Parallel pattern of differentiation at a genomic island shared between clinal and mosaic hybrid zones in a complex of cryptic seahorse lineages

This preprint has been reviewed and recommended by Peer Community In Evolutionary Biology (https://dx.doi.org/10.24072/pci.evolbiol.100056).

Running title: Genetic parallelism in seahorse lineages

Authors: 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]

Author's Affiliations:

[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, John Krebs Field Station, Wytham, OX2 8QJ, 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

Corresponding author: Institut des Sciences de l'Evolution de Montpellier, Université Montpellier, Montpellier, France; Florentine Riquet: flo.riquet@gmail.com

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 reduce gene flow between groups. Rarely are these two patterns of spatial distribution reported in the same study system. Here we report that the long-snouted seahorse *Hippocampus guttulatus* is subdivided into discrete panmictic entities by both types of hybrid zones. Along the European Atlantic coasts, a northern and a southern lineage meet in the southwest of France where they coexist in sympatry, -*i.e.* in the same geographical zone- with little hybridization. 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 contrasts with isolation maintained in sympatry or close parapatry at a fine scale. A high variation in locus-specific introgression rates provides additional evidence that partial reproductive isolation must be maintaining the divergence. We find that fixed differences between lagoon and marine populations in the Mediterranean Sea belong to the most differentiated SNPs between the two Atlantic lineages, against the genome-wide pattern of structure that mostly follow geography. These parallel outlier SNPs cluster on a single chromosome-wide island of differentiation. Since Atlantic lineages do not map to lagoon-sea habitat variation, genetic parallelism at the genomic island suggests a shared genetic barrier contributes to reproductive isolation in contrasting contexts -*i.e.* spatial *vs*. ecological. We discuss how a genomic hotspot of parallel differentiation could have evolved and become associated both with space and 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 their relationship with environmental variation was long thought to offer great promises to unravel the nature and origin of species. Though each taxon may be genetically homogeneous over large distances, they often 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 (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 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. Contrasting with this long-standing dichotomy, hybrid zones of both types have now been recognized to be multifactorial and maintained 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 still 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, mosaic distributions suggest local adaptation occurs, and genomic regions of high genetic differentiation between habitats are often interpreted as evidence that repeated local

92 selection increased differentiation in a parallel fashion (Nosil and Feder 2012). This
93 dichotomy, although just two combinations among a multitude (Kirkpatrick and Ravigné
94 2002), is anchored by emblematic study systems for which decades of research allow
95 support for such interpretations. For instance, the hybrid zone between the mice *Mus*
96 *musculus musculus* and *M. m. domesticus* in central Europe (Boursot et al. 1993) has
97 been well demonstrated to be maintained by selection against hybrid genotypes (Britton-
98 Davidian et al. 2005, Good et al. 2008) after secondary contact (Duvaux et al. 2011).
99 Although the position of the hybrid zone was initially found to be associated with rainfall
100 (Hunt and Selander 1973), local adaptation is not considered to contribute much to the
101 isolation. At the other extreme, local adaptation of the same genetic variants has
102 repeatedly allowed marine three-spined sticklebacks to evolve into a freshwater ecotype
103 (Jones et al. 2012). Intrinsic selection is thought absent in the marine-freshwater
104 sticklebacks system (e.g. Dalziel et al. 2012), although selection against hybrids can be
105 strong and hybrids tend to reside in salinity ecotones (Vines et al. 2016). It would be
106 misleading, however, to suggest the two alternative spatial contexts and relations to
107 environmental variation may correspond to alternative routes to speciation (e.g. mutation-
108 order *vs.* ecological speciation). The list of hybrid zones maintained both by local
109 adaptation and intrinsic reproductive isolation is also long. *Bombina* toads (Szymura and
110 Barton 1986), *Gryllus* crickets (Rand and Harrison 1989, Larson et al. 2014), or *Mytilus*
111 mussels (Bierne et al. 2003) are well-known examples of mosaic hybrid zones maintained
112 by exogenous and endogenous selection. However, parallel genetic divergence associated
113 with contrasting environmental conditions (e.g. marine/freshwater, highland/lowland,
114 host races) remains a strong hallmark of ecologically-driven divergence (Bierne et al.

115 2013). Given this -hybrid zone- context, in this paper we aim to provide an example of a
116 genetic parallelism with a lack of apparent ecological convergence. However, as we
117 discovered this pattern by serendipity in a newly studied complex of cryptic genetic
118 backgrounds in a non-model system, we have to describe this system first.

119 We studied the population genetics of the long-snouted seahorse, *Hippocampus*
120 *guttulatus,* across a large part of its geographic range. We developed an assay of 286
121 informative SNPs chosen from more than 2,500 SNPs identified in a population
122 transcriptomic study (Romiguier et al. 2014). *Hippocampus guttulatus* displays poor
123 dispersal abilities (e.g. site-fidelity, weak swimming performance, lack of dispersive
124 stage) and inhabits fragmented coastal habitats along its distribution range (from the
125 English Channel through the Mediterranean and Black Seas, Lourie and Vincent 2004).
126 In addition, most populations are small and have patchy distributions. Given these
127 biological characteristics, a strong genetic structure could have been expected. A very
128 low genetic diversity was observed in *H. guttulatus* when compared to 75 non-model
129 animal species (Romiguier et al. 2014), which could be related to the possibly low
130 population size of this species. However, genetic differentiation proved to be very weak
131 over very large geographic distances based on the two genetic studies conducted to date
132 with microsatellite loci (e.g. from the United Kingdom to North of Spain, Woodall et al.
133 2015, or across the Cape Finisterre oceanographic barrier, López et al. 2015). Four well-
134 differentiated genetic clusters, each distributed over extended regions, were delineated by
135 genetic discontinuities corresponding to usual delimitations between vicariant marine
136 lineages (Woodall et al. 2015) – between the Iberian Peninsula and the Bay of Biscay in
137 the North Eastern Atlantic, between the Atlantic Ocean and the Mediterranean Sea, and

between the Mediterranean and Black Seas. Although such genetic differentiation matches well with oceanographic barriers and was interpreted as spatial differentiation, this pattern is also concordant with the existence of reproductively isolated cryptic lineages, with boundaries that were trapped by exogenous barriers. In this latter interpretation, although the location of genetic breaks would be due to exogenous factors (e.g. temperature, salinity or oceanic fronts), the barrier to gene flow would mainly be driven by barrier loci that restrict gene flow on a large fraction of the genome (*i.e.* the coupling hypothesis; Bierne et al. 2011, Gagnaire et al. 2015, Ravinet et al. 2017). This hypothesis is receiving increasing support (e.g. Le Moan et al. 2016, Rougeux et al. 2016, Rougemont et al. 2016, Saarman et al. 2018) and could well explain the genetic structure observed in the long-snouted seahorse.

Using newly developed SNP-markers spread along the genome and a more extensive sampling along the *H. guttulatus* distribution range compared to Woodall et al. (2015) and López et al. (2015), we challenged the initial interpretation of barriers to dispersal against the alternative hypothesis of reproductive isolation between semi-isolated genetic backgrounds coinciding with oceanographic barriers. We describe five cryptic semi-isolated lineages: two lineages in the Atlantic Ocean with a parapatric distribution, two lineages in the Mediterranean Sea with a patchy fine-grained environment association (lagoon *vs.* marine environments), and one in the Black Sea. We find that 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. Furthermore, the North Atlantic lineage was related to the lagoon ecotype at this genomic island, against the genome-wide

pattern of structure. However, the two Atlantic lineages inhabit both marine and lagoon habitats. We argue that the *H. guttulatus* complex could become one of a few systems where a clinal and a mosaic hybrid zone are observed concomitantly, and a valuable new counter-example that provides evidence of genetic parallelism in absence of ecological convergence.

Materials and Methods

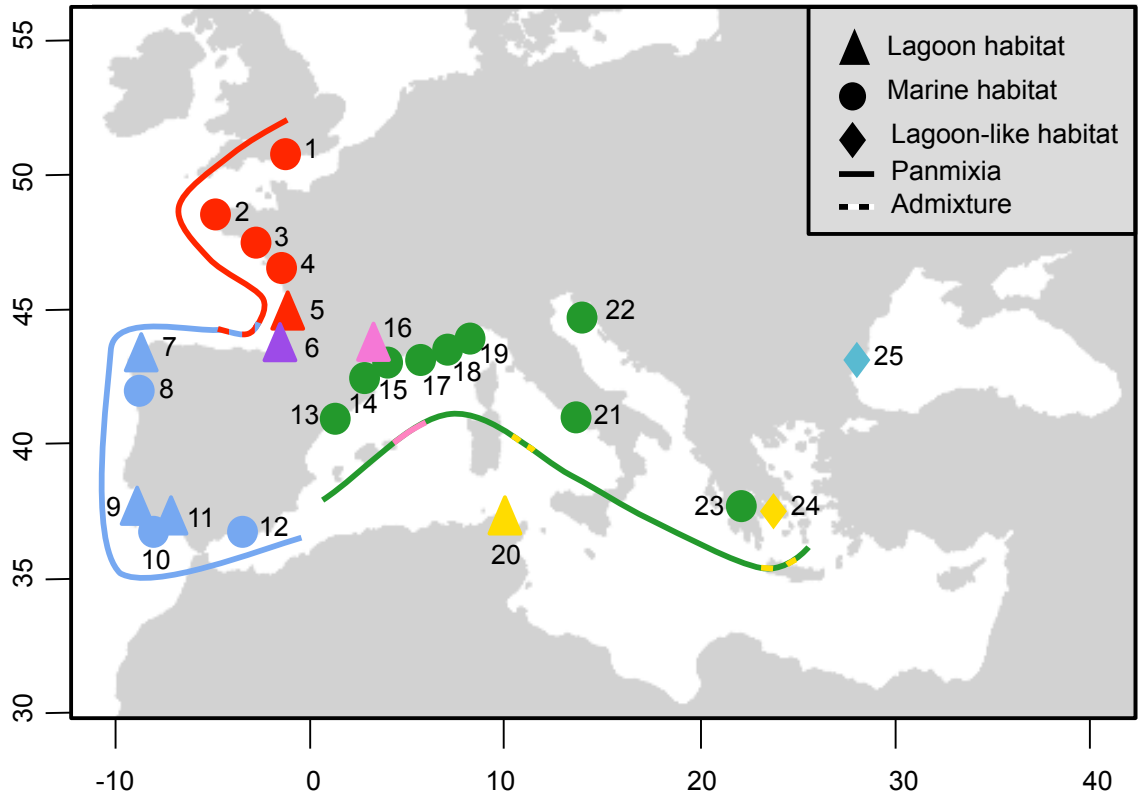


Figure 1 Sampling locations of *Hippocampus guttulatus*. Each study site is labeled as follow: 1- Poole, United Kingdom, 2- Brest, France, 3- Le Croisic, France, 4- Ré Island, France, 5- Arcachon, France, 6- Hossegor, France, 7- Coruña, Spain, 8- Vigo, Spain, 9- Portimão, Portugal, 10- Faro (maritime site), Portugal, 11- Faro (lagoon site), Portugal, 12- Málaga, Spain, 13- Tossa de Mar, Spain, 14- Leucate, France, 15- Sète (maritime site), France, 16- Thau lagoon, France, 17- La Ciotat, France, 18- Le Brusc, France, 19- Cavalaire-sur-Mer, France, 20- Bizerte lagoon, Tunisia, 21- Naples, Italy, 22- Rovinj, Croatia, 23- Kalamaki, Greece, 24- Halkida, Greece, and 25- Varna, Bulgaria. Lagoon habitats are represented by triangles, lagoon-like habitats by diamonds and maritime habitats by circles. Red, blue, green, pink and turquoise symbols stand respectively for the North Atlantic, South Atlantic, Mediterranean Sea, Mediterranean lagoon,

and Varna cluster, all of them showing panmixia (solid lines). Habitat with the co-occurrence of two lineages is colored in violet (site 6) while habitats with admixed individuals are in yellow (20 and 24); dashed lines symbolized these contact zones along the distribution range.

Sampling and DNA extraction

Hippocampus guttulatus samples were collected alive from 25 sites (Fig. 1) using a variety of methods (snorkeling, scuba diving, trawling nets, aquarium, 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. In three sites (sites 14, 18 and 19 in Fig. 1), dorsal fins were clipped from captive-bred seahorse held at the Mare Nostrum Aquarium, France, recently sampled in their natural habitats. Each individual sample was preserved and stored in 96% ethanol for subsequent genetic analyses.

Whole genomic DNA was extracted following either Woodall et al. (2015), López et al. (2015), using a standard CetylTrimethyl Ammonium Bromide (CTAB) Chloroform:Isoamyl alcohol (24:1) protocol (Doyle and Doyle 1987) or 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).

Data mining for SNP markers

A set of 12,613 contigs was examined to identify SNPs. This included one mitochondrial contig (GenBank accession number: AF192664) and 12,612 contigs from Romiguier et al. (2014). Briefly, Romiguier et al. (2014) produced high-coverage transcriptomic data (RNAseq) for six *H. guttulatus* from three locations (Le Croisic, Atlantic Ocean, France; Faro, Atlantic Ocean, Portugal; Thau lagoon, Mediterranean Sea, France), and two *H. hippocampus* from two locations (Sète, Mediterranean Sea, France;

207 Bizerte, Mediterranean Sea, Tunisia). *De novo* transcriptome assembly based on Illumina
208 reads was performed using a combination of the programs ABySS (Simpson et al. 2009)
209 and Cap3 (Huang and Madan 1999), then mapped to predicted cDNAs (contigs) with
210 BWA (Li and Durbin 2009). Based on these 12,613 contigs, SNPs were identified using
211 the bioinformatic pipeline described in Bouchemousse et al. (2016). SNPs were called
212 with Read2SNPs (Gayral et al. 2013) and filtered out according to the following criteria
213 to exclude: 1) SNPs showing more than two alleles, 2) SNPs failing to be sequenced in at
214 least one location, 3) SNPs present in only one individual (*i.e.* singletons), 4) SNPs
215 identified as paralogs using the paraclean option of Read2SNPs, and 5) SNPs closer than
216 20 bp from a contig extremity or an exon limit when blasted against the stickleback, cod
217 and tilapia genomes. This resulted in 2,684 selected SNPs screened with the Illumina
218 Assay Design Tool (ADT) software to select 384 SNPs on the basis of their quality index
219 (ADT score > 0.6). An Illumina BeadXpress® with Veracode™ technology
220 (GoldenGate® Genotyping Assay) was then used to genotype the 384 selected SNPs.
221 To identify their chromosomal positions, the template sequences of the targeted
222 SNP were blasted against i) the genome of the Gulf pipefish (*Syngnathus scovelli*, Small
223 et al. 2016) and ii) the scaffolds of the tiger tail seahorse (*Hippocampus comes*, Lin et al.
224 2016). *H. comes* scaffolds being unplaced, we blasted these scaffolds against *S. scovelli*
225 genome and, for more consistency, aligned them against seven well-assembled fish
226 genomes using BLAT searches (Bhagwatt et al. 2012); zebrafish (*Dano rerio*, Howe et al.
227 2013), fugu (*Takifugu rubripes*, Kai et al. 2011), tetraodon (*Tetraodon nigroviridis*,
228 Jaillon et al. 2004), Nile tilapia (*Oreochromis niloticus*, Brawand et al. 2014), medaka

(*Oryzias latipes*, Kasahara et al. 2007), stickleback (*Gasterosteus aculeatus*, Jones et al. 2012), and European seabass (*Dicentrarchus labrax*, Tine et al. 2014).

SNPs were polarized using *H. hippocampus* as an outgroup to identify the most parsimonious ancestral variant, which allowed the derived allele state to be identified. The Joint Site-Frequency Spectrum (JSFS is the most informative summary statistic regarding inter-population polymorphism; Wakeley 2008) obtained from the original transcriptome-wide SNP dataset (Romiguier et al. 2014) was compared to the JSFS obtained by the subset of 384 SNPs to investigate the extent of ascertainment bias potentially induced by our marker selection. In order to compare the JSFS obtained with both datasets, the JSFS dimension was projected down to a 5x5 matrix, which was the dimension of the Romiguier et al. (2014) dataset.

To detect recombination events, the number of non-overlapping recombinant intervals (NbRec) was estimated by RNAseqFGT using the four-gamete test (FGT) on unphased sequences (see Galtier et al. 2018). Briefly, for each locus, RNAseqFGT performs the FGT on all possible pairs of bi-allelic SNPs to detect recombination events, which required three to four individuals (of the six fish sequenced) with informative genotypes. This analysis was run on all *H. guttulatus* contigs available. R_{fgr} ("Four-Gamete Rule", a good indicator for intragenomic variation in recombination rate; Galtier et al. 2018), defined as the ratio of total number of inferred recombination events by contig length, was subsequently calculated. Finally, each contig, of which R_{fgr} was calculated, was blasted against *S. scovelli* genome (Small et al. 2016), so that mean R_{fgr} was estimated per chromosome or in sliding windows along a chromosome.

