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

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

[1] Institut des Sciences de l'Evolution de Montpellier, Université Montpellier, Montpellier, France
[2] CNRS Institut des Sciences de l'Evolution, UMR5554 UM-CNRS-IRD-EPHE, Station Marine OREME, Sète, France
[3] Department of Zoology, University of Oxford, 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

Abstract

Diverging semi-isolated lineages either meet in narrow clinal hybrid zones, or have a mosaic distribution associated with environmental variation. Intrinsic reproductive isolation is often emphasized in the former and local adaptation in the latter, although both can contribute to isolation. Rarely these two patterns of spatial distribution are reported in the same study system. 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 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. Surprisingly, 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. These parallel outlier SNPs cluster on a single chromosome-wide island of differentiation. Since Atlantic lineages do not match the 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 either with space or with a patchy environment in a single study system.

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

Introduction

The spatial context of contact zones between partially isolated taxa and 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, local adaptation is often emphasized to drive mosaic distributions which are interpreted as evidence that selection at local adaptation genes is acting to increase differentiation between habitats in a parallel fashion (Nosil and Feder 2012). This sketchy dichotomy is anchored by emblematic study systems for which decades of research allow support for such interpretations. For instance, the hybrid zone between the mice *Mus musculus musculus* and *M. m. domesticus* in central Europe (Boursot et al. 1993) has been well demonstrated to be maintained by selection against hybrid genotypes (Britton-Davidian et al. 2005, Good et al. 2008) after secondary contact (Duvaux et al. 2011). Although the position of the hybrid zone was initially found to be associated with rainfall (Hunt and Selander 1973), local adaptation is not considered to contribute much to the isolation. At the other extreme, local adaptation of the same genetic variants has repeatedly allowed marine three-spined sticklebacks to evolve into a freshwater ecotype (Jones et al. 2012). Intrinsic selection is thought absent in the marine-freshwater sticklebacks system, although hybrids perform poorly in both environment and tend to reside in salinity ecotones (Vines et al. 2016). It would be misleading, however, to suggest the two alternative spatial contexts and relations to environmental variation may correspond to alternative routes to speciation (e.g. mutation-order *vs.* ecological speciation). The list of hybrid zones maintained both by local adaptation and intrinsic reproductive isolation is also long. *Bombina* toads (Szymura and Barton 1986), *Gryllus* crickets (Rand and Harrison 1989, Larson et al. 2014), or *Mytilus* mussels (Bierne et al. 2003) are well-known examples of mosaic hybrid zones maintained by exogenous and endogenous selection. However, we do not usually expect intrinsic incompatibilities to be reused in a parallel fashion. Besides, parallel genetic divergence associated with contrasting environmental conditions (e.g. marine/freshwater, highland/lowland, host races) remains a strong hallmark of ecologically-driven divergence (Bierne et al. 2013). Given this -hybrid zone- context, in

this paper we aim to provide an example of a genetic parallelism with a lack of apparent ecological convergence. We expect this counter-example could contribute to break the inductive reasoning that genetic parallelism means ecological convergence. However, as we discovered this pattern by serendipity in a newly studied complex of cryptic genetic backgrounds in a non-model system, we have to describe this system first.

We studied the population genetics of the long-snouted seahorse, *Hippocampus guttulatus,* across a large part of its geographic range. We developed an assay of 286 informative SNPs chosen from more than 2,500 SNPs identified in a population transcriptomic study (Romiguier et al. 2014). *H. guttulatus* displays poor dispersal abilities (e.g. site-fidelity, weak swimming performance, lack of dispersive stage) and inhabits fragmented coastal habitats along its distribution range (from the English Channel through the Mediterranean and Black Seas, Lourie and Vincent 2004). In addition most populations are small and patchy. Given these biological characteristics, a strong genetic structure could have been expected. In agreement, a very low genetic diversity was observed in *H. guttulatus* when compared to 75 non-model animal species (Romiguier et al. 2014). However, genetic differentiation proved to be very weak over very large geographic distances based on the two genetic studies conducted to date with microsatellite loci (e.g. from the United Kingdom to North of Spain, Woodall et al. 2015, or across the Cape Finisterre oceanographic barrier, López et al. 2015). Four well-differentiated genetic clusters, each distributed over extended regions, were delineated by genetic discontinuities corresponding to usual delimitations between vicariant marine lineages (Woodall et al. 2015) – between the Iberian Peninsula and the Bay of Biscay in the North Eastern Atlantic, between the Atlantic Ocean and the Mediterranean Sea, and between the Mediterranean and Black Seas. Although 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, which boundaries 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) 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, we challenged the initial interpretation of barriers to dispersal against the alternative hypothesis of reproductive isolation between semi-isolated genetic backgrounds. We described 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. Surprisingly, a shared genomic island of clustered outlier loci was involved both in the isolation between the two parapatric lineages in the Atlantic Ocean and between the marine and lagoon ecotypes in the Mediterranean Sea. 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

