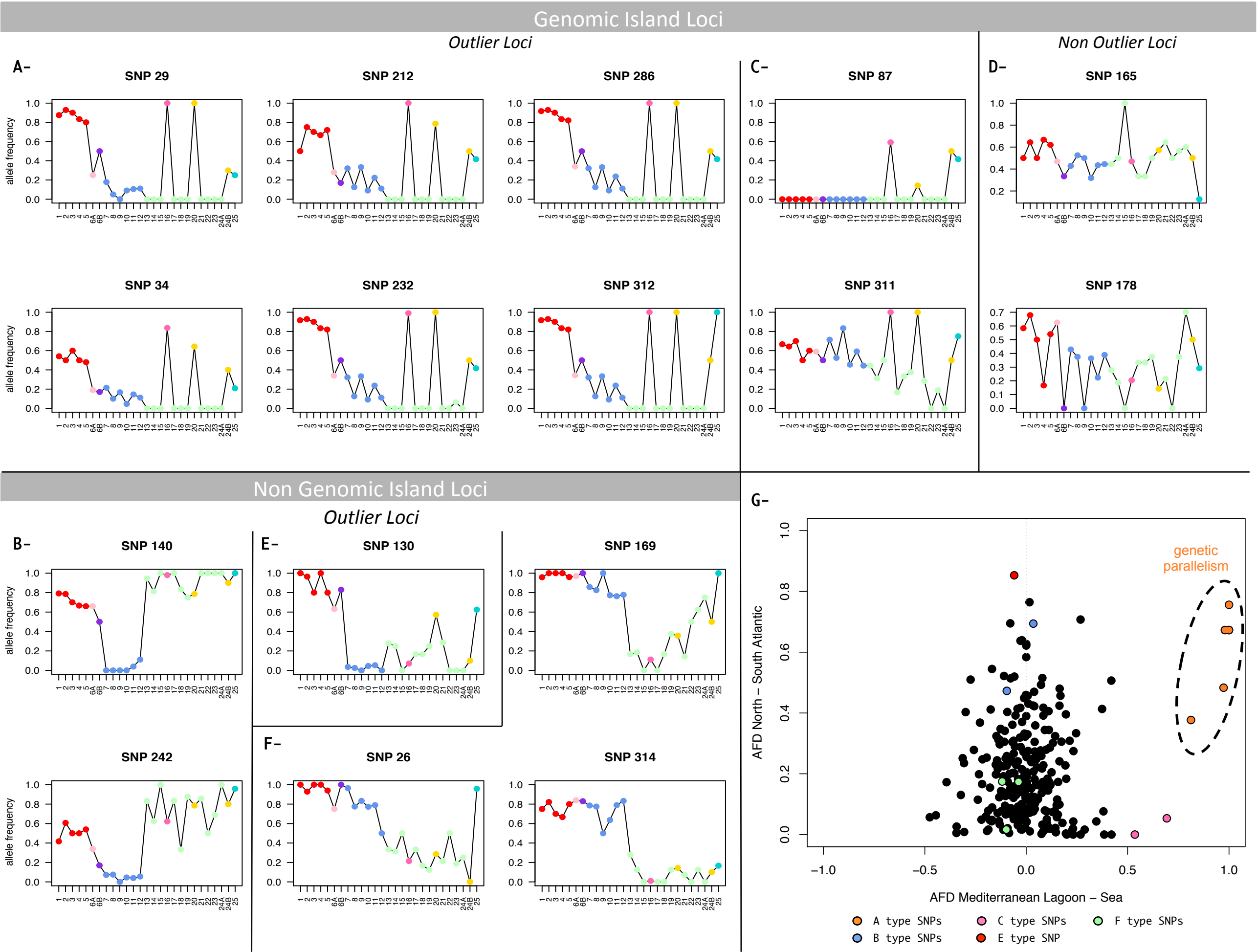
D- Neighbor-joining tree based on all markers