252 *Analyses of Genetic Diversity and Genetic Structure*

253 Allelic frequencies, expected heterozygosity (H_e) and fixation index (F_{IS}) were
254 estimated using GENEPOP on the web (Raymond and Rousset 1995, Rousset 2008). To
255 get a genome-wide picture from the multi-locus genotype dataset summarizing inter-
256 population polymorphism, the raw SNP data were visualized by INTROGRESS
257 (Gompert and Buerkle 2010). Alleles derived from each of the two parental populations
258 (*i.e.* the two populations considered as the source or origin for the population and
259 assumed to be fixed for different alleles at most sampled markers) were counted at the
260 individual level and converted into a matrix of counts, then used to visualize the
261 multilocus genotype of each individual.

262 Genetic structure among sampling sites was depicted using both Principal
263 Component Analysis (PCA) computed on the matrix of genotypes, and an individual-
264 based Bayesian clustering method. The latter method is a model-based approach with
265 strong priors and hypotheses (Hardy-Weinberg equilibrium, no linkage disequilibrium)
266 and contrasted with PCA, a distance-based approach for which few (nearly no)
267 assumptions may be violated. Comparing results from both analyses using different
268 statistical approaches allows us to make solid assumptions about our data. The PCAs
269 were carried out using the R package ADEGENET 1.4–2 (Jombart 2008, Jombart et al.
270 2011). The individual-based Bayesian clustering analysis was performed with the
271 software STRUCTURE 2.3.4 (Pritchard et al. 2000, Falush et al. 2003). For each value of
272 *K* (ranging from 1 to 25), 30 replicate chains of 150,000 Markov Chain Monte Carlo
273 (MCMC) iterations were run after discarding 50,000 burn-in iterations. An admixture
274 model with correlated allele frequencies was applied with *a priori* information on sample

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, CLUMPP (Jakobsson and Rosenberg 2007) was used and individuals' assignment visualized in the R software.

Once the different genetic clusters were identified (and cross-validated using both methods), genetic homogeneity among samples belonging to a cluster was checked, allowing subsequent pooling of samples into clusters, which was done with a minimum of seven individuals per cluster. 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.

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 population 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

298 trees while testing for gene flow between diverged populations at the same time. Samples
299 with too small sample size (N<7) or without random mating (see Results) were removed
300 as they generate erroneous results (Pickrell and Pritchard 2012). Using the total data set,
301 *i.e.* 286 loci (see below for the selection of the 286 SNPs out of the 384 SNPs), five
302 migration events were sequentially added to look for the best tree to fit the data, and we
303 retained the number of migration events at which an asymptotic likelihood was reached.
304
305 *Outlier detection*
306 Demographic processes similarly affect neutral markers, but markers linked to
307 loci targeted by natural selection display atypical patterns of variation, key to understand
308 the history of speciation (Bierne et al. 2013). The use of several independent methods is
309 often recommended to improve accuracy of outlier loci detection (Pérez-Figueroa et al.
310 2010, de Villemereuil et al. 2014). Each method is differently impacted by the genetic
311 structure and/or demographic history of the study species, which is usually unknown,
312 leading to frequent inconsistencies across methods (Gagnaire et al. 2015). To cope with
313 these problems, outlier loci were detected using four different methods. First,
314 BAYESCAN (Foll and Gaggiotti 2008) is a Bayesian method that uses a logistic
315 regression model to estimate directly the posterior probability that a given locus is under
316 selection. We used default parameter values in our analyses to detect outliers among the
317 clusters previously identified. The second approach (Duforet-Frebourg et al. 2014)
318 implemented in the R package PCAdapt (Luu et al. 2016) is based on a hierarchical
319 model where population structure is first depicted using K factors. No *a priori* hypothesis
320 for the genetic structure (and thus, no clustering) is 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 one to ten), ten replicate chains of 150,000 MCMC iterations were performed, and we discarded the first 5,000 iterations as burn-in. The third approach uses the estimated co-ancestry matrix to compute an extension of the original Lewontin and Krakauer statistic (Lewontin and Krakauer 1973) that accounts for the history of populations under a model of pure drift (Bonhomme et al. 2010). Finally we used the custom simulation test described in Fraïsse et al. (2014). The idea of this test is to use simulations of the best-supported demographic model to obtain the neutral envelope of the joint distribution of pairwise F_{ST} in a four-population analysis. This test uses the fact that it is easy to have false positives in each of two pairwise comparisons but that outliers in both comparisons, against the genome-wide structure, are more likely to be true positives. Roux et al. (2016) found that an Isolation-with-Migration (IM) model fitted the seahorse data well and that more parameterized models did not improve the fit. These authors also found that the time of divergence (Tsplit) was very similar in each pairwise comparison. We therefore used a four-population IM model with the population sizes and migration rates inferred by Roux et al. (2016) on the transcriptome data of Romiguier et al. (2014). The two latter tests are very similar in their spirit, although one uses a simple history of divergence with drift and the other intends to explore more complex demographic histories. They were here used to identify parallel SNPs that provide a grouping of populations that goes against the genome-wide trend.

Genotype discordance among loci in admixed samples (genomic cline framework) was tested using Barton's concordance method as described in Macholán et

al. (2011). 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. Instead of testing the deviation from the diagonal, our aim was to compare genomic clines (*i.e.* regression curves) between geographic samples, as 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: an efficient genotyping method in a protected species with low genetic diversity

Of the 384 SNPs, a total of 318 SNPs amplified successfully. Of these 318 SNPs, 32 SNPs were removed from the final dataset, being either monomorphic in *H. guttulatus* (the four SNPs used to diagnostically distinguish *H. hippocampus* to *H. guttulatus*) or with a minor allele frequency below 5% (28 SNPs). Two *H. hippocampus* initially identified as *H. guttulatus* were removed from the initial dataset, resulting in a final dataset of 292 *H. guttulatus* genotyped for 286 SNPs.

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) -assumed freed from

368 ascertainment bias- with the 314-SNP dataset, *i.e.* excluding only monomorphic SNPs in
369 *H. guttulatus*. JSFSs are detailed in Supporting Information 1 (Fig. SI1). Briefly,
370 singletons represented 35-40% of the SNPs in the Romiguier et al. (2014) dataset (Fig.
371 SI1A). This proportion was reduced twofold in our 314-SNP dataset (Fig. SI1B).
372 Identifying fewer singletons in the 314-SNP dataset would affect differently the genetic
373 structure. However, we efficiently removed rare variants without biasing the frequency
374 spectrum too much: the deficit of singletons was homogeneously compensated by every
375 other cell of the JSFSs. An even representation over the entire allele frequency range was
376 indeed observed based on our dataset. The comparison of the two JSFSs (Fig. SI1C)
377 reveals that very few cells apart from singletons have an excess above 5%, suggesting
378 limited ascertainment bias in our SNP panel, except with rare alleles as expected.
379 With a limited ascertainment bias, with rare alleles being underrepresented, no
380 missing data, and constraints on our model study (small amount of DNA available with
381 the use of non-lethal fin-clipping sampling techniques), selecting SNPs characterized
382 from a preliminary population transcriptomic survey proved to be a straightforward
383 strategy for genome-wide investigation of the spatial distribution in this species
384 compared to classical genotyping-by-sequencing approaches.
385
386 *A strong genetic structure delineating five broadly distributed panmictic genetic clusters*
387 Estimates of expected heterozygosity (H_e), and departure from Hardy-Weinberg
388 equilibrium (HWE; F_{IS}), for each study site and each genetic cluster identified, are
389 presented in Supporting Information 2. The gene diversity was similar among populations

with no significant departure from HWE observed (with the exception of site 6 which is a zone of co-existence of two lineages, see below).

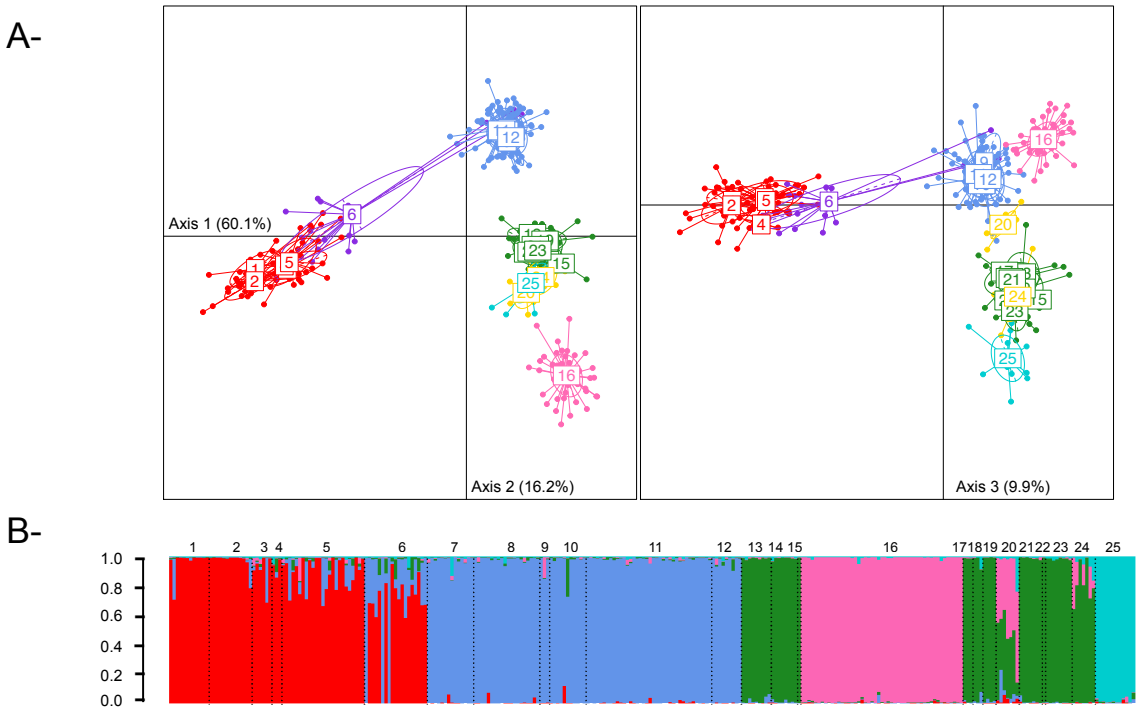


Figure 2 Genetic population structure based on 286 SNP markers analyzed by A- Principal Component Analyses depicting axis 1 (explaining 60.1% of the variance) and axis 2 (explaining 16.2% of the variance; left panel) and axes 1 and 3 (explaining 9.9% of the variance; right panel) with each label showing the barycenter of each study site; and B- Individual Bayesian ancestry proportions determined using STRUCTURE with K=5 clusters identified. Dotted black lines separate each study site. The five clusters identified are distinguished by the same colors and numbers as used in Fig. 1. Each individual is depicted as a vertical bar with colors distinguishing its ancestries to the five clusters.

The Principal Component Analysis (PCA) revealed clear differentiation separating four clusters along the first two axes (60.1% and 16.2 % of the variance explained; Fig. 2A, left panel). A clear differentiation was shown between North Atlantic (sites 1-5, in red), South Atlantic (sites 7-12, in blue), Mediterranean Sea (sites 13-25, lagoon site 16 excluded, in green) and Mediterranean 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

411 out along the third axis (9.9% of the variance explained, right panel), with Bizerte (site
412 20, in gold) in between the South-Atlantic (sites 7-12) and all Mediterranean populations
413 (sites 13-24), and Varna (Black Sea, site 25, in turquoise) standing out from the
414 Mediterranean group on the other hand.

415 A similar genetic structure was detected by the STRUCTURE analysis, of which
416 the output composed of five clusters is presented in Figure 2B. Different values of K
417 were explored (from two to 10) and K=5 provided the best meaningful result. Clustering
418 from STRUCTURE is similar to clustering revealed by PCA, which is a distance-based
419 method that makes no assumptions on our data). Geographic groups comprised: 1- the
420 North Atlantic sites (sites 1-5), 2- the South Atlantic sites (sites 7-12), 3- the
421 Mediterranean Sea sites (sites 13-25, lagoon site 16 excluded), 4- the Mediterranean Thau
422 lagoon (site 16) and 5- Varna (site 25). Similar to the PCA (Fig. 2A), a gradient of
423 introgression is visible along the Atlantic coasts, with a decreasing proportion of South
424 Atlantic cluster ancestry North of Hossegor (site 6; Fig. 2B). Most Hossegor individuals
425 belong to the North Atlantic genetic background, though with mixed ancestry suggesting
426 local introgression. Five individuals were genetically assigned to the North Atlantic
427 genetic background with an ancestry rate higher than 0.82 and three individuals were
428 assigned to the South Atlantic genetic background with ancestry rate higher than 0.98.
429 Bizerte (site 20), and to a lesser extent Halkida (site 24), proved to have mixed ancestries
430 from both Mediterranean lagoon and Sea clusters. Bizerte also appears to have Atlantic
431 ancestries, especially one individual with an Atlantic ancestry rate of 23%. Increasing the
432 K-value to 6 resulted in the addition of Bizerte (site 20) as a new cluster, and increasing
433 the K-value to 7 resulted in the addition of Halkida (site 24) as a new cluster (Figure SI3).

This illustrates introgressive hybridization does not produce strong departure from Hardy-Weinberg and linkage equilibrium in these admixed clusters. Further increases in K did not result in new meaningful geographic clusters.

Altogether, distance-based (PCA) and model-based (Structure) analyses supported the identification of five clusters, a pattern also showed by the visualization of raw mutli-locus genotype data (Supporting Fig. SI4). This representation illustrates that most markers contribute to the signal of five genetic clusters.

Importantly, no significant departure from panmixia was observed within each cluster (SI2). Furthermore, genetic homogeneity was observed between sites within each cluster (SI2). In contrast, Fisher's exact tests revealed significant differences in allelic frequencies among clusters (*p*-value < 0.001; 0.09 ± 0.02 < mean F_{ST} ± sd < 0.26 ± 0.04), with significant differentiation being observed for all pairwise comparisons (SI2-2).

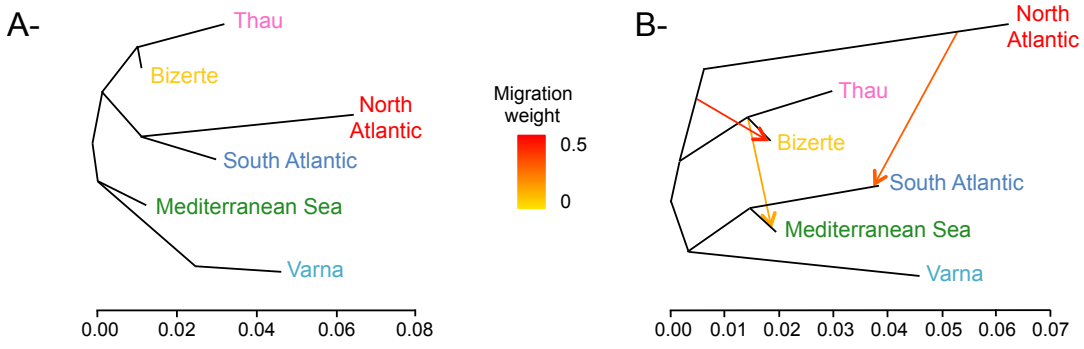


Figure 3 Population trees inferred by TREEMIX (A) without or (B) with 3 migration events. Admixture arrows are colored according to the migration weight. The model including three admixture events significantly improved the fit as compared to a situation without migration (*p*-value < 0.001). Panmictic clusters are colored according to Fig. 1.

Finally, the population tree inferred using TREEMIX without accounting for migration (Fig. 3A) was highly consistent with all above analyses. Atlantic clusters branching together on one hand, and Mediterranean lagoons (Thau and Bizerte) branching together

458 on the other hand. Interestingly, three admixture events significantly improved the model
459 as compared to a scenario without migration (p-value < 0.001; Fig. 3B). This population
460 tree indicated significant gene flow among four *H. guttulatus* clusters, between the north
461 and the south genetic clusters in the Atlantic coasts, in concordance with the gradient of
462 introgression along the Atlantic coasts (Fig. 2), between marine and lagoon samples in
463 the Mediterranean Sea, and finally between the North Atlantic and Mediterranean lagoon
464 samples. Note that, though arrows should indicate directionality of gene flow, when
465 migration is between closely related populations without outgroups, and introgression is
466 heterogeneous in the genome, inferred directions of migration arrows could be erroneous.
467
468 *Signature of selection and genetic parallelism*
469 Nine SNPs out of 286 (3.15%) were consistently identified to depart from
470 neutrality with the four tests (BayeScan, PCAdapt, FLK and custom simulation test; Fig.
471 4). Interestingly, six of them showed very similar allele frequency patterns, distinguishing
472 North Atlantic sites (sites 1-5), Mediterranean Thau lagoon (site 16), Bizerte (site 20)
473 and, in a lesser extent Halkida (site 24) and Varna (site 25) from South Atlantic and
474 Mediterranean Sea sites (Fig. 4), and pointing out genetic parallelism (*i.e.* convergence of
475 allele frequency patterns) between these lineages.