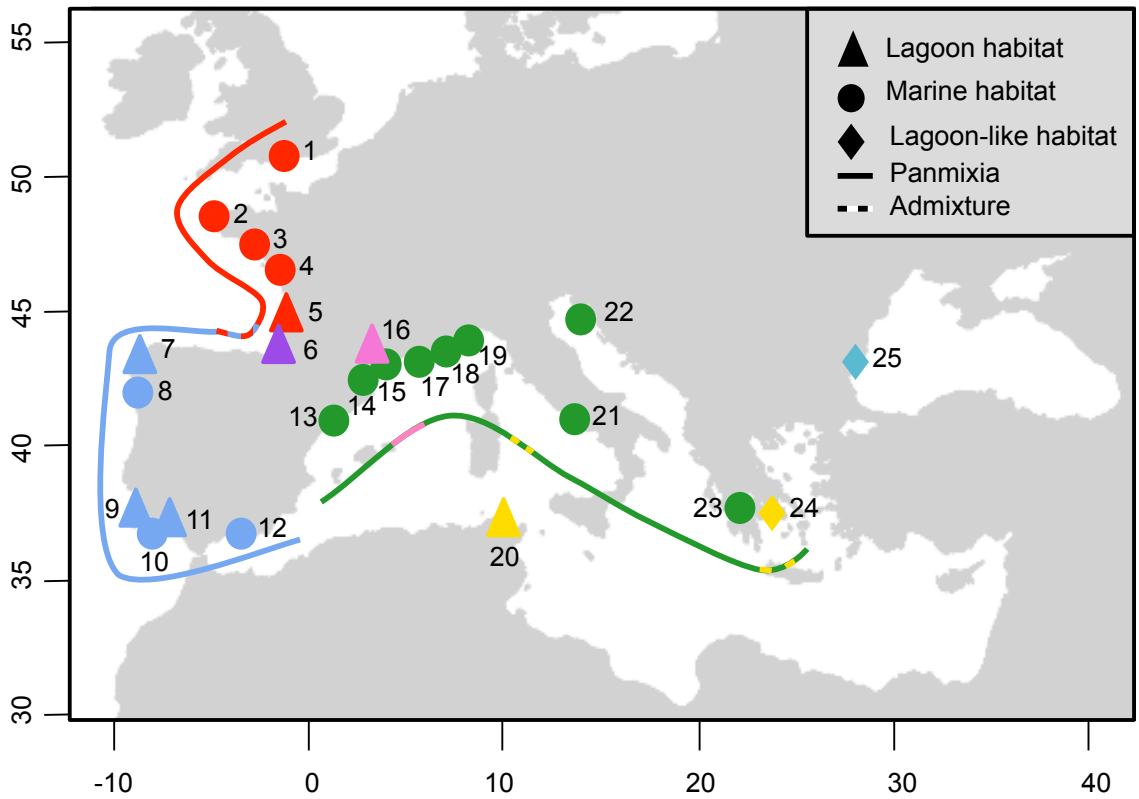


Figure 1 Sampling locations of *Hippocampus guttulatus*. Each study site is labeled as follow: 1- Poole, United Kingdom, 2- Brest, France, 3- Le Croisic, France, 4- Ré Island, France, 5- Arcachon, France, 6- Hossegor, France, 7- Coruna, Spain, 8- Vigo, Spain, 9- Portimao, Portugal, 10- Faro (maritime site), Portugal, 11- Faro (lagoon site), Portugal, 12- Malaga, 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). Dashed lines stand for contact zones (sites 6, 20 and 23).