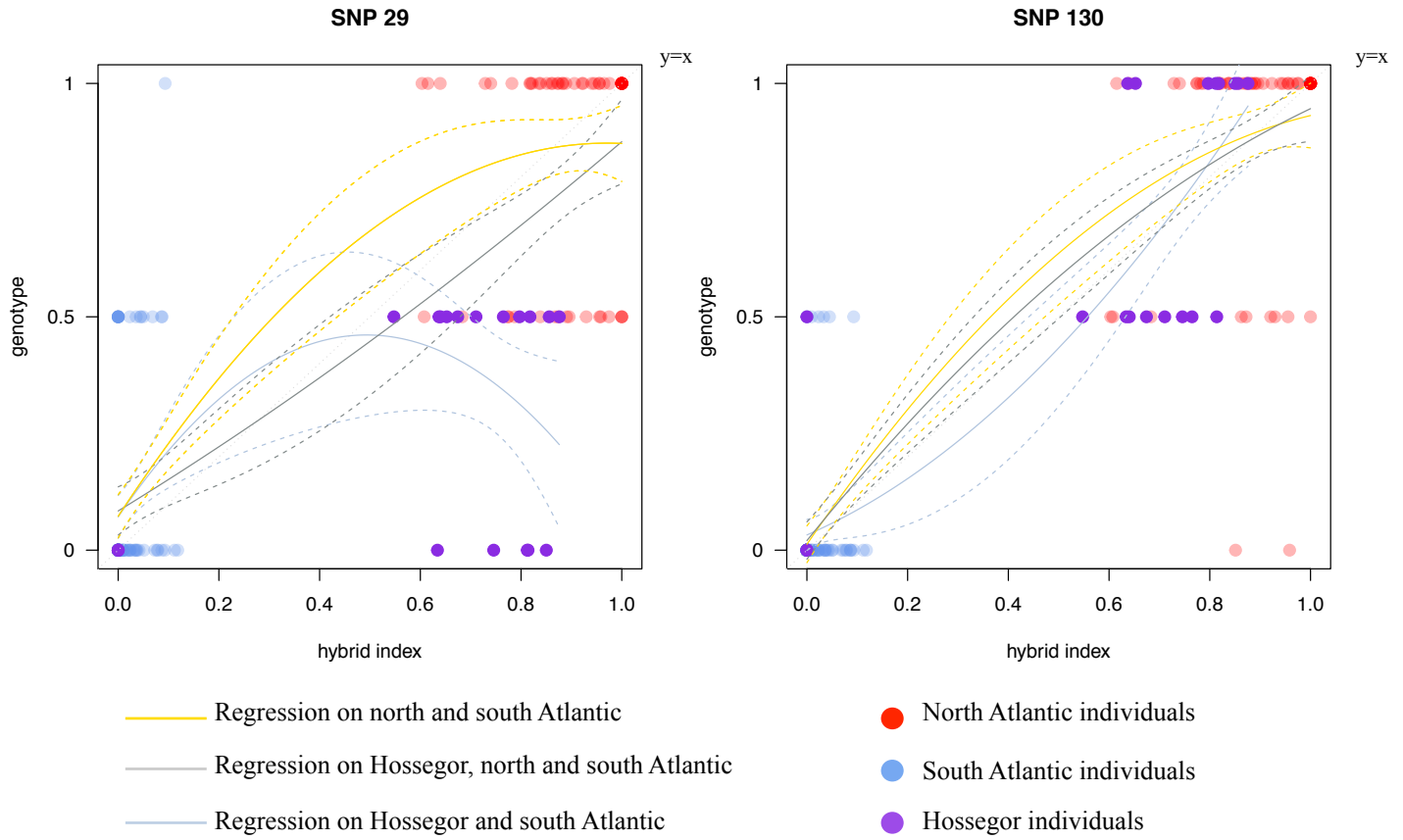



A-
Axis 1 (60.1%)
Axis 2 (16.2%)
Axis 3 (9.9%)
B-
1.0
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
C-
SNPs 286, 312, 29, 232, 212, 34, 140, 242
SNP 130
SNPs 26, 169, 314
SNP 311
SNP 87
D-
Thau
Bizerte
North Atlantic
South Atlantic
Mediterranean Sea
Varna
0.00 0.02 0.04 0.06 0.08
Migration weight
0.5
0
North Atlantic
Thau
Bizerte
South Atlantic
Mediterranean Sea
Varna
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
LK statistic


A-
Long-snouted seahorse (Hippocampus guttulatus)
Tiger tail seahorse (Hippocampus comes)
Gulf pipefish (Syngnathus scovelli)
European seabass (Dicentrarchus labrax)
Stickleback (Gasterosteus aculeatus)
Medaka (Oryzias latipes)
Nile tilapia (Oreochromis niloticus)
Fugu (Takifugu rubripes)
Tetraodon (Tetraodon nigroviridis)
Zebrafish (Danio rerio)
B-
0
5Mb
LG15:
LG17:
ChrXVIII:
Chr24:
LG15:
Chr16:
Chr14:
Chr20:
C-


Genomic Island Loci
Outlier Loci
Non Outlier Loci
A-
SNP 29
SNP 212
SNP 286
SNP 34
SNP 232
SNP 312
C-
SNP 87
SNP 311
D-
SNP 165
SNP 178
allele frequency
Non Genomic Island Loci
Outlier Loci
B-
SNP 140
SNP 242
E-
SNP 130
SNP 169
F-
SNP 26
SNP 314
G-
genetic parallelism
AFD North – South Atlantic
AFD Mediterranean Lagoon – Sea
A type SNPs
B type SNPs
C type SNPs
E type SNP
F type SNPs


SNP 29
SNP 130
y=x
y=x
genotype
hybrid index
0
0.5
1
0.0
0.2
0.4
0.6
0.8
1.0
Regression on north and south Atlantic
Regression on Hossegor, north and south Atlantic
Regression on Hossegor and south Atlantic
North Atlantic individuals
South Atlantic individuals
Hossegor individuals