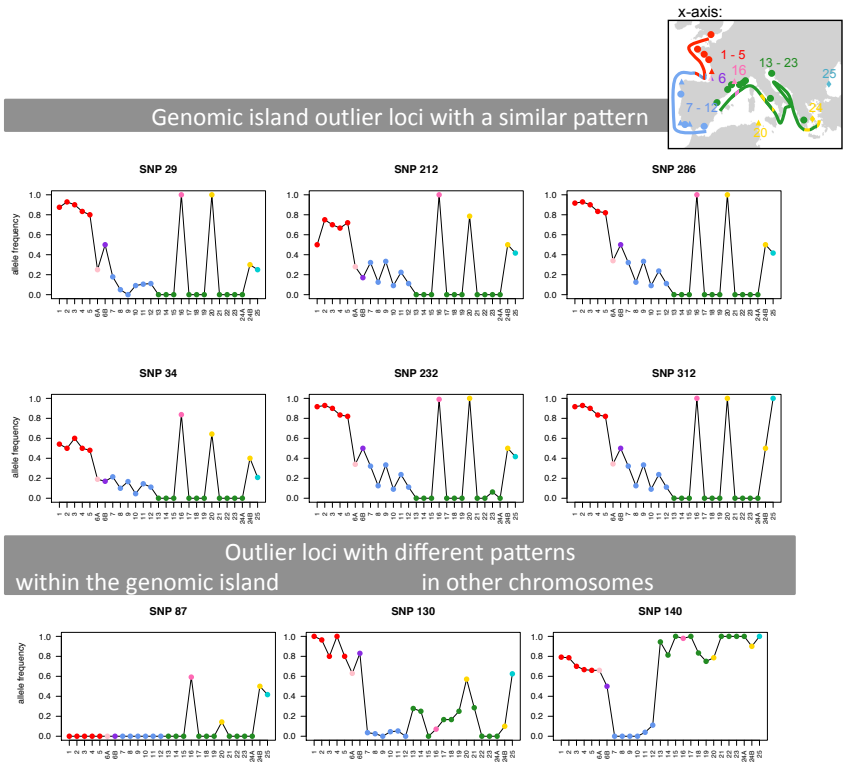


Figure 4 *Hippocampus guttulatus* allele frequencies (y-axis) for the nine outlier SNPs. The six outliers shown on the top of the panel are characterized by a very similar high allele frequency along with a location on a unique chromosome, while the three outliers below are characterized by various allele frequency, different from the six outliers previously mentioned. Each study site (x-axis) is labeled and colored according to Fig. 1, reminded by a simplified map on the top of the figure. Hossegor was separated in 6A and 6B and Halkida in 24A and 24B according to their North or South Atlantic / Mediterranean Thau lagoon or Sea genetic background, respectively (see Fig. 2, SI4).

These six outliers located on different *H. guttulatus* contigs were located on different *Hippocampus comes* scaffolds, except SNPs 29 and 286 mapping to a unique *H. comes* scaffold (Fig. 5A). Interestingly, these scaffolds –that contain outliers- consistently mapped to a unique chromosome in *Syngnathus scovelli* (LG15, Fig. 5A). Results were similar when directly blasting these six *H. guttulatus* contigs against *S. scovelli* genome, but with SNP 29 mapping to an unplaced scaffold (Fig. 5A). A unique chromosome is still involved when blatting *H. guttulatus* outlying contigs against seven additional well-

494 assembled fish genomes, in agreement with a well-conserved synteny of fishes (detailed
495 in SI5).
496

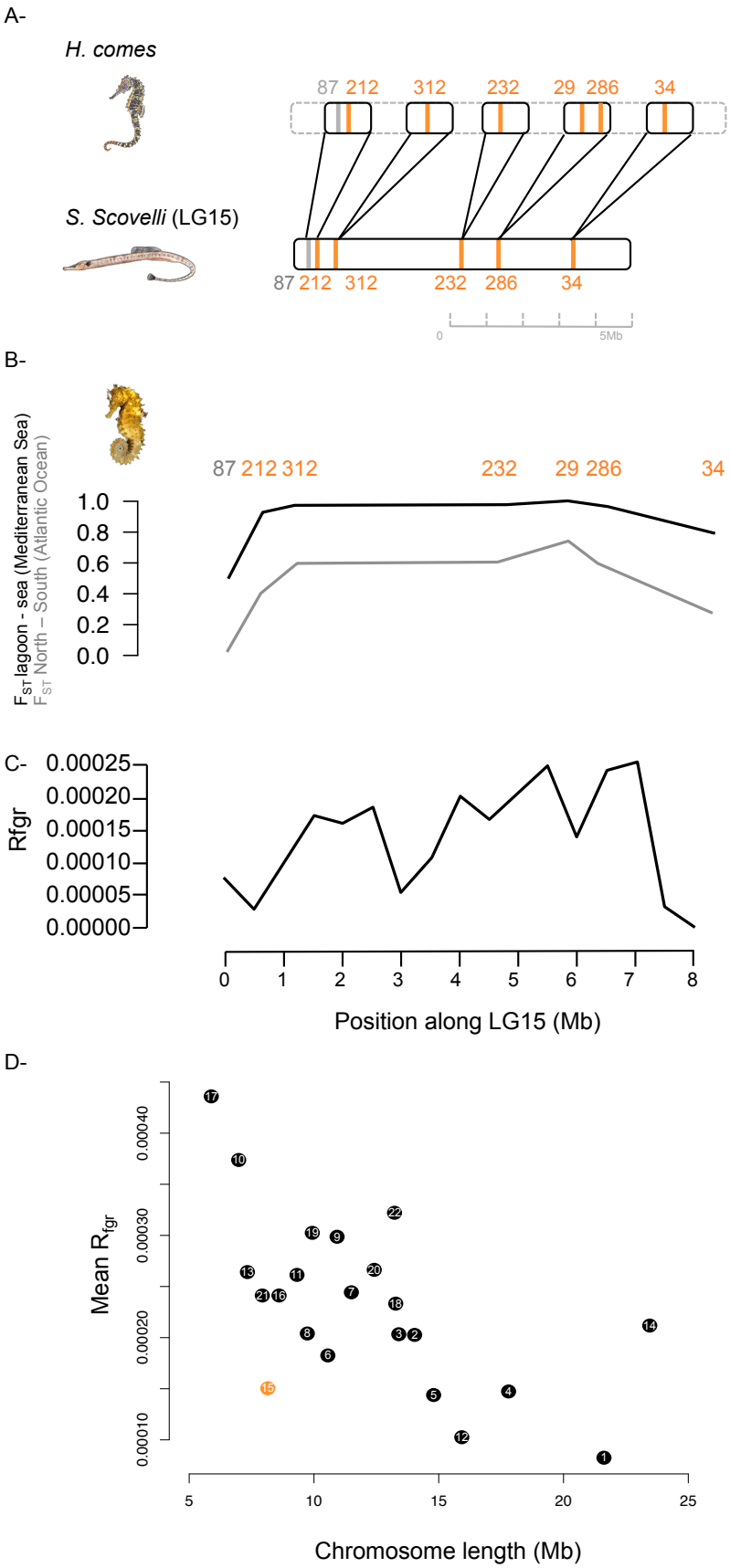

A-
H. comes
S. Scovelli (LG15)
87 212 312 232 29 286 34
87 212 312 232 286 34
0
5Mb
B-
87 212 312 232 29 286 34
1.0
0.8
0.6
0.4
0.2
0.0
F_ST lagoon - sea (Mediterranean Sea)
F_ST North – South (Atlantic Ocean)
C-
Rfgr
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
0 1 2 3 4 5 6 7 8
Position along LG15 (Mb)
D-
Mean R_fgr
0.00040
0.00030
0.00020
0.00010
5 10 15 20 25
Chromosome length (Mb)

Figure 5 Outlier cross-mapping (A-), F_{ST} between the Mediterranean lagoon and Sea lineages (in black), and between the North and South Atlantic lineages (in grey) along the unique chromosome (B-), the number of inferred recombination events (R_{fgr}) alongside LG15, *i.e.* the chromosome carrying the genomic island of differentiation (C-), and the ratio of the total number of inferred recombination events by contig length averaged per chromosome (R_{fgr}; x-axis), used as a proxy of the population recombination rate, plotted against chromosome length (y-axis; D-). As *H. comes* scaffolds are unplaced, *H. guttulatus* SNPs were first blasted on *H. comes* scaffolds, then each *H. comes* scaffold was mapped to *S. scovelli* genome. A putative *H. comes* chromosome was hence reconstructed (A-). The order of the scaffolds and SNPs according to the blasts was conserved. Outlier SNPs displaying parallel differentiation between Atlantic lineages and Mediterranean ecotypes are colored in orange, while the grey outlier showed high differentiation between the Mediterranean Thau lagoon and other sites. The chromosome of the genomic island of differentiation (blasted against S. scovelli LG15) is colored in orange in panel D-.

Genetic parallelism was visualized by plotting F_{ST} of Mediterranean lagoon and sea locations (x-axis in Fig. 6) against F_{ST} between North and South Atlantic clusters (y-axis in Fig. 6). Outliers showing genetic parallelism appeared in the top right of Figure 6, showing elevated genetic differentiation between Mediterranean lagoon and marine sites (x-axis) on the one hand, and between North and South Atlantic sites (y-axis) on the other hand.

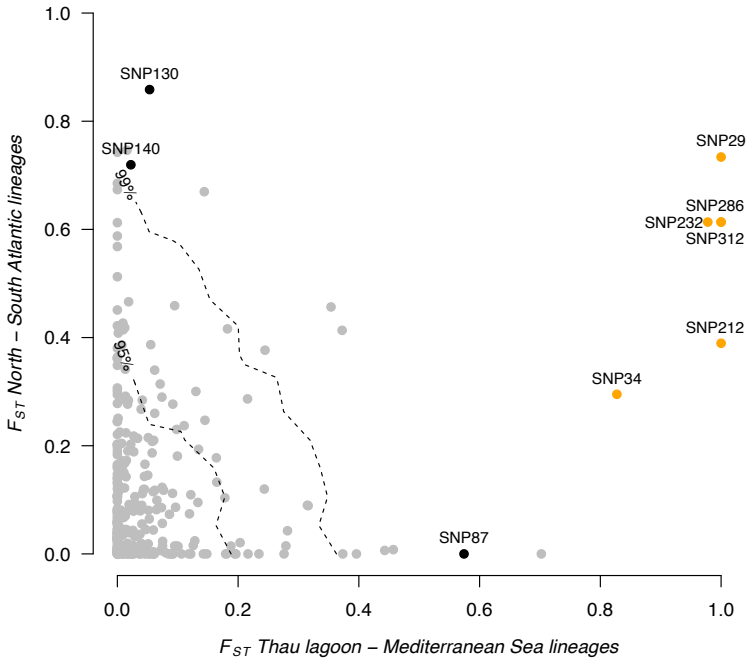


Figure 6 Genome scan of infra-specific differentiation in *H. guttulatus*. Dashed lines represent the 95% and 99% quantiles of the neutral envelope of F_{ST} obtained following Fraïsse et al. (2014). Loci identified by all methods as outliers are colored in orange – the six outliers that displayed parallel differentiation between Atlantic lineages and Mediterranean ecotypes – and in black – the three other outliers.

Three other outliers were also consistently identified (SNPs 87, 130 and 140; Fig. 4 and 6) using all four methods. They distinguished either the Mediterranean Thau lagoon (*i.e.* high genetic differentiation along the x-axis in Fig. 6; SNP 87), the North Atlantic sites or the South Atlantic sites (*i.e.* high genetic differentiation along the y-axis in Fig. 6; SNPs 130 and 140) from all other sites. By using a similar approach for convergent outliers, we only observed SNP 87 that consistently mapped in the unique (and putative) chromosome.

Of the 21,416 contigs from Galtier et al. (2018), we obtained conjointly the position along *S. scovelli* genome and the ratio of total number of inferred recombination events

by contig length (R_{fgr}) for 2,112 contigs. A sliding window analysis of R_{fgr} along LG15 (the genomic island) did not reveal a strong heterogeneity of the population recombination rate within this chromosome (Fig. C). Mean R_{fgr} and the chromosome length were significantly correlated (Spearman's rank correlation ρ= -0.62, *p*-value = 0.003; Fig. 5D) as expected. LG15 appeared as an outlier with a lower recombination rate than expected from its size, and the correlation with chromosome length improved when LG15 was excluded (Spearman's rank correlation ρ= -0.70, *p*-value = 0.0006; Fig. 5D), suggesting a lower population recombination rate within the chromosome with the genomic island of differentiation (*i.e.* the one with outliers blasting against *S. scovelli* LG15).

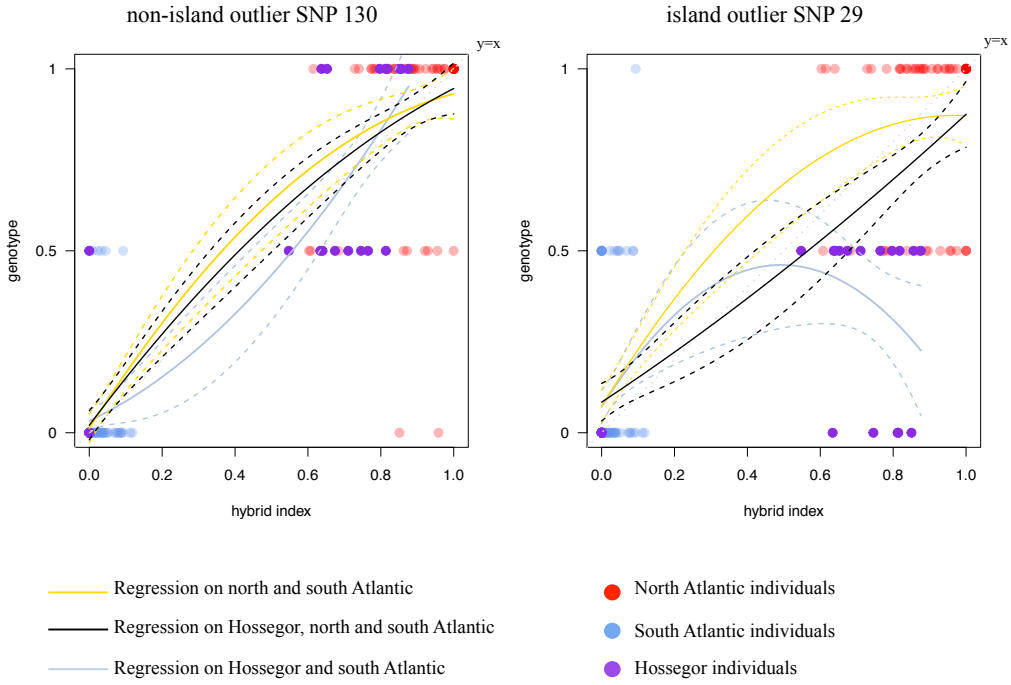


Figure 7 Genomic cline plots for two illustrative markers showing extreme level of differentiation (SNPs 29 and 130). Genomic clines were conducted on North and South Atlantic lineages following polynomial regressions on all Atlantic individuals without Hossegor (in yellow), on all Atlantic individuals (in black) and on all Atlantic individuals, without North Atlantic (in blue-grey). 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).

Figure 7 shows the genomic cline analysis obtained with SNP 29, an outlier that clustered on the unique chromosome (Fig. 5), and SNP 130, an outlier between the northern and the southern Atlantic lineage but that mapped to another chromosome and was not differentiated in the Mediterranean Sea (no genetic parallelism). Regressions were found to be different at SNP 29 and other outliers that clustered on the same chromosome, when the Hossegor sample was included in the analysis or not (Fig. 7), while regressions were always close to the diagonal with SNP 130 and other outliers that mapped to other chromosomes. This analysis reveals a clear discordance between genomic island loci and the rest of the genome in the Hossegor sample. The allele frequency of genomic island loci is found similar in North Atlantic-like seahorses of Hossegor and South Atlantic populations (high frequency of South Atlantic alleles).

Note that loci with extreme level of differentiation could not be related to any sex differences. Of all the 292 fish genotyped, 168 were sexed, a sampling that comprised individuals from the five lineages as well as the hybrid zones, with a balanced sex ratio within each site. In addition, the analysis of gene ontology terms for outlier loci did not reveal any significant functional enrichment.

Discussion

Genetic analyses of the long-snouted seahorse revealed cryptic discrete panmictic genetic lineages that meet either in a narrow contact zone in the Atlantic, or display a mosaic distribution associated with environmental variation in the Mediterranean Sea. Despite limited dispersal abilities and seemingly small population sizes (but see Curtis

584 and Vincent 2006), each lineage showed remarkable genetic homogeneity over very large
585 distances, with genetic panmixia observed within each lineage. This spatial structure,
586 with strong and sharp genetic subdivisions, is not expected if random genetic drift was
587 predominantly responsible for the genetic differentiation between these populations.