Sampling and DNA extraction

H. 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

177 releasing back the individual into its natural habitat. In three sites (sites 14, 18 et 19 in
178 Fig. 1), dorsal fins were clipped from captive-bred seahorse held at the Mare Nostrum
179 Aquarium, France, recently sampled in their natural habitats. Each individual sample was
180 preserved and stored in 96% ethanol for subsequent genetic analyses.
181 Whole genomic DNA was extracted following either Woodall et al. (2015), López
182 et al. (2015) or using a standard CetylTrimethyl Ammonium Bromide (CTAB)
183 Chloroform:Isoamyl alcohol (24:1) protocol (Doyle and Doyle 1987). For samples from
184 Vigo (Spain), the whole genome was amplified using the GenomiPhi kit (GE HealthCare)
185 according to the manufacturer's protocol. Quality and quantity of DNA extraction was
186 checked on an agarose gel, and normalized to 35 $ng.\mu L^{-1}$ using Qubit Fluorometric
187 Quantitation (Invitrogen).
188
189 *Data mining for SNP markers*
190 A set of 12,613 contigs was examined to identify SNPs. This included one
191 mitochondrial contig (GenBank accession number: AF192664) and 12,612 contigs from
192 Romiguier et al. (2014). Briefly, Romiguier et al. (2014) produced high-coverage
193 transcriptomic data (RNAseq) for six *H. guttulatus* from three locations (Le Croisic,
194 Atlantic Ocean, France; Faro, Atlantic Ocean, Portugal; Thau lagoon, Mediterranean Sea,
195 France), and two *H. hippocampus* from two locations (Sète, Mediterranean Sea, France;
196 Bizerte, Mediterranean Sea, Tunisia). *De novo* transcriptome assembly based on Illumina
197 reads was performed using a combination of the programs ABySS (Simpson et al. 2009)
198 and Cap3 (Huang and Madan 1999), then mapped to predicted cDNAs (contigs) with
199 BWA (Li and Durbin 2009). Based on these 12,613 contigs, SNPs were identified using
200 the bioinformatic pipeline described in Bouchemousse et al. (2016). SNPs were called
201 with Read2SNPs (Gayral et al. 2013) and filtered out according to the following criteria
202 to exclude: 1) SNPs showing more than two alleles, 2) SNPs failing to be sequenced in at
203 least one location, 3) SNPs present in only one individual (*i.e.* singletons), 4) SNPs
204 identified as paralogs using the paraclean option of Read2SNPs, and 5) SNPs closer than
205 20 bp from a contig extremity or an exon limit when blasted against the stickleback, cod
206 and tilapia genomes. This resulted in 2,684 selected SNPs screened with the Illumina
207 Assay Design Tool (ADT) software to select 384 SNPs on the basis of their quality index
208 (ADT score > 0.6). An Illumina BeadXpress® with Veracode™ technology
209 (GoldenGate® Genotyping Assay) was then used to genotype the 384 selected SNPs.
210 To identify their chromosomal positions, the template sequences of the targeted
211 SNP were blasted against i) the genome of the Gulf pipefish (*Syngnathus scovelli*, Small
212 et al. 2016) and ii) the scaffolds of the tiger tail seahorse (*Hippocampus comes*, Lin et al.
213 2016). *H. comes* scaffolds being unplaced, we blasted these scaffolds against *S. scovelli*
214 genome and, for more consistency, blatted (Bhagwatt et al. 2012) them against seven
215 well-assembled fish genomes; zebrafish (*Dano rerio*, Howe et al. 2013), fugu (*Takifugu*
216 *rubripes*, Kai et al. 2011), tetraodon (*Tetraodon nigroviridis*, Jaillon et al. 2004), Nile
217 tilapia (*Oreochromis niloticus*, Brawand et al. 2014), medaka (*Oryzias latipes*, Kasahara
218 et al. 2007), stickleback (*Gasterosteus aculeatus*, Jones et al. 2012), and European
219 seabass (*Dicentrarchus labrax*, Tine et al. 2014).
220 SNPs were polarized using *H. hippocampus* as an outgroup to identify the most
221 parsimonious ancestral variant, which allowed the derived allele state to be identified.
222 The Joint Site-Frequency Spectrum (JSFS is the more informative summary statistics