588 The spatial organization of the different genetic backgrounds proved to be an
589 unusual combination of geographic subdivision in the Atlantic Ocean, and genetic
590 structure related to the sea-lagoon ecological contrast in the Mediterranean Sea. We find
591 that these two subdivisions partly relied on the same genetic architecture. We observed
592 genetic parallelism at some markers showing extreme levels of differentiation between
593 habitats in the Mediterranean Sea, but also between geographic lineages in the Atlantic
594 Ocean. Intriguingly, all the loci showing convergent allele frequency patterns mapped to
595 a unique chromosome. Genomic divergence is highly variable along a genome, and the
596 use of low-density genome scans was here efficient to characterize this single but large
597 genomic region.

598 We suggest the existence of a shared evolutionary history between Atlantic
599 parapatric lineages and Mediterranean ecotypes, with the Mediterranean lagoon ecotype
600 anciently related to the North-Atlantic lineage. The underlying reproductive isolation
601 mechanisms may involve a combination of intrinsic and extrinsic genetic barriers, where
602 relative contributions may differ between the two contexts.

603

604 *1- Genome scans in hybrid zones*

605 Our SNP panel allowed us to define five discrete panmictic genetic clusters, two
606 in the Atlantic Ocean, two in the Mediterranean Sea, and one in the Black Sea, which

607 cannot be morphologically distinguished with reliability so far (*i.e.* cryptic lineages). The
608 average genetic differentiation between clusters and its associated variance were strong
609 ($0.09 \pm 0.02 < F_{ST} \pm \text{sd} < 0.26 \pm 0.04$). In this context, identifying outlier loci is a
610 complex task with problems of false positives (Lotterhos and Whitlock 2014, 2015). It is
611 also increasingly being acknowledged that a signal of local adaptation should not be
612 easily captured without a broad sampling of the genome in a standard infra-specific low-
613 linkage disequilibrium context (Hoban et al. 2016). The discovery of nine well-supported
614 outlier loci in a fairly small SNP dataset suggests strong variance in differentiation levels
615 associated with the existence of cryptic genetic barriers involving many selected loci
616 (Bierne et al. 2011). Extensive linkage disequilibrium is also maintained in this complex
617 of semi-isolated genetic backgrounds when compared to a standard infra-specific context.
618 However, this is not expected to be a rare situation as genomic studies have provided
619 accumulating evidence that semi-permeable barriers to gene flow are widespread (Roux
620 et al. 2016) and affect a substantial proportion of genomes (Harrison and Larson 2016).
621 In this study, six of the nine outliers showed a signature of genetic parallelism
622 (Fig. 2, Fig. SI4). These six outliers not only proved to be the most differentiated loci
623 using a pairwise comparison between Northern and Southern lineages in the Atlantic
624 Ocean and between the lagoon and marine ecotypes in the Mediterranean Sea (Fig. 6),
625 but they also displayed a genetic structure that is not compatible with the sample
626 covariance matrix inferred with the full set of loci (FLK test, Supporting Information SI6)
627 or with simulations under the best-supported demographic model inferred from
628 transcriptome data (Fig. 6). Finally, these six outliers (3.15% of the SNPs analyzed) in
629 high linkage disequilibrium (Fig. 4) proved to map to a unique chromosome (4.5% of the

630 genome) in available fish genomes, although the order on the chromosome was not so
631 well conserved in distantly related species (SI5). This result suggests the existence of a
632 large genomic island of differentiation as already reported in other fishes (e.g.
633 sticklebacks: Jones et al. 2012, cod: Hemmer-Hansen et al. 2013, Berg et al. 2017,
634 seabass: Duranton et al. 2018). Reduced recombination, a chromosomal inversion or
635 another form of recombination suppression would likely be involved (Gagnaire et al.
636 2018, Martin et al. 2018, Roesti et al. 2013). Such clusters of divergent loci are more
637 likely to form through genomic rearrangements bringing coadapted loci close together
638 (Yeaman and Whitlock 2011, Yeaman 2013). Our analysis of recombination rates indeed
639 revealed that the population recombination rate is reduced in the chromosome that bears
640 the genomic island (Fig. 5D). However, the enhanced structure of this chromosome can
641 partly explain this observation. Indeed, this chromosome is more structured between
642 populations than the others; a reduced population recombination rate ($N_e r$) is expected
643 even if the recombination rate (r) is not. In addition, recombination could be disrupted
644 only between inversions or other structural variations while being efficient within them.
645 This chromosome-wide island of differentiation also explains why we detected the signal
646 of parallel outlying differentiation with a moderate number of loci analyzed. The
647 genomic island could have been easily missed with a few microsatellites but our SNP
648 panel had enough coverage that it contained a few SNPs on LG15. These preliminary
649 results call for genome sequence analysis and genetic mapping in order to better
650 characterize the genomic island and its structural variation, and to investigate the genome
651 localization of additional barrier loci with a smaller chromosomal footprint.
652

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

In the Atlantic Ocean we confirm the existence of two well-differentiated genetic clusters previously identified by Woodall et al. (2015) with five microsatellites and two mitochondrial genes. A genetic homogeneity was observed over large distances within each lineage (Fig. 1), contrasting with a very strong and abrupt genetic differentiation between them. Inter-lineage genetic divergence was not only captured by the most differentiated SNP that were nearly fixed between the two lineages (SNP 130 in Fig. 4, F_{ST} = 0.9), but also, by many other SNPs (SI4), suggesting isolating mechanisms and loci with genome-wide effects. This mechanism is also suggested by the discovery of the two parental lineages coexisting in the same lagoon in South West of France (Hossegor lagoon, site 6 in Fig. 1). Such zone of co-existence in sympatry contrasting with genetic homogeneity at a large spatial scale within each lineage suggests reproductive isolation (Jiggins and Mallet 2000), involving either pre-zygotic isolation or selection against hybrids at early life stages.

Alternatively, we could also have sampled the first-generation migrants without any genomic modifications. However, recent migration of the two diverged genetic backgrounds in this lagoon cannot alone explain the co-existence of the two lineages; a mechanism of reproductive isolation should maintain divergence between the two lineages. Indeed, hybridization has occurred to some extent as suggested by local introgression of the North-Atlantic seahorse in this lagoon (Fig. 3). The decreasing proportion of south lineage ancestry from Hossegor to the English Channel also provides indirect evidence for asymmetric introgression (Fig. 2). In addition, further support for partial reproductive isolation is evidenced by the genomic island outlier loci pattern in

676 Hossegor (site 6). The North Atlantic lineage mostly carries the southern alleles at
677 genomic island loci (pale pink dots in Fig. 4, see also SI4). To produce such North
678 Atlantic-type individuals with a South-Atlantic island, recombination between parental
679 backgrounds is required, suggesting repeated hybridization over several generations.

680 In Barton's concordance analysis (Fig. 7) the geographic information was
681 included to contrast the results obtained with or without the Hossegor sample. The
682 objective of this analysis was to infer the existence of a local discordance in this site. The
683 swamping of North Atlantic fishes by the southern allele at the genomic island suggests
684 epistatic or coupling interactions among loci implied in reproductive isolation.
685 Endogenous post-zygotic selection against hybrids does not result in a stable
686 polymorphism in a single isolated population, but instead in transient polymorphisms.
687 Also known as bistable variants, the maintenance of underdominant or epistatically
688 interacting genetic incompatibilities requires a migration-selection balance in a spatially
689 subdivided population (Barton and Turelli 2011), or frequency-dependent selection
690 (Barton and de Cara 2009). The system may have fixed one state of a bistable variant in
691 the Hossegor lagoon. Alternatively, the southern allele may confer a selective advantage
692 in this lagoon whatever the genetic background. However the latter hypothesis does not
693 seem coherent in Hossegor as the northern allele is always in higher frequency in lagoon-
694 like habitats elsewhere, though we may have missed more subtle ecological parameters
695 that would have driven this pattern.

696 In the southern Atlantic, outlier alleles tend to follow marine/lagoon
697 environments, with the northern allele frequencies in lagoon populations slightly higher
698 than in the marine populations (Fig. 4). Local adaptation genes could be embedded in the

699 island and escape coupling with intrinsic isolation through rare recombination events.
700 These alleles might also come from Mediterranean lagoons by introgression through the
701 Atlantic-Mediterranean divide. Alternatively, as northern alleles enter the southern
702 background through a lagoon, they might be better propagated through lagoons thanks to
703 habitat resistance created by local adaptation at unscored loci.

704 Overall, the Atlantic contact zone possesses all the characteristics of standard
705 clinal hybrid zones that follow the tension zone model (Barton and Hewitt 1985), *i.e.* a
706 secondary contact zone maintained by a balance between migration and intrinsic
707 reproductive isolation. Exogenous selection may also contribute, although the two
708 Atlantic lineages both inhabit indifferently lagoon and sea habitats (Fig. 1), and co-exist
709 in syntopy in Hossegor, suggesting a limited role of ecological contrast in their genetics.
710 Hossegor is a small lagoon, only 2.3 km long, 300 m wide and no more than 2 m deep,
711 which strongly limits the opportunity for microparapatry. In any case, only strong
712 intrinsic reproductive isolation can guarantee a genome-wide barrier to gene flow
713 explaining the co-existence of the two lineages in the Hossegor lagoon.

714

715 *3- Sea and lagoon ecotypes in the Mediterranean Sea*

716 Contrasting with the Atlantic hybrid zone, the two cryptic lineages identified in
717 the Mediterranean Sea were associated with lagoon/sea ecosystem variation. Our broad
718 genomic and spatial sampling revealed two crucial observations. First, while the marine
719 lineage was surprisingly homogeneous over the whole Mediterranean Sea, from Greece
720 to Spain (F_{IT} = 0.0078 n.s.), lagoon-like samples, especially the Thau lagoon, showed a
721 strong and genome-wide genetic differentiation from them. Samples from two lagoons

722 (Thau in France and Bizerte in Tunisia, sites 16 and 20 in Fig. 1) were sufficient to
723 reveal an association with the environment that was previously unseen, the Thau lagoon
724 being the only sample from Western Mediterranean basin in Woodall et al. (2015).
725 Second, fixed differences between lagoon and marine samples were observed, although
726 they were sampled only few kilometers apart (e.g. sites 14-16 in Fig. 2, 4). A single but
727 important seahorse sampled on the seaside of the Thau lagoon (site 15 in Fig. 1), plus
728 seven others sampled on the seaside of another lagoon of the region (site 14 in Fig. 1)
729 proved to belong to the marine genetic cluster, without any sign of introgression.
730 Likewise, no evidence of introgression was observed in the Thau sample (Fig. 3). Once
731 again, despite genetic homogeneity over large area, such strong and abrupt genetic
732 differentiation suggests partial reproductive isolation between these two lineages. In this
733 case there are obvious ecological drivers, *i.e.* habitat specialization, so that the entire
734 Mediterranean Sea could be viewed as a mosaic hybrid zone, with one parental form
735 (defined by outlier SNPs under the genic view of species delineation, Wu 2001)
736 inhabiting lagoons in Thau (site 16), Bizerte (site 20) and Halkida (site 24), and another
737 parental form inhabiting the sea. Indeed, the Bizerte lagoon (site 20 in Fig. 1), which is
738 ecologically similar to the Thau lagoon (Sakka Hlaili et al. 2008), has a population
739 genetically similar to marine Mediterranean samples at most loci but share the genetic
740 composition of the Thau lagoon at the genomic island loci (Fig. 4 and SI4). In addition, a
741 subsample of the Halkida population (Greece, site 24B in Fig. 4) was composed of five
742 individuals heterozygous at the genomic island. The environmental parameters at this
743 location are hypothesized to be more lagoon-like, being a secluded bay beyond the
744 northern end of the Euipus Strait. This sample only provides evidence that the genomic

745 island polymorphism and the mosaic spatial structure extends to the eastern basin
746 without really providing further clues about the role of the environment. More
747 Mediterranean lagoons and estuaries will need to be sampled in the future to better
748 characterize the association of these two ecotypes with environmental variation.

749 Our results in the Mediterranean Sea (*i.e.* lagoon/marine system) resemble those
750 obtained in the emblematic three-spined sticklebacks marine/freshwater system (Jones et
751 al. 2012), or more recently in the marine-migratory/freshwater-resident lampreys
752 (Rougemont et al. 2016) and coastal/marine ecotypes in European anchovies (Le Moan
753 et al. 2016). When a shared divergent genomic island is observed –*i.e.* genetic
754 parallelism, as here in long-snouted seahorse among the Mediterranean lagoons, or in the
755 examples cited above, there could be three possible interpretations: (i) parallel gene
756 reuse from a shared ancestral polymorphism present in the marine supposedly ancestral
757 population (Jones et al. 2012), (ii) the spread of a locally adapted allele (*i.e.* the
758 'transporter hypothesis'; Schluter and Conte 2009) or (iii) secondary contact followed by
759 spatial reassortment of the divergent lineages and extensive introgression swamping such
760 that only selected loci and their chromosomal neighborhood retain the history of
761 adaptation (Bierne et al. 2013). The three scenarios are difficult to discriminate as they
762 converge toward a similar pattern (Johannesson et al. 2010, Bierne et al. 2013, Welch
763 and Jiggins 2014). Here, as for the lampreys (Rougemont et al. 2016), the Thau lagoon
764 provides a possible support for the secondary contact model because the differentiation,
765 although stronger at the genomic island, is genome-wide. The Bizerte lagoon however
766 can either be interpreted as a marine lineage introgressed by the lagoon allele at the
767 genomic island, or as a lagoon lineage (defined by adaptive/speciation genes) massively

768 introgressed by neutral marine alleles. Incorporating heterogeneous migration rates in
769 demographic inference methods allowed Le Moan et al. (2016), Rougeux et al. (2016)
770 and Rougemont et al. (2016) to identify the signal of a secondary contact history carried
771 by islands of differentiation in lampreys, white fishes and anchovies. Unfortunately our
772 286-SNPs dataset does not allow performing such historical demographic reconstruction.
773 Nonetheless the TREEMIX analysis reveals that episodes of secondary admixtures
774 strongly improve the fit to the sample covariance matrix, but adaptive introgression or
775 massive introgression swamping can both explain them. Anyhow, demographic
776 reconstruction does not completely refute the 'transporter hypothesis' which stipulates
777 lagoon alleles spread from lagoon to lagoon (or freshwater allele from river to river) and
778 is a scenario that produces a very similar genomic pattern of differentiation to the one
779 produced by a standard secondary contact (Bierne et al. 2013, Rougemont et al. 2016). In
780 the case of the seahorse complex, however, we made the new observation that genetic
781 parallelism is observed with the Atlantic populations where the structure is geographic
782 and independent of the lagoon-sea habitats, which offers a new twist to the debate with
783 complementary arguments.
784
785 *4- Genetic parallelism in two different spatial/ecological contexts*
786 The most astonishing result of our genetic analysis was genetic parallelism
787 between the Mediterranean lagoon ecotype and the north Atlantic lineage at a large
788 genomic island. Parallel evolution is thought to imply distinct but repetitive ecological
789 characteristics (e.g. Butlin et al. 2014). In the present study, we found that the genomic
790 island was associated with the sea-lagoon ecological contrast in the Mediterranean Sea,

791 while there was no such genetic differentiation between lagoon and sea samples in the
792 Atlantic Ocean. The North and South lineages inhabit indifferently lagoons and seas, so
793 that what seemed obvious in the Mediterranean Sea regarding the divergence of the two
794 lineages, *i.e.* habitat specialization, was not observed along the Atlantic Ocean where the
795 divergence seems uncorrelated to the ecological contrast that explains the two
796 Mediterranean ecotypes. No analogy was also observed regarding abiotic parameters,
797 such as temperature or salinity, between the North Atlantic and Mediterranean lagoon
798 lineages, making hard to correlate these two lineages to ecological drivers too. Although
799 we may have missed putative ecological drivers of such genetic parallelism, parallel gene
800 reuse driven by ecological convergence seems here unlikely. A shared history of
801 divergence retained at outlier loci in the North Atlantic and Thau lineages would
802 nonetheless be compatible with isolation after postglacial warming. Indeed, while moving
803 lengthwise at each glacial cycle, species can be trapped in Mediterranean pockets of cold
804 waters due to the particular geography of the Mediterranean Sea, *i.e.* perpendicular to
805 north-south population displacements (e.g. Borsa et al. 1997, Debes et al. 2008). Isolated
806 populations of Atlantic-derived lineages trapped within the Mediterranean Sea could have
807 adapted to new environments such as lagoons or the Black Sea. Shared variations with
808 the North-Atlantic lineage would only be visible at regions of the genome protected from
809 gene flow by local selection and reproductive isolation.

810 Shared ancestral polymorphism sieved by adaptation in a patchy environment
811 (Bierne et al. 2013) and incipient speciation (Guerrero and Hahn 2017) would then
812 explain our data. The association with habitat in the Mediterranean Sea and with space in
813 the Atlantic Ocean could be explained by a secondary evolution of locally adapted genes

814 within the genomic island in the Mediterranean Sea, benefiting from the barrier to gene
815 flow imposed by intrinsic selection (divergence hitchhiking; Via 2012). Alternatively, the
816 genomic island could have coupled with local adaptation polymorphisms localized
817 elsewhere in the genome in the Mediterranean Sea (Bierne et al. 2011), while it would
818 have been trapped by a barrier to dispersal in the Atlantic Ocean (Barton 1979, Barton
819 and Hewitt 1985). Without further data and the true genomic position of loci in the
820 seahorse genome, it is difficult to disentangle the two hypotheses.