223 regarding inter-population polymorphism; Wakeley 2008) obtained from the original
224 transcriptome-wide SNP dataset (Romiguier et al. 2014) was compared to the JSFS
225 obtained by the subset of 384 SNPs to investigate the extend of ascertainment bias
226 potentially induced by our marker selection. In order to compare the JSFS obtained with
227 both datasets, the JSFS dimension was projected down to a 5x5 matrix, which was the
228 dimension of the Romiguier et al. (2014) dataset.
229
230 *Analyses of Genetic Diversity and Genetic Structure*
231 The mean number of alleles per locus (N_{all}) and the allelic richness (A_r, *i.e.* the
232 expected number of alleles corrected for sampling size based on a rarefaction method)
233 were estimated using the function divBasic implemented in the R package diveRsity
234 (Keenan et al. 2013). Allelic frequencies, expected heterozygosity (H_e) and fixation index
235 (F_{IS}) were estimated using GENEPOP on the web (Raymond and Rousset 1995, Rousset
236 2008). To get a genome-wide picture from the multi-locus genotype dataset summarizing
237 inter-population polymorphism, the raw SNP data were visualized by INTROGRESS
238 (Gompert and Buerkle 2010), although its primary goal is to estimate genomic clines, but
239 allows in a simple manner to plot raw data to investigate any genomic signal. Alleles
240 derived from each of the parental populations were counted at the individual level and
241 converted into a matrix of counts, then used to visualize the multilocus genotype of each
242 individual.
243 Genetic structure among sampling sites was depicted using both Principal
244 Component Analysis (PCA) computed on the matrix of genotypes, and an individual-
245 based Bayesian clustering method. The latter being a model-based approach with strong
246 priors and hypotheses (Hardy-Weinberg equilibrium, no linkage disequilibrium)
247 contrasted with PCA, a distant-based approach when few (nearly no) assumptions may be
248 violated. Comparing results from both analyses using different statistical approaches,
249 allows us to make solid assumptions about our data. The PCAs were carried out using the
250 R package ADEGENET 1.4–2 (Jombart 2008, Jombart et al. 2011). The individual-based
251 Bayesian clustering analysis was performed with the software STRUCTURE 2.3.4
252 (Pritchard et al. 2000, Falush et al. 2003). For each value of K (ranging from 1 to 25), 30
253 replicate chains of 150,000 Monte Carlo Markov Chain (MCMC) iterations were run
254 after discarding 50,000 burn-in iterations. An admixture model with correlated allele
255 frequencies was applied with *a priori* information on samples origin. Note that this
256 method makes the assumption of homogeneous admixture rate in the genome (neutral
257 admixture) and therefore return a sort of weighted average admixture rate when
258 introgression is heterogeneous across the genome. To determine individual ancestry
259 proportions (q-values) that best matched across all replicate runs, CLUMPP (Jakobsson
260 and Rosenberg 2007) was used and individuals' assignment visualized in the R software.
261 Once the different genetic clusters were identified (and cross-validated using both
262 methods), genetic homogeneity among samples belonging to a cluster was checked,
263 allowing subsequently pooling of samples into clusters with a minimum of seven
264 individuals per cluster. Genetic structure was computed among clusters by calculating
265 global and pairwise F_{ST} (Weir and Cockerham 1984) using GENEPOP on the web
266 (Rousset 2008). Exact tests for population differentiation (10,000 dememorization steps,
267 500 batches and 5,000 iterations per batch) were carried out to test for differences in
268 allele frequencies. s, defined as the adjusted p-values using an optimized False Discovery