821 One locus (SNP 87) localized in the same chromosome as parallel SNPs (Fig. 5)
822 is differentiated between the marine and lagoon ecotypes in the Mediterranean Sea, but is
823 not differentiated between the northern and southern lineages in the Atlantic Ocean. At
824 first sight, it could be interpreted as evidence for a possible secondary local sweep in
825 Mediterranean lagoons. However, according to the gene order inferred from the closest
826 species (Gulf pipefish and Tiger tail seahorse; Fig. 5A) along with a barrier to gene flow
827 less effective in the Atlantic (see SNPs 34 and 212 in Fig. 4 and 5B), it could also be
828 interpreted as being localized in the island "shoulders" (*i.e.* loci in the vicinity of the
829 regions harboring local adaptation and/or reproductive isolation loci; Gagnaire et al.
830 2015, Le Moan et al. 2016) in which a stronger introgression rate has erased the
831 differentiation faster in the Atlantic than in the Mediterranean, in favor of the alternative
832 interpretation. This would mean that recombination disruption, provided it exists, would
833 not be as strong at the end of the chromosome arm. Importantly, whatever the explanation
834 -divergence hitchhiking or coupling- it requires invoking interaction between intrinsic
835 and ecological selection and not ecological selection alone (Bierne et al. 2011, Kulmuni
836 and Westram 2017). Furthermore, intrinsic isolation has most probably evolved first in

837 this system as no genetic parallelism was observed in outliers discriminating lagoon to
838 marine ecotypes, which would contradict the predominant view that ecological selection
839 is necessarily the initial catalyzer of the chain of accumulation of barriers in ecological
840 speciation.
841
842 **Conclusions**
843 Analyzing the population genetics of the long-snouted seahorse *Hippocampus guttulatus*
844 revealed a complex of panmictic genetic backgrounds subdivided by sharp semi-
845 permeable hybrid zones. This is now standard observation in marine species (Knowlton
846 1993, Pante et al. 2015, Sheets et al. 2018) where morphological stasis might be more
847 widespread than in terrestrial species. The subdivision of species by hybrid zones is a
848 long-lasting observation in the terrestrial realm (Hewitt 1989) but arguably more readily
849 detected by morphological differences. This is nonetheless important knowledge to
850 further inform captive breeding and *in-situ* management decisions. We also easily found
851 outlier loci despite a moderate number of loci analyzed, and the clustering of these outlier
852 loci in a single genomic region that showed depleted population recombination rates,
853 which, henceforth, seems a standard observation of the recent hybrid zone literature (e.g.
854 in stickelbacks, Jones et al. 2012; jaera isopods, Ribardière et al. 2017; cods, Hemmer-
855 Hansen et al. 2013; littorina snails, Westram et al. 2018; saltmarsh beetles, Van
856 Belleghem et al. 2018).
857 However, we also made two additional observations that are less commonly reported and
858 deserve a broader interest outside the study of seahorse themselves. First, we found the
859 two, usually opposed, standard spatial structures of the hybrid zone literature, namely
860 clinal and mosaic hybrid zones, in the same study system. This result calls for further

investigations with lab and fieldwork in order to better understand the mechanisms of reproductive isolation at play and their genetic architecture. Secondly, we found a parallel pattern of differentiation at the genomic island in the two spatial/ecological contexts. Although this result will also need to be substantiated by follow-up genomic studies, it nonetheless reveals that the hallmark of ecologically driven adaptive divergence can be observed in absence of obvious ecological convergence. We argue that alternative scenarios involving secondary introgression swamping and intrinsic isolation should be more seriously considered as valid alternatives and the seahorse complex could become an interesting flagship system in the debate.

Author contributions: F.R. analyzed the data and wrote the article. C.L.-H. performed molecular experiments. L.W. performed sampling of most fishes. C.B., P.L., B.H., F.O.-F., P.A., V.B., O.B., T.E.-A., and S.H. also contributed in sampling. K.B. provided accurate computational solutions for bioinformatics analyses. S.A.-H. and P.-A.G. wrote the article. N.B. designed the work, analyzed data and wrote the article.

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, Laurent Duret and Christelle Fraïsse for computational advice. 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. We are also grateful to Prof Noor, Dr Flaxman and the reviewers for very helpful comments.

Data archival location: SNPs data have been deposited at DRYAD. DOI: 10.5061/dryad.mq122fv.

Literature

Barton, N. 1979. 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.

Berg, P. R., B. Star, C. Pampoulie, I. R. Bradbury, P. Bentzen, J. A. Hutchings, S. Jentoft, and K. S. Jakobsen. 2017. Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. Heredity, doi: 10.1038/hdy.2017.54.

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., Gagnaire, P.-A., and P. David. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. Curr Zool, 59: 72–86.

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.

Belleghem, S. M. V., C. Vangestel, K. D. Wolf, Z. D. Corte, M. Möst, P. Rastas, L. D. Meester, and F. Hendrickx. 2018. Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. PLOS Genet. 14:e1007796.

Bonhomme, M., C. Chevalet, B. Servin, S. Boitard, J. Abdallah, S. Blott, and M. SanCristobal. 2010. Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics 186:241–262.

Borsa, P., Blanquer, A., & Berrebi, P. 1997. Genetic structure of the flounders *Platichthys flesus* and *P. stellatus* at different geographic scales. Mar Biol, 129:233-246.

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.

945 Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y.
946 Ng, Z. W. Lim, E. Bezault, J. Turner-Maier, J. Johnson, R. Alcazar, H. J. Noh, P. Russell,
947 B. Aken, J. Alföldi, C. Amemiya, N. Azzouzi, J.-F. Baroiller, F. Barloy-Hubler, A.
948 Berlin, R. Bloomquist, K. L. Carleton, M. A. Conte, H. D'Cotta, O. Eshel, L. Gaffney, F.
949 Galibert, H. F. Gante, S. Gnerre, L. Greuter, R. Guyon, N. S. Haddad, W. Haerty, R. M.
950 Harris, H. A. Hofmann, T. Hourlier, G. Hulata, D. B. Jaffe, M. Lara, A. P. Lee, I.
951 MacCallum, S. Mwaiko, M. Nikaido, H. Nishihara, C. Ozouf-Costaz, D. J. Penman, D.
952 Przybylski, M. Rakotomanga, S. C. P. Renn, F. J. Ribeiro, M. Ron, W. Salzburger, L.
953 Sanchez-Pulido, M. E. Santos, S. Searle, T. Sharpe, R. Swofford, F. J. Tan, L. Williams,
954 S. Young, S. Yin, N. Okada, T. D. Kocher, E. A. Miska, E. S. Lander, B. Venkatesh, R.
955 D. Fernald, A. Meyer, C. P. Ponting, J. T. Streelman, K. Lindblad-Toh, O. Seehausen,
956 and F. Di Palma. 2014. The genomic substrate for adaptive radiation in African cichlid
957 fish. Nature 513:375–381.
958
959 Britton-Davidian, J., F. Fel-Clair, J. Lopez, P. Alibert, and P. Boursot. 2005. Postzygotic
960 isolation between the two European subspecies of the house mouse: estimates from
961 fertility patterns in wild and laboratory-bred hybrids. Biol. J. Linn. Soc. 84:379–393.
962
963 Butlin, R. K., M. Saura, G. Charrier, B. Jackson, C. André, A. Caballero, J. A. Coyne, J.
964 Galindo, J. W. Grahame, J. Hollander, P. Kemppainen, M. Martínez-Fernández, M.
965 Panova, H. Quesada, K. Johannesson, and E. Rolán-Alvarez. 2014. Parallel evolution of
966 local adaptation and reproductive isolation in the face of gene flow. Evolution 68:935–
967 949.
968
969 Christe, C., K. N. Stölting, L. Bresadola, B. Fussi, B. Heinze, D. Wegmann, and C.
970 Lexer. 2016. Selection against recombinant hybrids maintains reproductive isolation in
971 hybridizing *Populus* species despite F1 fertility and recurrent gene flow. Mol. Ecol.
972 25:2482-2498.
973
974 Conesa, A., and S. Götz. 2008. Blast2GO: A comprehensive suite for functional analysis
975 in plant genomics. Int. J. Plant Genomics 2008.
976
977 Curtis, J. M. R., and A. C. J. Vincent. 2006. Life history of an unusual marine fish:
978 survival, growth and movement patterns of *Hippocampus guttulatus* Cuvier 1829. J. Fish
979 Biol. 68:707–733.
980
981 Dalziel, A. C., T. H. Vines, and P. M. Schulte. 2012. Reductions in prolonged swimming
982 capacity following freshwater colonization in multiple threespine
983 stickleback populations. Evolution 66:1226–1239.
984
985 Debes, P., Zachos, F., and R. Hanel. 2008. Mitochondrial phylogeography of the
986 European sprat (Sprattus sprattus L., Clupeidae) reveals isolated climatically vulnerable
987 populations in the Mediterranean Sea and range expansion in the northeast Atlantic. Mol.
988 Ecol. 17 :3873–3888.
989

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.

Duranton, M., F. Allal, C. Fraïsse, N. Bierne, F. Bonhomme, and P.-A. Gagnaire. 2018. The origin and remolding of genomic islands of differentiation in the European sea bass. Nat. Commun. 9:2518.

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.

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.

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.

Fraïsse, C., C. Roux, J. J. Welch, and N. Bierne. 2014. Gene-flow in a mosaic hybrid zone: is local introgression adaptive? Genetics 197:939–951.

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.

Gagnaire, P.-A., J.-B. Lamy, F. Cornette, S. Heurtebise, L. Dégremont, E. Flahauw, P. Boudry, N. Bierne, and S. Lapègue. 2018. Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters Crassostrea gigas and Crassostrea angulata. Genome Biol. Evol. 10:2518–2534.

Galtier, N., C. Roux, M. Rousselle, J. Romiguier, E. Figuet, S. Glémin, N. Bierne, and L. Duret. 2018. Codon Usage Bias in Animals: Disentangling the Effects of Natural Selection, Effective Population Size, and GC-Biased Gene Conversion. Mol. Biol. Evol. 35:1092–1103.

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.

1035 Bernard, and N. Galtier. 2013. Reference-free population genomics from next-generation
1036 transcriptome data and the vertebrate–invertebrate gap. PLOS Genet. 9:e1003457.
1037
1038 Gompert, Z., and C. Alex Buerkle. 2010. INTROGRESS: a software package for
1039 mapping components of isolation in hybrids. Mol. Ecol. Resour. 10:378–384.
1040
1041 Good, J. M., M. A. Handel, and M. W. Nachman. 2008. Asymmetry and polymorphism
1042 of hybrid male sterility during the early stages of speciation in house mice. Evol. Int. J.
1043 Org. Evol. 62:50–65.
1044
1045 Guerrero, R., and M. W. Hahn. 2017. Speciation as a sieve for ancestral polymorphism.
1046 bioRxiv 155176.
1047
1048 Harrison, R. G. 1993. Hybrid zones and the evolutionary process. Oxford University
1049 Press.
1050
1051 Harrison, R. G., and E. L. Larson. 2016. Heterogeneous genome divergence, differential
1052 introgression, and the origin and structure of hybrid zones. Mol. Ecol. 25:2454:2466.
1053
1054 Hedrick, P. W. 2013. Adaptive introgression in animals: examples and comparison to
1055 new mutation and standing variation as sources of adaptive variation. Mol. Ecol.
1056 22:4606–4618.
1057
1058 Hemmer-Hansen, J., E. E. Nielsen, N. O. Therkildsen, M. I. Taylor, R. Ogden, A. J.
1059 Geffen, D. Bekkevold, S. Helyar, C. Pampoulie, T. Johansen, FishPopTrace Consortium,
1060 and G. R. Carvalho. 2013. A genomic island linked to ecotype divergence in Atlantic
1061 cod. Mol. Ecol. 22:2653–2667.
1062
1063 Hewitt, G. M. 1988. Hybrid zones-natural laboratories for evolutionary studies. Trends
1064 Ecol. Evol. 3:158–167.
1065
1066 Hewitt, G. M. 1989. The subdivision of species by hybrid zones. *Speciation and its*
1067 *Consequences*, 85-110.
1068
1069 Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L.
1070 Poss, L. K. Reed, A. Storfer, and M. C. Whitlock. 2016. Finding the genomic basis of
1071 local adaptation: pitfalls, practical solutions, and future directions. Am. Nat. 188:379–
1072 397.
1073
1074 Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E.
1075 Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C.
1076 Churcher, C. Scott, J. C. Barrett, R. Koch, G.-J. Rauch, S. White, W. Chow, B. Kilian, L.
1077 T. Quintais, J. A. Guerra-Assunção, Y. Zhou, Y. Gu, J. Yen, J.-H. Vogel, T. Eyre, S.
1078 Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd,
1079 E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M.
1080 Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis,

C. Clee, K. Oliver, R. Clark, C. Riddle, D. Elliot, D. Eliott, G. Threadgold, G. Harden, D. Ware, S. Begum, B. Mortimore, B. Mortimer, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Lloyd, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. D. Cooper, J. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C. M. Dooley, Z. Ersan-Ürün, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn, J. Konantz, M. Konantz, M. Oberländer, S. Rudolph-Geiger, M. Teucke, C. Lanz, G. Raddatz, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, S. C. Schuster, N. P. Carter, J. Harrow, Z. Ning, J. Herrero, S. M. J. Searle, A. Enright, R. Geisler, R. H. A. Plasterk, C. Lee, M. Westerfield, P. J. de Jong, L. I. Zon, J. H. Postlethwait, C. Nüsslein-Volhard, T. J. P. Hubbard, H. Roest Crollius, J. Rogers, and D. L. Stemple. 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498–503.

Huang, X., and A. Madan. 1999. CAP3: A DNA Sequence assembly program. Genome Res. 9:868–877.

Hunt, W. G., and Selander, R. K. 1973. Biochemical genetics of hybridisation in European house mice. Heredity, 31, 1133

Jaillon, O., J.-M. Aury, F. Brunet, J.-L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot, S. Nicaud, D. Jaffe, S. Fisher, G. Lutfalla, C. Dossat, B. Segurens, C. Dasilva, M. Salanoubat, M. Levy, N. Boudet, S. Castellano, V. Anthouard, C. Jubin, V. Castelli, M. Katinka, B. Vacherie, C. Biémont, Z. Skalli, L. Cattolico, J. Poulain, V. de Berardinis, C. Cruaud, S. Duprat, P. Brottier, J.-P. Coutanceau, J. Gouzy, G. Parra, G. Lardier, C. Chapple, K. J. McKernan, P. McEwan, S. Bosak, M. Kellis, J.-N. Volff, R. Guigó, M. C. Zody, J. Mesirov, K. Lindblad-Toh, B. Birren, C. Nusbaum, D. Kahn, M. Robinson-Rechavi, V. Laudet, V. Schachter, F. Quétier, W. Saurin, C. Scarpelli, P. Wincker, E. S. Lander, J. Weissenbach, and H. Roest Crollius. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946–957.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinforma. Oxf. Engl. 23:1801–1806.

Jiggins, C. D., and J. Mallet. 2000. Bimodal hybrid zones and speciation. Trends Ecol. Evol. 15:250–255.

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.

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.

Kirkpatrick, M., and V. Ravigné. 2002. Speciation by natural and sexual selection: models and experiments. Am. Nat. 159 Suppl 3:S22–35.

Knowlton, N. 1993. Sibling species in the sea. Annu. Rev. Ecol. Syst. 24:189–216.

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.

Larson, E. L., T. A. White, C. L. Ross, and R. G. Harrison. 2014. Gene flow and the maintenance of species boundaries. Mol. Ecol. 23:1668–1678.

Le Moan, A., P.-A. Gagnaire, and F. Bonhomme. 2016. Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. Mol. Ecol. 25:3187-3202.

Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74: 175–195.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 25:1754-60.

Lin, Q., S. Fan, Y. Zhang, M. Xu, H. Zhang, Y. Yang, A. P. Lee, J. M. Woltering, V. Ravi, H. M. Gunter, W. Luo, Z. Gao, Z. W. Lim, G. Qin, R. F. Schneider, X. Wang, P. Xiong, G. Li, K. Wang, J. Min, C. Zhang, Y. Qiu, J. Bai, W. He, C. Bian, X. Zhang, D. Shan, H. Qu, Y. Sun, Q. Gao, L. Huang, Q. Shi, A. Meyer, and B. Venkatesh. 2016. The seahorse genome and the evolution of its specialized morphology. Nature 540:395–399.

López, A., M. Vera, M. Planas, and C. Bouza. 2015. Conservation genetics of threatened *Hippocampus guttulatus* in vulnerable habitats in NW Spain: temporal and spatial stability of wild populations with flexible polygamous mating system in captivity. PLoS ONE 10:e0117538.

Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol. Ecol. 23:2178–2192.

Lotterhos, K. E., and M. C. Whitlock. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol. Ecol. 24:1031–1046.

Lourie, S. A., and A. C. J. Vincent. 2004. Using biogeography to help set priorities in marine conservation. Conserv. Biol. 18:1004–1020.

Luu, K., E. Bazin, and M. G. B. Blum. 2016. PCAdapt: an R package to perform genome scans for selection based on principal component analysis. bioRxiv 056135.

Macholán, M., S. J. E. Baird, P. Dufková, P. Munclinger, B. V. Bímová, and J. Piálek. 2011. Assessing multilocus introgression patterns: a case study on the mouse X chromosome in central Europe. Evol. Int. J. Org. Evol. 65:1428–1446.

Martin, S. H., J. Davey, C. Salazar, and C. Jiggins. 2018. Recombination rate variation shapes barriers to introgression across butterfly genomes. bioRxiv 297531.

Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. Phil Trans R Soc B 367:332–342.

1219 Pante, E., N. Puillandre, A. Viricel, S. Arnaud-Haond, D. Aurelle, M. Castelin, A.
1220 Chenuil, C. Destombe, D. Forcioli, M. Valero, F. Viard, and S. Samadi. 2015. Species are
1221 hypotheses: avoid connectivity assessments based on pillars of sand. Mol. Ecol. 24:525–
1222 544.
1223
1224 Pérez-Figueroa, A., M. J. García-Pereira, M. Saura, E. Rolán-Alvarez, and A. Caballero.
1225 2010. Comparing three different methods to detect selective loci using dominant markers.
1226 J. Evol. Biol. 23:2267–2276.
1227
1228 Pickrell, J. K., and J. K. Pritchard. 2012. Inference of population splits and mixtures from
1229 genome-wide allele frequency data. PLoS Genet 8:e1002967.
1230
1231 Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure
1232 using multilocus genotype data. Genetics 155:945–959.
1233
1234 R Development Core Team (2011) R: A language and environment for statistical
1235 computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-
1236 0, http://www.R-project.org/.
1237
1238 Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone:
1239 mitochondrial, nuclear, and reproductive differentiation of crickets by soil type.
1240 Evolution 43:432–449.
1241
1242 Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M. a. F. Noor,
1243 B. Mehlig, and A. M. Westram. 2017. Interpreting the genomic landscape of speciation: a
1244 road map for finding barriers to gene flow. J. Evol. Biol. 30:1450–1477.
1245
1246 Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population genetics
1247 software for exact tests and ecumenicism. J. Hered. 86:248–249.
1248
1249 Ribardière, A., C. Daguin-Thiébaut, C. Houbin, J. Coudret, C. Broudin, O. Timsit, and T.
1250 Broquet. 2017. Geographically distinct patterns of reproductive isolation and
1251 hybridization in two sympatric species of the *Jaera albifrons* complex (marine isopods).
1252 Ecol. Evol. 7:5352–5365.
1253
1254 Roesti, M., D. Moser and D. Berner. 2013. Recombination in the threespine stickleback
1255 genome-patterns and consequences. Mol Ecol. 22:3014–3027.
1256
1257 Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari,
1258 R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G.
1259 Tsagkogeorga, A. a.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, and N.
1260 Galtier. 2014. Comparative population genomics in animals uncovers the determinants of
1261 genetic diversity. Nature 515:261–263.
1262
1263 Rougemont, Q., P.-A. Gagnaire, C. Perrier, C. Genthon, A.-L. Besnard, S. Launey, and
1264 G. Evanno. 2016. Inferring the demographic history underlying parallel genomic

1265 divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Mol. Ecol.
1266 26:142–162.
1267
1268 Rougeux, C., L. Bernatchez, and P.-A. Gagnaire. 2016. Modeling the multiple facets of
1269 speciation-with-gene-flow towards inferring the divergence history of lake whitefish
1270 species pairs (*Coregonus clupeaformis*). bioRxiv 068932.
1271
1272 Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software
1273 for Windows and Linux. Mol. Ecol. Resour. 8:103–106.
1274
1275 Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016.
1276 Shedding light on the grey zone of speciation along a continuum of genomic divergence.
1277 PLOS Biol. 14:e2000234.
1278
1279 Saarman, N. P., R. Opiro, C. Hyseni, R. Echodu, E. A. Opiyo, K. Dion, T. Johnson, S.
1280 Aksoy, and A. Caccone. 2018. The population genomics of multiple tsetse fly (Glossina
1281 fuscipes fuscipes) admixture zones in Uganda. Mol. Ecol., doi: 10.1111/mec.14957.
1282
1283 Sakka Hlaili, A., B. Grami, N. Niquil, M. Gosselin, D. Hamel, M. Troussellier, and H.
1284 Hadj Mabrouk. 2008. The planktonic food web of the Bizerte lagoon (south-western
1285 Mediterranean) during summer: I. Spatial distribution under different anthropogenic
1286 pressures. Estuar. Coast. Shelf Sci. 78:61–77.
1287
1288 Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad.
1289 Sci. U. S. A. 106 Suppl 1:9955–9962.
1290
1291 Sheets, E. A., P. A. Warner, and S. R. Palumbi. 2018. Accurate population genetic
1292 measurements require cryptic species identification in corals. Coral Reefs 37:549–563.
1293
1294 Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, and İ. Birol. 2009.
1295 ABySS: A parallel assembler for short read sequence data. Genome Res. 19:1117–1123.
1296
1297 Small, C. M., S. Bassham, J. Catchen, A. Amores, A. M. Fuiten, R. S. Brown, A. G.
1298 Jones, and W. A. Cresko. 2016. The genome of the Gulf pipefish enables understanding
1299 of evolutionary innovations. Genome Biol. 17:258.
1300
1301 Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S.
1302 Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and
1303 P. Nosil. 2014. Stick insect genomes reveal natural selection's role in parallel speciation.
1304 Science 344:738–742.
1305
1306 Storey, J. D. 2002. A direct approach to false discovery rates. J. R. Stat. Soc. Ser. B Stat.
1307 Methodol. 64:479–498.
1308

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. Evolution 40:1141–1159.

Teeter, K. C., B. A. Payseur, L. W. Harris, M. A. Bakewell, L. M. Thibodeau, J. E. O'Brien, J. G. Krenz, M. A. Sans-Fuentes, M. W. Nachman, and P. K. Tucker. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. 18:67–76.

Tine, M., H. Kuhl, P.-A. Gagnaire, B. Louro, E. Desmarais, R. S. T. Martins, J. Hecht, F. Knaust, K. Belkhir, S. Klages, R. Dieterich, K. Stueber, F. Piferrer, B. Guinand, N. Bierne, F. A. M. Volckaert, L. Bargelloni, D. M. Power, F. Bonhomme, A. V. M. Canario, and R. Reinhardt. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5.

Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367:451–460.

Vines, T., A. Dalziel, A. Albert, T. Veen, P. Schulte, and D. Schluter. 2016. Cline coupling and uncoupling in a stickleback hybrid zone. Evolution 70:1023-38.

Wakeley, J. 2008. Coalescent theory: an introduction. Roberts & Company Publishers, Greenwood Village, Colorado.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. Evolution 38:1358–1370.

Welch, J. J., and C. D. Jiggins. 2014. Standing and flowing: the complex origins of adaptive variation. Mol. Ecol. 23:3935–3937.

Westram, A. M., M. Rafajlović, P. Chaube, R. Faria, T. Larsson, M. Panova, M. Ravinet, A. Blomberg, B. Mehlig, K. Johannesson, and R. Butlin. 2018. Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. Evol. Lett. 2:297–309.

Woodall, L. C., R. Jones, B. Zimmerman, S. Guillaume, T. Stubbington, P. Shaw, and H. J. Koldewey. 2012. Partial fin-clipping as an effective tool for tissue sampling seahorses, *Hippocampus* spp. J. Mar. Biol. Assoc. U. K. 92:1427–1432.

Woodall, L. C., H. J. Koldewey, J. T. Boehm, and P. W. Shaw. 2015. Past and present drivers of population structure in a small coastal fish, the European long snouted seahorse *Hippocampus guttulatus*. Conserv. Genet. 1–15.

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

1355 Yeaman, S. 2013. Genomic rearrangements and the evolution of clusters of locally
1356 adaptive loci. Proc. Natl. Acad. Sci. 110:E1743–E1751.
1357
1358 Yeaman, S., and M. C. Whitlock. 2011. The genetic architecture of adaptation under
1359 migration-selection balance. Evol. Int. J. Org. Evol. 65:1897–1911.

Supporting Information

Supporting Information SI1 Joint Site-Frequency Spectra (JSFS) based on Romiguier et al. (2014; A-), our dataset (B-) and the difference between them (C-) among Le Croisic (France), Faro (Portugal), and Thau (France) *H. guttulatus*. JSFS is a bidimensional representation of allelic frequencies spectra for two populations (Ewens 1972). It is a $(2n_1+1)$ x $(2n_2+1)$ dimension matrix, with n_1 the number of individuals in population 1 and n_2 the number of individuals in population 2, where each entry S(i,j) gives the number of SNPs for which the derived allele was found i and j times in population 1 and 2, respectively. The occurrence of biallelic polymorphism for which the derived allele was found in both populations is written as a percentage in each entry. For instance, the entry S(1, 0) shows the number of polymorphism for which the derived allele was observed one time in population 1, but not observed in population 2. In the 1st plot of Fig. SI1A, 26.6% of the derived allele was observed in only one individual at Thau (x-axis) but never observed in Le Croisic individuals (y-axis).

A)

0.9	1.2	1.5	4.2	0
2.6	2.0	2.1	2.1	2.5
5.9	4.1	4.5	2.6	1.4
16.1	5.0	2.7	2.1	1.4
0	26.6	4.8	2.3	1.3

Croisic

Thau

0.5	1.0	1.9	2.6	0
7.9	1.3	1.8	2.0	3.6
5.0	2.9	4.8	1.9	2.0
21.7	4.4	4.2	2.0	7.4
0	13.5	4.8	2.2	0.6

Faro

Croisic

0.4	1.3	1.1	4.1	0
7.1	1.6	1.6	1.8	2.6
5.0	2.5	3.5	2.4	2.0
19.7	5.7	3.4	1.6	6.1
0	20.6	3.4	1.3	0.8

Faro

Thau

B)

1.6	2.2	2.8	3.8	0
3.9	4.1	3.9	3.7	3.9
6.8	5.1	4.0	3.1	2.5
11.8	7.1	4.8	2.8	1.9
0	9.1	6.0	3.5	1.8

Croisicc

Thau

1.2	2.1	2.3	3.3	0
2.5	3.7	3.6	4.0	3.8
5.1	5.9	5.1	4.7	3.2
8.9	7.7	5.6	4.6	2.5
0	10.0	5.6	3.4	1.4

Faro

Croisic

1.6	1.5	2.3	3.6	0
3.0	3.5	3.8	3.9	3.7
5.9	6.3	5.3	3.9	2.9
10.4	6.3	5.5	3.2	2.2
0	9.2	6.6	3.0	1.8

Faro

Thau

C)

-0.77	-0.97	-1.31	0.44	0
-1.34	-2.05	-1.78	-1.55	-1.35
-0.86	-0.98	0.42	-0.48	-1.09
4.39	-2.10	-2.09	-0.77	-0.56
0	12.52	-1.18	-1.15	-0.44

Croisic

Thau

-0.68	-1.08	-0.411	-0.76	0
5.41	-2.39	-1.83	-1.95	-0.15
-0.11	-3.00	-0.27	-2.83	-1.18
12.84	-3.22	-1.43	-2.57	4.96
0	3.48	-0.77	-1.23	-0.82

Faro

Croisic

-1.18	-0.15	-1.21	0.56	0
4.18	-1.86	-2.22	-2.02	-1.06
-0.83	-3.75	-1.76	-1.47	-0.90
9.37	-2.59	-2.13	-1.61	3.94
0	11.42	-2.12	-1.68	-0.93

Faro

Thau

Legend:

0-5	5-10	10-15	15-20	20-25	25-30

18

Supporting Information SI2 *H. guttulatus* genetic diversity and structure

Table SI2-1 *Hippocampus guttulatus* sample information and genetic diversity indices of the study samples based on 286 SNP markers.

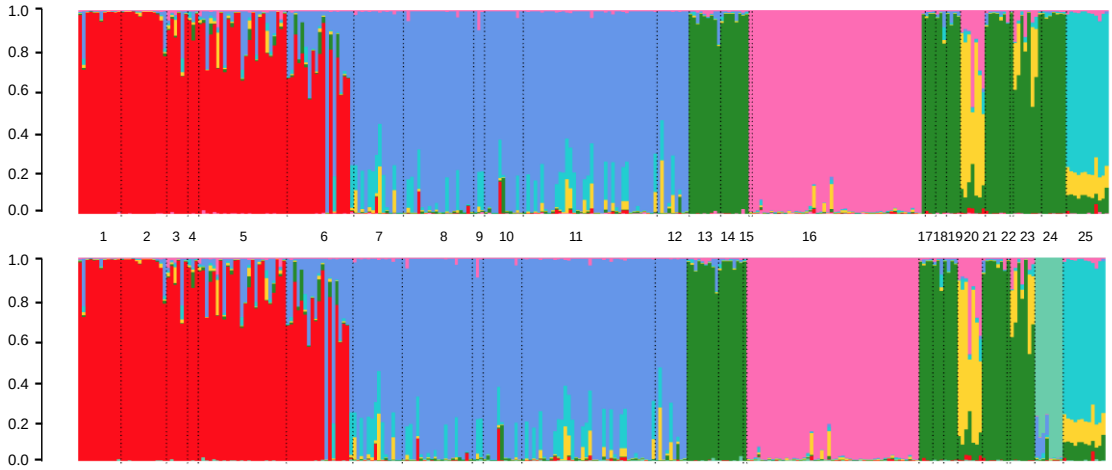
N: number of individuals successfully genotyped, H_e: expected heterozygosity, F_{IS}: fixation index

label	country	location	*N*	H_e	F_{IS}
North Atlantic cluster (sites 1 to 5)			*59*	*0.331*	*0.019*
1	UK	Poole	12	0.304	0.013
2	France	Brest	14	0.305	-0.019
3	France	Le Croisic	5	0.327	-0.013
4	France	Ré Island	3	0.331	0.018
5	France	Arcachon	25	0.330	-0.013
6	France	Hossegor	19	0.351	0.042
South Atlantic cluster (sites 7 to 12)			*95*	*0.337*	*-0.011*
7	Spain	Corogne	14	0.337	-0.019
8	Spain	Vigo	20	0.338	-0.027
9	Portugal	Portimao	3	0.343	0.009
10	Portugal	Faro (sea)	11	0.339	0.019
11	Portugal	Faro (lagoon)	38	0.334	-0.014
12	Spain	Malaga	9	0.335	-0.009
Mediterrannean cluster (sites 13 to 23 without sites 16 and 20)			*44*	*0.336*	*-0.002*
13	Spain	Tossa	9	0.343	0.018
14	France	Leucate	8	0.335	-0.023
15	France	Sète	1	-	-
16	France	Thau	49	0.316	0.017
17	France	La Ciotat	3	0.328	0.068
18	France	Le Brusc	3	0.345	-0.020
19	France	Cavalaire-sur-Mer	4	0.339	-0.029
20	Tunisia	Bizerte	7	0.344	0.039
21	Italy	Naples	7	0.336	-0.025
22	Croatia	Croatia	1	-	-
23	Greece	Kalamaki	8	0.319	-0.035
24	Greece	Halkida	7	0.324	0.015
25	Bulgaria	Varna	12	0.255	-0.014
				0.374	*0.127*

Table SI2-2 Genetic structure (pairwise F_{ST} estimates) between the clusters identified. Probability values for exact tests, corrected for multiple comparisons, not provided here, were all lower than 0.001.

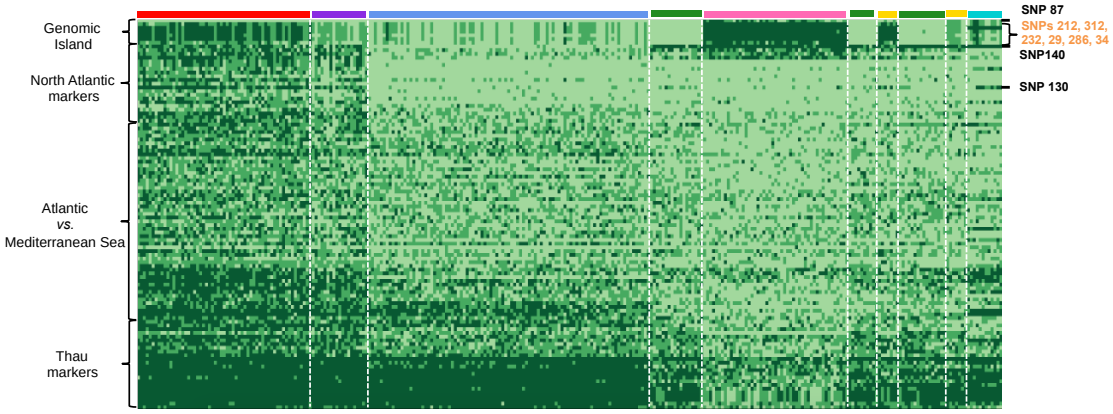
	North Atlantic	South Atlantic	Med. lagoon (Thau)	Med. marine sites	Bizerte	Halkida	Varna
North Atlantic							
South Atlantic	0.178						
Med. lagoon	0.223	0.160					
Med. marine sites	0.200	0.087	0.115				
Bizerte	0.158	0.114	0.067	0.070			
Halkida	0.187	0.116	0.105	0.024	0.054		
Varna	0.260	0.186	0.196	0.150	0.171	0.154	

Supporting Information SI3 Genetic population structure based on 286 SNP markers analyzed by Individual Bayesian ancestry proportions determined using STRUCTURE with K=6 and K=7 clusters identified. Dotted black lines separate each study site. The five clusters identified are distinguished by the same colors and numbers as used in Fig. 1, the 6th cluster in gold and the 7th cluster in orange. Each individual is depicted as a vertical bar with colors distinguishing its ancestries to the clusters.