269 Rate approach, were computed using the QVALUE package in the R software (Storey
270 2002) to correct for multiple testing.
271 Evolutionary history of genetic clusters was also investigated under a model of
272 divergence and admixture events using the population graph approach implemented in
273 the TREEMIX software (Pickrell and Pritchard 2012). This software uses the covariance
274 matrix of allele frequency between pairs of populations to infer both population splits and
275 gene flow. A maximum likelihood population tree is first generated under the hypothesis
276 of an absence of migration, and admixture events are sequentially added, improving (or
277 not) the tree model. This statistical method shows the benefit of constructing population
278 trees while testing for gene flow between diverged populations at the same time. Samples
279 with too small sample size (N<7) or without random mating (see Results) were removed
280 as they generate erroneous results (Pickrell and Pritchard 2012). Using the total data set,
281 *i.e.* 286 loci, five migration events were sequentially added to look for the best tree to fit
282 the data and the number of migration events that reached an asymptotic likelihood was
283 kept.
284
285 *Outlier detection*
286 The use of several independent methods is often recommended to improve
287 accuracy of outlier loci detection (Pérez-Figueroa et al. 2010, de Villemereuil et al.
288 2014). Each method is differently impacted by the genetic structure and/or demographic
289 history of the study species, which is usually unknown, leading to frequent
290 inconsistencies across methods (Gagnaire et al. 2015). To cope with these problems,
291 outlier loci were detected using four different methods. First, BAYESCAN (Foll and
292 Gaggiotti 2008) is a Bayesian method that uses a logistic regression model to estimate
293 directly the posterior probability that a given locus is under selection. We used default
294 parameter values in our analyses to detect outliers among the clusters previously
295 identified. The second approach (Duforet-Frebourg et al. 2014) implemented in the R
296 package PCAdapt (Luu et al. 2016) is based on a hierarchical model where population
297 structure is first depicted using K factors. No *a priori* on the genetic structure (and thus,
298 no clustering) is required in advance. Loci that are atypically related to population
299 structure, measured by the K factors, are identified as outliers. For each value of K
300 (ranging from one to ten), ten replicate chains of 150,000 Monte Carlo Markov Chain
301 iterations were performed and we discarded the first 5,000 iterations as burn-in. The third
302 approach uses the estimated co-ancestry matrix to compute an extension of the original
303 Lewontin and Krakaeur statistic that account for the history of populations under a model
304 of pure drift (Bonhomme et al. 2010). Finally we used the custom simulation test
305 described in Fraïsse et al. (2014). The idea of this test is to use simulations of the best-
306 supported demographic model to obtain the neutral envelope of the joint distribution of
307 pairwise F_{ST} in a four-population analysis. This test uses the fact that it is easy to have
308 false positives in each of two pairwise comparisons but that outliers in both comparisons,
309 against the genome-wide structure, are more likely to be true positives. Roux et al. (2016)
310 found that an Isolation-with-Migration (IM) model fitted the seahorse data well and that
311 more parameterized models did not improve the fit. These authors also found that the
312 time of divergence (Tsplit) was very similar in each pairwise comparison. We therefore
313 used a four-population IM model with the population sizes and migration rates inferred
314 by Roux et al. (2016) on the transcriptome data of Romiguier et al. (2014). The two latter

315 tests are very similar in their spirit, although one use a simple history of divergence with
316 drift and the other intend to explore more complex demographic histories. They were
317 here used to identify parallel SNPs that provide a grouping of populations that goes
318 against the genome-wide trend.
319
320 Genotype discordance among loci in admixed samples (*i.e.* genomic cline
321 analysis) was tested using Barton's concordance method as described in Macholán et al.
322 (2011). The method fits a quadratic function to the relationship between a single locus
323 hybrid index (0 or 1 if homozygous, 0.5 if heterozygous) and the genome-wide hybrid
324 index. The function parameters measure the deviation from x = y (*i.e.* the expectation of
325 homogeneous genome-wide introgression) as a function of the expected heterozygosity.
326 Instead of testing the deviation from the diagonal, our aim was to compare genomic
327 clines (*i.e.* regression curves) between geographic samples, as discordance was observed
328 in one population and not others.
329 Finally, we tested for gene ontology (GO) terms enrichment to determine if
330 outlier loci displayed functional enrichment, compared to the full dataset, using Fisher's
331 Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg 1995)
332 implemented in the software Blast2GO (Conesa and Götz 2008).
333
334 **Results**
335 *SNPs characterization/calling: an efficient genotyping method in a protected species with*
336 *low genetic diversity*
337 Of the 384 SNPs, including five SNPs that diagnostically distinguished *H.*
338 *guttulatus* from *H. hippocampus* (one mitochondrial and four nuclear), a total of 318
339 SNPs amplified successfully. Of these 318 SNPs, 286 showed a minor allele frequency
340 above 5%. Two *H. hippocampus* initially identified as *H. guttulatus* were removed from
341 the initial dataset, resulting in a final dataset of 292 *H. guttulatus* genotyped for 286
342 SNPs.
343 In order to evaluate the extent of ascertainment bias that may be induced by our
344 procedure for selecting markers, we compared the Joint Site-Frequency Spectra (JSFS)
345 obtained with the original dataset of Romiguier et al. (2014) -assumed freed from
346 ascertainment bias- with our 314-SNP dataset. JSFSs are detailed in Supporting
347 Information 1 (Fig. SI1). Briefly, as expected, the data of Romiguier et al. (2014) is
348 principally composed of singletons that represented 35-40% of the SNPs (Fig. SI1A),
349 while this proportion was reduced twofold in our 314-SNP dataset (Fig. SI1B). We
350 efficiently removed rare variants without biasing the frequency spectrum too much as the
351 deficit of singletons was homogeneously compensated by every other cells of the JSFSs.
352 An even representation over the entire allele frequency range was indeed observed based
353 on our dataset. The comparison of the two JSFSs (Fig. SI1C) reveals that very few cells
354 apart from singletons have an excess above 5%, suggesting limited ascertainment bias in
355 our SNP panel, except with rare alleles as expected.
356 With a limited ascertainment bias, with rare alleles being underrepresented, no
357 missing data, and constraints on our model study (small amount of DNA available with
358 the use of non-lethal fin-clipping sampling techniques), selecting SNPs characterized
359 from a preliminary population transcriptomic survey proved to be a straightforward

strategy for genome-wide investigation of the spatial distribution in this species compared to classical genotyping-by-sequencing approaches.

A strong genetic structure delineating five broadly distributed panmictic genetic clusters

Estimates of the mean number of alleles per locus (N_{all}), allelic richness (A_r), expected heterozygosity (H_e), and departure from Hardy-Weinberg equilibrium (HWE; F_{IS}), for each study site and each genetic cluster identified, are 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 hybrid population, see below).

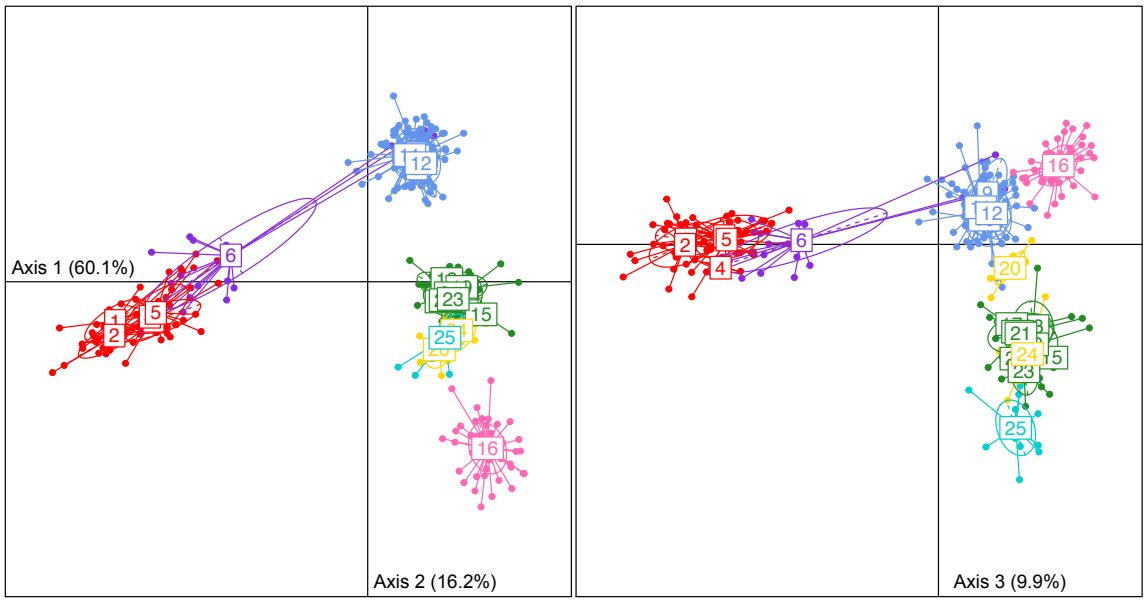


Figure 2 Genetic population structure based on 286 SNP markers analyzed by Principal Component Analyses depicting axis 1 (explaining 60.1% of the variance) and axis 2 (explaining 16.2% of the variance; left panel) and axes 1 and 3 (explaining 9.9% of the variance; right panel). Samples are numbered and colored according to Fig. 1. Each label shows the barycenter of each study site.

The Principal Component Analysis (PCA) revealed clear differentiation separating five clusters (Fig. 2). Along the first two axes (left panel), explaining 60.1% and 16.2% of the variance, the PCA showed clear differentiation between North Atlantic (sites 1-5, in red), South Atlantic (sites 7-12, in blue), Mediterranean Sea (sites 13-25, 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 out along the third axis (9.9% of the variance explained, right panel), with Bizerte (site 20, in gold) being closer to the Mediterranean Thau lagoon on the one hand, and Varna (Black Sea, site 25, in turquoise) standing out from the Mediterranean group on the other hand.