Supporting Information SI4 Genetic population structure based on 286 SNP markers

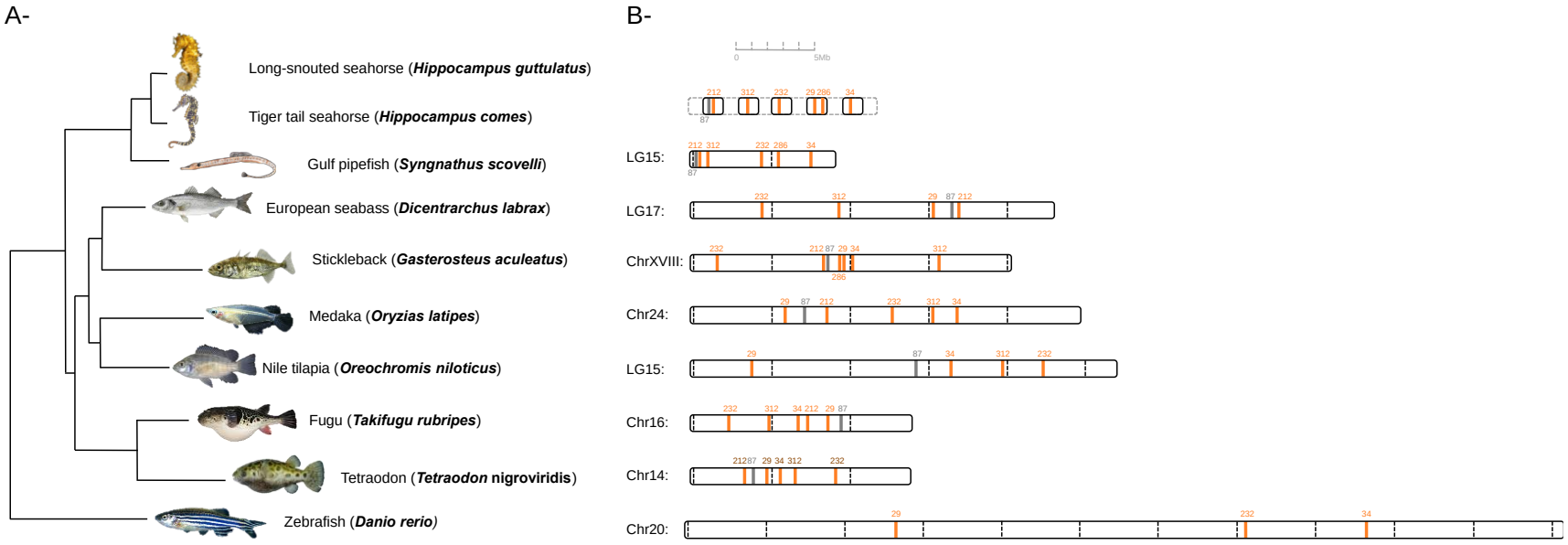
The raw genotype matrix for a subset of 105 most differentiated SNP markers and all individuals. Individuals are ordered based on their locations (x-axis). Each rectangle denotes an individual's genotype at a given locus: dark green (homozygote for the allele most frequent in the North Atlantic cluster), green (heterozygotes) and light green (homozygote for the minor allele in the North Atlantic cluster). The names of the outlier SNPs are mentioned at the right of the matrix. Convergent SNPs are represented in orange while the other outliers are in black. Each cluster is colored according to Fig. 1 and Fig. 2 above the raw genotype matrix.



Supporting Information SI5 Blat against seven well-assembled fish genomes

Although *H. guttulatus* genome is unknown, seven well-assembled genomes of closely related fishes with placed scaffolds are available. By blatting/blasting *H. guttulatus* contigs comprising the outliers against these seven well-assembled genomes, six to seven of the outliers were recurrently reported on a unique chromosome in each species, and consistently in four of the seven fish genomes (Fig. and Table SI4). Four of these outlier loci (SNPs 232, 312, 87) mapped to this chromosome for all the seven fish genomes (colored in orange and grey in Fig. SI4, Table SI4).

Figure SI5 Phylogenomic tree (A-), and outlier cross-mapping (B-). 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. 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 are colored in orange, while SNP 87, an outlier highly differentiated between Mediterranean ecotypes is colored in grey.



Contig	SNP	mapping on the seabass genome	mapping on the stickelback genome	mapping on the medaka genome	mapping on the tilapia genome	mapping on the fugu genome	mapping on the tetraodon genome	mapping on the zebrafish genome	mapping on the tiger tailed seahorse unplaced genomic scaffold	mapping on the gulf pipefish genome
Cont4384	1	UN	Un	5	GL831288-1	19	11	4	scaffold68	LG20
Cont28960	2	NO MATCHES	VII	ultracontig115	LG22	15	7	NO MATCHES	scaffold33	LG11
Cont17577	3	LG1B	V	no matches	LG8-24	no matches	2	NO MATCHES	scaffold330	LG10
Cont76931	4	LG19	XIV	12	LG7	6	4	10	scaffold145	LG8
Cont18820	5	LG11	VI	no matches	LG13	4	17	13	scaffold36	LG22
Cont10979	6	LGx	IV	23	LG17	18	19	4	scaffold63	LG17
Cont1508	7	LG11	VI	15	LG13	4	17	13	scaffold76	LG22
Cont36293	8	NO MATCHES	Un	8	LG4	HE594104	17	20	scaffold314	LG16
Cont14853	9	LG8	XI	8	LG4	5	Un_random	3	scaffold314	LG16
Cont16331	10	LG19	XIV	12	LG7	6	4	5	scaffold183	LG8
Cont24807	11	NO MATCHES	XV	scaffold1184	LG15	no matches	Un_random	19	scaffold221	no matches
Cont2239	12	LG16	XX	no matches	LG11	7	8	16	scaffold383	LG3
Cont21934	13	LG8	XI	no matches	LG4	5	Un_random	4	scaffold314	LG16
Cont14860	15	NO MATCHES	VIII	4	LG2	HE592589	17	20	scaffold384	LG9
Cont55395	16	LG12	XV	22	LG19	2	10	17	scaffold331	LG19
Cont7542	17	NO MATCHES	V	scaffold2590	GL831743-1	14	2	13	scaffold620	LG10
Cont12824	19	LG4	VIII	4	GL831204-1	22	15	17	scaffold132	LG9
Cont14921	20	LG8	XI	8	LG4	5	3	3	scaffold93	LG16
Cont11766	21	LG16	XX	ultracontig182	LG11	7	8	16	scaffold94	LG3
Cont34018	22	LG13	I	13	LG14	11	16	13	scaffold125	LG18
Cont99369	23	LG12	XV	22	LG19	2	10	2	scaffold92	LG19
Cont20099	24	NO MATCHES	III	no matches	GL831354-1	22	Un_random	2	scaffold117	LG6
Cont70320	25	LG4	VIII	4	LG23	HE591825	2	5	scaffold22	LG9
Cont10895	26	LG15	XVI	21	LG16-21	1	2	15	scaffold111	LG7

Cont25919	27	NO MATCHES	XII	7	AERX01073340-2	16	Un_random	9	scaffold174	LG1
Cont83384	29	LG17	XVIII	24	LG15	16	14	20	scaffold154	no matches
Cont15642	30	LG7	IX	1	LG6	17	18	1	scaffold51	LG5
Cont3263	31	LG1B	V	19	LG8-24	1	2	12	scaffold17	LG10
Cont8597	32	LG10	III	17	LG18	HE591745	15	2	scaffold85	LG6
Cont23724	33	LG1A	XVII	ultracontig62	LG5	19	Un_random	24	C16644750	LG20
Cont2140	34	LG6	XVIII	24	LG15	16	14	20	scaffold100	LG15
Cont9669	35	LG7	IX	1	LG6	17	18	12	scaffold51	LG5
Cont28226	36	LG9	X	11	LG22	12	21	16	scaffold163	LG8
Cont1200	37	LG18-21	XXI	20	LG9	10	15	NO MATCHES	scaffold501	LG21
Cont5402	38	NO MATCHES	VI	9	LG13	4	17	13	scaffold76	LG22
Cont598	39	LG20	XIII	9	LG12	21	12	1	scaffold155	LG12
Cont16697	40	UN	VII	14	GL831331-1	15	7	5	scaffold66	LG11
Cont77955	41	LG16	XX	16	LG11	7	8	16	scaffold327	LG3
Cont8845	42	LG7	IX	1	LG6	HE591689	Un_random	1	scaffold158	LG8
Cont13226	43	LG19	XIV	12	LG7	6	4	10	scaffold185	LG8
Cont26323	44	LG16	XX	16	LG11	7	8	19	scaffold94	LG3
Cont20336	45	NO MATCHES	IX	no matches	LG6	8	Un_random	1	scaffold7	LG5
Cont84033	46	LG20	XIII	9	LG12	HE591823	2	12	scaffold147	LG12
Cont2647	48	LG6	XIX	6	LG7	9	13	25	scaffold104	LG2
Cont959	49	LG22-25	XII	7	LG20	3	9	1	scaffold174	LG20
Cont8405	50	NO MATCHES	XI	8	LG4	HE593597	1_random	13	scaffold427	LG4
Cont17342	51	LG20	XIII	9	LG12	21	Un_random	5	scaffold322	LG12
Cont5748	52	LG8	XI	6	LG5	8	9	21	scaffold171	LG16
Cont31658	53	LG22-25	XII	7	LG20	3	9	23	scaffold14	LG1
Cont53822	54	NO MATCHES	VIII	ultracontig115	LG5	no matches	Un_random	3	scaffold275	no matches
Cont12257	55	LG14	VII	14	LG10	15	7	11	scaffold57	LG11
Cont15081	56	LG6	XIX	6	LG7	HE592015	Un_random	18	scaffold112	LG2

Cont15386	57	NO MATCHES	XII	11	LG18	no matches	9	24	scaffold124	no matches
Cont15656	58	LG12	XV	22	LG19	2	10	13	scaffold176	LG19
Cont7271	59	LG10	IX	1	LG6	HE593635	21	17	scaffold86	LG5
Cont26255	60	NO MATCHES	IX	1	LG6	17	18	1	scaffold7	LG5
Cont18545	61	LG2	IV	10	LG2	22	1	14	scaffold156	LG14
Cont9410	62	LG8	XI	8	LG4	HE592480	19	1	C16899674	no matches
Cont109	63	LG4	IX	20	GL831336-1	10	Un_random	14	scaffold490	LG21
Cont48307	64	LG10	III	17	LG18	HE591799	Un_random	21	scaffold255	LG6
Cont33723	65	LG7	IX	1	LG6	17	18	1	scaffold51	LG5
Cont8390	66	LG5	II	3	LG1	13	5	7	scaffold150	LG4
Cont5102	67	LG24	I	2	GL831144-1	8	3	1	scaffold345	LG1
Cont28256	68	LG14	VII	14	LG10	15	7	20	scaffold57	LG11
Cont25932	69	LG9	X	scaffold794	LG22	12	21_random	19	scaffold174	LG1
Cont8790	70	LG8	XI	8	LG4	5	Un_random	3	scaffold10	LG16
Cont6242	71	LG24	I	2	GL831324-1	HE591766	14	17	scaffold345	LG15
Cont14556	72	NO MATCHES	III	17	LG18	HE591799	Un_random	2	scaffold255	LG6
Cont27390	73	LG19	VII	12	LG7	6	14	19	scaffold232	LG8
Cont19945	74	LG9	X	no matches	LG22	12	21	19	scaffold244	LG13
Cont21029	75	LG20	XIII	no matches	LG12	21	12	10	scaffold155	LG12
Cont14962	76	LG13	I	13	LG14	4	2	1	scaffold29	LG18
Cont23817	77	LG4	VIII	17	GL831204-1	20	1	10	scaffold132	LG9
Cont8589	78	LG16	XX	16	LG11	7	Un_random	16	scaffold315	LG3
Cont16601	79	NO MATCHES	VII	20	AERX01074718-1	8	Un_random	4	scaffold18	LG14
Cont26079	80	UN	XI	scaffold2279	LG4	HE591950	3	19	scaffold99	LG16
Cont33822	81	LG20	XIII	9	LG12	HE591722	Un_random	5	scaffold45	LG12
Cont41945	83	LG7	IX	1	LG6	17	18	1	scaffold86	LG5
Cont23307	84	LG12	XV	no matches	GL831422-1	9	14	17	scaffold107	LG19
Cont25415	85	LG16	XX	8	LG11	7	8	10	scaffold315	LG3

Cont26534	86	LG2	IV	10	LG2	14	1	9	scaffold9	LG14
Cont9143	87	LG17	XVIII	24	LG15	16	14	8	scaffold8	LG15
Cont87996	90	LG20	XIII	9	LG12	21	12	5	scaffold147	LG12
Cont21350	91	LG19	XIV	no matches	LG7	6	4	10	scaffold145	LG8
Cont5243	92	LG1A	Un	scaffold3606	LG5	19	11	6	scaffold68	LG20
Cont4052	93	NO MATCHES	II	no matches	GL831436-1	no matches	Un_random	12	scaffold154	no matches
Cont10293	94	LG2	IV	10	GL831552-1	HE591614	Un_random	14	scaffold203	LG14
Cont34439	95	LG22-25	XII	ultracontig49	GL831308-1	3	Un_random	23	scaffold174	LG1
Cont13746	96	LG9	X	11	LG22	12	21	19	scaffold16	LG11
Cont4345	97	LG20	XIII	9	LG12	21	12	10	scaffold155	LG12
Cont6023	98	UN	Un	ultracontig89	LG20	3	Un_random	8	scaffold245	LG20
Cont12506	99	LG12	XV	15	LG19	2	Un_random	2	scaffold176	LG19
Cont11299	101	LG12	XV	22	LG19	2	10	20	scaffold32	LG19
Cont93017	102	NO MATCHES	IV	13	LG13	5	3	NO MATCHES	scaffold78	LG12
Cont13559	103	LG19	Un	12	LG7	HE595610	Un_random	21	scaffold183	LG8
Cont14019	104	NO MATCHES	IX	scaffold521	LG6	17	18	11	scaffold86	LG5
Cont11526	105	LG10	III	17	LG18	22	15	6	scaffold317	LG6
Cont14070	106	UN	Un	20	GL831288-1	11	10	6	scaffold420	LG20
Cont6347	107	LG7	IX	1	LG6	17	18	11	scaffold86	LG5
Cont2549	108	LG15	XIII	scaffold3077	LG14	3	18	17	scaffold177	LG11
Cont3550	110	LG22-25	XII	7	LG20	HE591958	9	8	scaffold174	LG1
Cont30259	111	LG1A	Un	5	LG5	19	11	6	scaffold68	LG20
Cont69271	112	LG6	XIX	6	LG13	9	13	25	scaffold1501	LG2
Cont28490	113	LG4	VIII	4	LG23	20	1	11	C16361965	LG9
Cont13741	114	LG2	IV	10	GL831239-1	14	1_random	5	scaffold156	LG14
Cont15454	115	LG16	XX	10	LG11	7	8	23	scaffold327	LG3
Cont18687	116	NO MATCHES	IV	no matches	LG17	18	19	21	scaffold127	LG17
Cont26203	117	LG20	XIII	10	LG12	HE591882	10	17	scaffold179	no matches