This pattern of genetic structure was also detected by the STRUCTURE analysis, of which the output composed of five clusters is presented in Supporting Information S3. By selecting a K-value of five, we distinguished identical clusters composed by: 1- the North Atlantic sites, 2- the South Atlantic sites, 3- the Mediterranean Sea sites, 4- the Mediterranean Thau lagoon and 5- Varna. Similar to the PCA (Fig. 2), a gradient of introgression is visible along the Atlantic coasts, with a decreasing proportion of South Atlantic cluster ancestry North of Hossegor (site 6). While most Hossegor individuals belong to the North Atlantic genetic background, three individuals were genetically assigned to the South Atlantic genetic background. Bizerte (site 20), and to a lesser extent Halkida (site 24), proved to have mixed ancestries from both Mediterranean lagoon and Sea clusters. Increasing the K-value resulted in additional clusters composed by Bizerte (site 20, K=6), Halkida (site 23, K=7), but then (K>7) no individuals were completely assigned to any new cluster.

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. SI3B). 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 exact tests revealed significant differences in allelic frequencies among clusters (*p*-value < 0.001), with significant differentiation being observed for all pairwise comparisons (F_{ST} ranging from 0.024 between the Mediterranean Sea cluster and Halkida to 0.260 between the North Atlantic cluster and Varna).

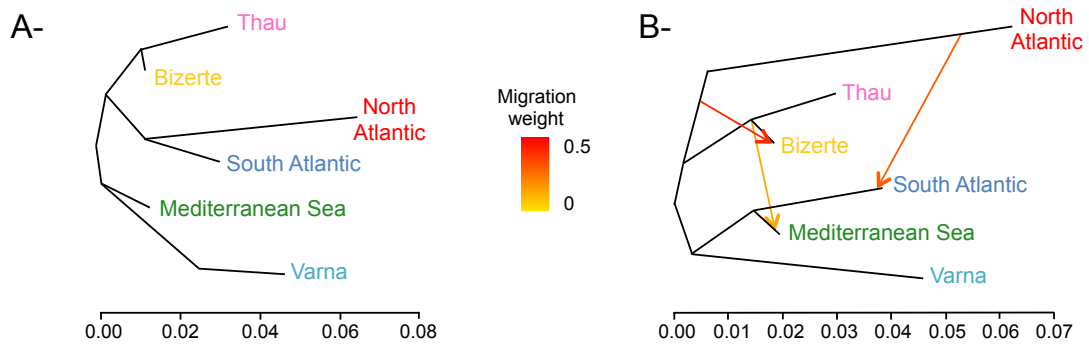
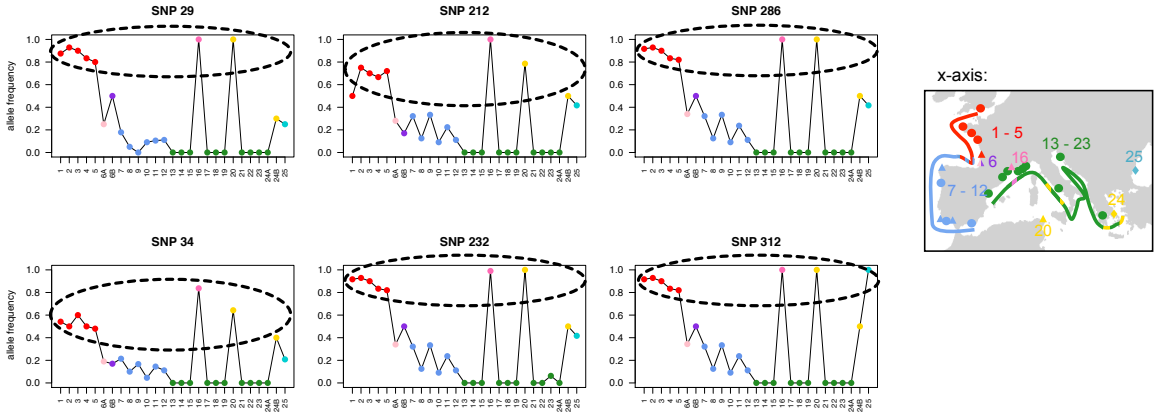


Figure 3 Population trees inferred by TREEMIX (A) without or (B) with 3 migration events. Admixture arrows are colored according to the migration weight. The model including three admixture events significantly improved the fit as compared to a situation without migration (*p*-value < 0.001). Panmictic 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 on the other hand. Interestingly, three admixture events significantly improved the model as compared to a scenario without migration (*p*-value < 0.001; Fig. 3B). This population tree indicated significant gene flow among four *H. guttulatus* clusters, between the north