Cont8075	118	UN	Un	11	GL831438-1	HE592393	Un_random	16	scaffold73	LG13
Cont20917	120	LG19	XIV	no matches	LG7	HE592064	Un_random	10	scaffold101	LG8
Cont14449	122	UN	XX	3	LG16-21	16	no matches	3	scaffold2808	LG3
Cont8353	123	LG22-25	XII	7	LG20	3	9	8	scaffold157	LG1
Cont10848	124	LG10	III	17	LG18	22	15	2	scaffold317	LG6
Cont1099	125	LG7	IX	1	LG6	17	18	1	scaffold51	LG5
Cont25313	126	LG22-25	XII	3	LG9	HE593724	Un_random	13	scaffold28	LG1
Cont11253	127	LG6	XIX	6	LG7	9	13	13	scaffold570	LG2
Cont18851	128	LG20	XIII	no matches	LG12	no matches	no matches	16	scaffold130	no matches
Cont9335	129	LG7	IX	1	LG6	HE591737	Un_random	14	scaffold51	LG5
Cont642	130	UN	X	ultracontig72	LG7	HE592842	13	10	scaffold37	LG2
Cont2041	132	NO MATCHES	IV	10	LG2	13	18	10	scaffold118	LG14
Cont3347	133	LG6	XIX	6	LG7	9	13	5	scaffold104	LG2
Cont32959	134	LG14	VII	14	LG10	HE591967	Un_random	10	scaffold177	LG11
Cont2853	135	NO MATCHES	IV	24	LG8-24	no matches	no matches	8	scaffold367	LG10
Cont35401	137	LG7	IX	1	LG6	HE591988	Un_random	1	scaffold51	LG5
Cont10193	138	NO MATCHES	XVI	10	GL831574-1	no matches	2	15	scaffold409	LG1
Cont96526	139	LG6	XIX	6	GL831310-1	HE592940	13	19	scaffold298	LG2
Cont1243	140	LG13	I	13	LG14	11	16	19	scaffold6	LG18
Cont68920	141	LG9	X	11	LG22	12	21_random	19	scaffold87	LG13
Cont7267	142	NO MATCHES	Un	20	LG9	10	Un_random	12	scaffold25	LG21
Cont32750	143	LG5	II	3	LG1	13	5	7	scaffold378	LG4
Cont8984	144	UN	II	18	LG17	17	Un_random	13	scaffold55	LG3
Cont32113	145	NO MATCHES	IX	7	LG22	15	18	14	scaffold51	LG5
Cont32643	147	UN	Un	8	GL831404-1	5	Un_random	3	scaffold187	LG16
Cont11887	148	LG8	XI	ultracontig104	GL831541-1	5	15	1	scaffold257	LG16
Cont2465	149	LG19	XIV	12	LG7	6	4	1	scaffold30	LG8
Cont44942	150	UN	XIX	12	LG1	20	Un_random	5	scaffold150	LG4

Cont27645	151	LG19	XIV	1	LG7	6	4	2	scaffold347	LG8
Cont103950	153	LG6	XIX	6	GL831206-1	HE592015	Un_random	4	scaffold37	LG2
Cont25796	154	LG5	II	6	LG4	13	no matches	12	scaffold82	LG4
Cont4827	156	LG18-21	XXI	20	LG9	HE592042	Un_random	18	scaffold293	LG21
Cont14555	157	LG7	IX	22	LG6	17	Un_random	1	scaffold7	LG5
Cont31838	158	LG2	IV	10	LG2	14	1	Un	scaffold9	LG14
Cont9025	159	LG22-25	XII	7	LG20	3	Un_random	8	scaffold1	LG1
Cont4262	160	LG19	XIV	12	LG7	HE592442	Un_random	21	scaffold437	LG8
Cont9081	161	LG11	VI	15	LG13	HE591905	17	13	scaffold76	LG22
Cont11271	162	LG7	IX	1	LG6	17	18	1	scaffold7	LG5
Cont30153	163	LG22-25	XII	7	LG20	HE591793	Un_random	23	scaffold139	LG1
Cont3030	164	LG11	VI	18	GL831350-1	7	13	13	scaffold352	LG22
Cont7002	165	LG17	XVIII	24	LG15	1	17	2	scaffold144	LG15
Cont123	166	LG2	IV	10	LG2	14	1	14	scaffold9	LG14
Cont3077	167	LG16	XX	16	LG11	7	8	16	scaffold94	LG3
Cont12396	168	NO MATCHES	XV	2	GL831306-1	no matches	no matches	4	scaffold43	LG19
Cont35424	169	LG5	II	3	LG1	13	5	7	scaffold261	LG4
Cont19337	170	LG12	XXI	no matches	LG20	HE592690	18	21	scaffold174	LG1
Cont9720	171	LG4	VIII	4	GL831204-1	20	1	8	scaffold132	LG9
Cont35271	172	LG12	XV	scaffold997	LG19	HE591840	10	17	scaffold221	no matches
Cont25942	173	LG5	II	3	LG1	13	5	1	scaffold21	LG4
Cont5609	174	LG15	XVI	ultracontig257	LG16-21	1	5	9	scaffold111	LG7
Cont7337	176	LG14	IV	14	GL831184-1	15	7	17	scaffold59	LG11
Cont11323	177	LG19	XIV	scaffold5324	LG10	6	2	NO MATCHES	scaffold600	LG8
Cont21369	178	UN	XVIII	no matches	LG12	no matches	4	4	scaffold170	LG15
Cont34108	179	LG14	VII	14	GL831331-1	HE591963	7	21	scaffold177	LG11
Cont32218	180	NO MATCHES	XIX	6	LG13	9	13	12	scaffold62	LG2
Cont13604	181	NO MATCHES	XX	16	LG11	7	Un_random	7	C17013657	LG3

Cont1765	183	LG5	II	3	LG1	13	5	18	scaffold261	LG4
Cont2573	184	LG16	XX	16	LG11	7	8	2	scaffold242	LG3
Cont25894	185	NO MATCHES	XIV	6	LG13	HE591871	5	6	scaffold145	LG8
Cont15558	186	LG13	I	13	LG14	11	10	15	scaffold29	LG18
Cont17934	187	LG7	IX	1	LG6	HE591737	Un_random	8	scaffold51	LG5
Cont25867	188	LG13	I	11	LG14	11	16	13	scaffold72	LG18
Cont29792	189	LG6	XIX	6	LG13	9	13	24	scaffold62	LG2
Cont33785	190	NO MATCHES	II	3	LG17	14	Un_random	25	scaffold275	LG4
Cont21458	192	LG13	I	no matches	LG14	HE591604	16	5	scaffold122	no matches
Cont18822	193	UN	IV	no matches	LG1	18	no matches	16	scaffold81	LG17
Cont34240	194	LG22-25	XII	7	LG20	3	9	22	scaffold152	LG1
Cont10740	195	LG9	X	11	LG22	12	Un_random	19	scaffold401	LG13
Cont7133	196	UN	VII	14	GL831331-1	18	Un_random	5	scaffold66	LG11
Cont17712	197	LG10	III	17	LG18	22	15	2	scaffold228	no matches
Cont12627	198	LG18-21	XXI	20	LG9	10	Un_random	19	scaffold322	LG21
Cont76087	199	LG8	XI	8	LG4	5	Un_random	3	scaffold433	LG16
Cont28921	200	UN	III	17	GL831564-1	HE594843	Un_random	24	scaffold159	LG6
Cont3548	202	LG1B	V	19	LG8-24	1	2	12	scaffold74	LG10
Cont13811	203	LG8	XI	8	LG4	5	Un_random	3	scaffold343	LG16
Cont11788	204	LG1A	XVII	5	LG5	19	11	14	scaffold68	LG20
Cont11310	205	LG5	II	3	LG1	13	5	9	scaffold21	LG4
Cont13591	206	LG12	XV	22	LG19	2	10	18	scaffold739	LG19
Cont562	208	LG14	VII	20	LG22	HE591765	7	10	scaffold119	no matches
Cont521	209	LG6	XIX	6	LG7	HE591792	Un_random	25	scaffold110	LG2
Cont28169	210	LG18-21	XXI	20	LG9	10	6	5	scaffold181	LG21
Cont23608	211	UN	Un	no matches	LG8-24	5	Un_random	3	scaffold58	LG10
Cont16652	212	LG17	XVIII	24	GL831366-1	16	14	12	scaffold8	LG15
Cont34334	213	LG20	XIII	9	LG12	21	12	22	scaffold13	LG12

Cont24441	215	LG20	XIII	9	LG12	21	12	5	scaffold155	LG12
Cont9474	216	NO MATCHES	XII	scaffold794	LG20	2	10	9	scaffold174	LG1
Cont88381	217	LG22-25	X	7	LG20	2	9	3	scaffold41	LG1
Cont13357	218	LG1A	Un	5	LG5	19	11	6	scaffold40	LG20
Cont15610	219	NO MATCHES	VIII	22	LG23	no matches	1	8	scaffold229	LG9
Cont11435	220	NO MATCHES	III	17	LG16-21	HE591745	Un_random	10	scaffold85	LG6
Cont28359	222	UN	VII	14	GL831331-1	2	3	11	scaffold177	LG11
Cont15607	223	LG22-25	Un	7	LG20	HE591897	Un_random	8	scaffold174	LG1
Cont2754	224	LG1B	V	19	LG8-24	1	2	13	scaffold620	LG10
Cont792	225	LG4	VIII	4	LG17	20	1	6	scaffold282	no matches
Cont51071	226	LG10	III	17	LG18	HE591935	15	11	scaffold285	LG6
Cont21701	227	NO MATCHES	XVI	no matches	LG16-21	1	2	12	scaffold310	LG5
Cont32639	228	LG7	IX	22	LG22	16	6	10	scaffold158	no matches
Cont7523	229	LG16	XX	ultracontig182	LG11	7	8	16	scaffold335	LG3
Cont6068	230	UN	VI	15	LG13	4	17	13	scaffold348	LG21
Cont5773	231	LG10	III	1	LG18	22	15	24	scaffold105	LG6
Cont11263	232	LG17	XVIII	24	LG15	16	14	20	scaffold339	LG15
Cont7123	233	LG10	III	17	LG18	22	15_random	2	scaffold85	LG6
Cont1322	234	NO MATCHES	V	ultracontig223	LG8-24	HE591713	2	12	scaffold58	LG10
Cont15378	235	LG16	XX	16	LG11	7	Un_random	25	scaffold315	LG3
Cont10526	236	LG7	IX	scaffold1494	LG6	17	18	10	scaffold7	LG5
Cont32784	238	NO MATCHES	NO MATCHES	8	LG11	19	3	11	scaffold191	LG20
Cont7154	239	NO MATCHES	NO MATCHES	9	LG1	HE591723	no matches	16	scaffold249	no matches
Cont29967	242	LG16	XX	16	LG11	7	8	Un	scaffold327	LG3
Cont28892	243	NO MATCHES	XIX	6	LG7	9	13	9	scaffold186	LG2
Cont29410	244	LG7	IX	1	LG6	HE591939	Un_random	3	scaffold86	LG5
Cont9573	245	LG5	II	3	LG1	13	5	10	scaffold261	LG4
Cont14756	247	LG8	IX	21	LG5	HE592191	16	1	scaffold7	LG5

Cont35041	248	NO MATCHES	Un	scaffold4856	NO MATCHES	12	no matches	7	scaffold77	no matches
Cont21756	249	LG24	I	no matches	GL831262-1	8	3	9	scaffold283	LG1
Cont21948	250	LG22-25	XII	no matches	LG20	3	Un_random	23	scaffold1	LG1
Cont3690	251	LG7	IX	1	LG6	17	18	3	scaffold7	LG5
Cont6177	253	LG14	VII	14	LG10	15	7	6	scaffold57	LG11
Cont9851	254	LG7	IX	scaffold3161	GL831395-1	17	18	1	scaffold7	LG5
Cont24363	255	LG6	XIX	scaffold3797	LG7	9	13	18	scaffold104	LG2
Cont34776	256	LG2	IV	20	LG2	14	20	14	scaffold156	LG14
Cont16424	257	LG6	X	16	GL831310-1	9	13	20	scaffold319	LG2
Cont19623	258	LG4	VIII	no matches	LG23	20	1	22	scaffold151	LG9
Cont3718	259	LG5	II	3	LG1	13	5	7	scaffold2	LG4
Cont18756	260	LG4	VIII	no matches	GL831204-1	20	1	22	scaffold132	LG9
Cont5741	261	LG20	XIII	9	LG12	21	Un_random	10	scaffold147	LG12
Cont95374	262	LG18-21	XXI	20	LG9	10	Un_random	24	scaffold322	LG21
Cont2378	263	LG9	X	11	LG22	HE591892	8	19	scaffold166	LG13
Cont6695	264	LG22-25	XII	ultracontig90	GL831582-1	HE591827	9	23	scaffold174	LG1
Cont12415	265	LG11	VI	15	LG13	4	17	1	scaffold64	LG22
Cont6274	266	LG12	XV	22	LG19	2	10	17	scaffold443	LG19
Cont7452	267	UN	Un	22	LG19	2	Un_random	17	scaffold295	LG19
Cont9803	268	LG12	XV	22	LG19	2	10	17	scaffold112	LG19
Cont7637	269	LG1A	Un	5	LG5	19	11	11	scaffold68	LG20
Cont5882	270	NO MATCHES	VI	scaffold3613	LG13	4	Un_random	13	scaffold36	LG22
Cont17369	271	LG6	XIX	6	LG7	9	2	4	scaffold104	LG2
Cont190	274	LG6	XIX	no matches	GL831310-1	9	13	25	scaffold319	LG2
Cont33032	275	LG11	VI	15	LG13	4	17	13	scaffold76	LG22
Cont7134	276	LG18-21	XXI	ultracontig236	GL831564-1	10	1	10	scaffold680	LG21
Cont26989	277	LG10	III	17	GL831264-1	22	15	13	scaffold317	LG6
Cont32497	278	LG2	IV	10	LG2	14	1	17	scaffold9	LG14

Cont22067	279	NO MATCHES	XX	no matches	LG3	HE591751	Un_random	Un	scaffold53	no matches
Cont5181	280	LG1A	XVII	5	LG5	19	Un_random	15	scaffold68	LG20
Cont15293	281	LG18-21	XXI	20	LG9	10	6	7	scaffold451	LG21
Cont24693	282	LG22-25	XII	7	LG20	3	Un_random	23	scaffold14	LG1
Cont1305	283	LG8	XI	8	LG4	5	Un_random	3	scaffold193	LG16
Cont9662	284	LG20	XIII	9	LG12	HE591911	12	8	scaffold78	LG12
Cont17645	285	LG13	I	13	LG14	11	10	15	scaffold29	LG18
Cont60740	286	NO MATCHES	NO MATCHES	scaffold637	LG5	19	21	10	scaffold154	LG15
Cont1293	287	LG12	XV	22	LG19	2	10	13	scaffold587	LG19
Cont10382	288	LG13	I	13	LG14	22	16	18	scaffold72	LG18
Cont6703	289	NO MATCHES	IX	1	GL831609-1	17	18	18	scaffold158	LG5
Cont51421	290	LG7	III	2	GL831681-1	17	Un_random	1	scaffold301	LG5
Cont5084	291	LG1A	VII	5	LG5	19	11	11	scaffold68	LG20
Cont34339	292	LG14	VII	14	GL831235-1	15	7	21	scaffold109	LG11
Cont8013	293	LGx	IV	23	LG17	18	19	4	scaffold53	LG17
Cont16340	295	LG2	IV	10	LG2	14	1	1	scaffold118	LG14
Cont23957	296	LG6	XVIII	18	LG5	11	1_random	1	scaffold180	no matches
Cont10884	297	LG6	XIX	6	LG13	9	13	7	scaffold186	LG2
Cont93885	298	LG11	VI	22	LG13	4	1	15	scaffold57	LG22
Cont30303	299	LG1B	V	ultracontig221	LG8-24	HE591713	2	3	C16949222	LG10
Cont15330	300	LG5	II	3	LG1	13	5	11	scaffold261	LG4
Cont6569	301	LG12	XV	22	LG19	2	10	20	scaffold176	LG19
Cont6280	302	LG1B	V	19	LG8-24	1	2	24	scaffold416	LG10
Cont18078	303	LG10	III	no matches	LG18	22	15_random	13	scaffold85	LG6
Cont3646	304	LG15	XVI	21	LG16-21	1	2	23	scaffold237	LG7
Cont15961	305	LG14	VII	14	GL831331-1	15	7	5	scaffold57	LG11
Cont1392	306	LG13	Un	17	LG14	22	16	15	scaffold122	LG18
Cont6381	307	LG11	VI	15	LG13	4	17	13	scaffold108	LG22

Cont6287	308	LG16	VIII	18	LG11	7	8	8	scaffold153	LG3
Cont7606	310	LG13	I	13	LG14	11	10	15	scaffold29	LG18
Cont35431	311	LG17	XVIII	24	LG15	16	14	20	scaffold8	LG15
Cont11333	312	LG17	XVIII	24	LG15	16	14	10	scaffold397	LG15
Cont17350	313	LG6	XIX	6	LG7	9	13	18	scaffold104	LG2
Cont3764	314	NO MATCHES	Un	no matches	LG14	no matches	Un_random	16	scaffold122	LG18
Cont10557	315	LG6	XIX	ultracontig72	GL831206-1	HE592034	Un_random	25	scaffold37	LG2
Cont54629	316	LG12	XV	22	LG19	2	10	21	scaffold43	LG19
Cont30176	317	LG16	Un	16	LG11	7	Un_random	16	scaffold249	LG3
Cont1278	318	LG15	XVI	21	LG16-21	1	2	9	scaffold237	LG7

Supporting Information SI6 Genome scan of infra-specific differentiation in *H. guttulatus*. Solid and dashed lines respectively represent the 99% and 95% quantiles of the neutral envelope of F_{ST} obtained following Bonhomme et al. (2010) approach. Loci identified by all methods as outliers are colored in orange – the six outliers that displayed parallel differentiation between Atlantic lineages and Mediterranean ecotypes – and in black – the three other outliers.