429 and the south genetic clusters in the Atlantic coasts, in concordance with the gradient of
430 introgression along the Atlantic coasts (Fig. 2 and SI3), between marine and lagoon
431 samples in the Mediterranean Sea, and finally between the North Atlantic and
432 Mediterranean lagoon samples. We suggest that this latter observation is most probably
433 mainly driven by parallel outlier loci (see below), either as a result of adaptive
434 introgression or as a consequence of a shared history of divergence retained at outlier loci
435 against massive secondary gene flow.
436
437 *Signature of selection and genetic parallelism*
438 Nine SNPs out of 286 (3.15%) were consistently identified to depart from
439 neutrality with the four tests (BayeScan, PCAdapt, FLK and custom simulation test).
440 Interestingly, six of them showed very similar allele frequency patterns, distinguishing
441 North Atlantic sites, Mediterranean Thau lagoon, Bizerte (site 20) and, in a lesser extent
442 Halkida (site 23) and Varna (site 25) from South Atlantic and Mediterranean Sea sites
443 (Fig. 4), and pointing out genetic parallelism (*i.e.* convergence of allele frequency
444 patterns) between these lineages.
445



446
447 **Figure 4** *H. guttulatus* allele frequencies (y-axis) for the six outliers with very similar
448 high allele frequency. Each study site (x-axis) is labeled and colored according to Fig. 1,
449 reminded by a simplified map on the right. Hossegor was separated in 6A and 6B and Halkida in
450 24A and 24B according to their North or South Atlantic / Mediterranean Thau lagoon or Sea
451 genetic background, respectively (see Fig. 2, SI3).
452
453 These six outliers located on different *H. guttulatus* contigs were located on different
454 *Hippocampus comes* scaffolds, except SNPs 29 and 286 mapping to a unique *H. comes*
455 scaffold (Fig. 5A). Interestingly, these scaffolds –that contain outliers- consistently
456 mapped to a unique chromosome in *Syngnathus scovelli* (Fig. 5A). Results were similar
457 when directly blasting these six *H. guttulatus* contigs against *S. scovelli* genome, but with
458 SNP 29 mapping to an unplaced scaffold (Fig. 5A). A unique chromosome is still
459 involved when blatting *H. guttulatus* outlying contigs against seven additional well-
460 assembled fish genomes, in agreement with a well-conserved synteny of fishes (detailed
461 in SI5).
462

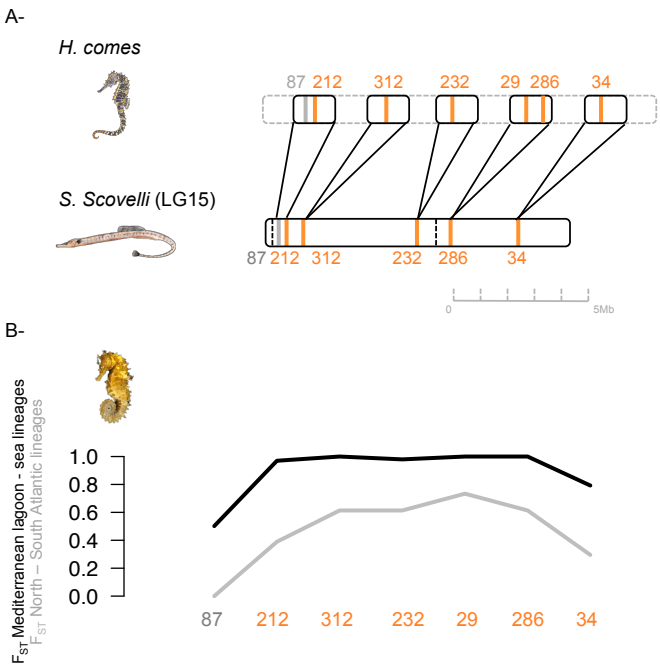


Figure 5 Outlier cross-mapping (A-) and F_{ST} between the Mediterranean lagoon and Sea lineages (in black), and between the North and South Atlantic lineages (in grey) along the unique chromosome (B-). As *H. comes* scaffolds are unplaced, *H. guttulatus* SNPs were first blasted on *H. comes* scaffolds, 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.

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. This F_{ST} - F_{ST} co-plot also showed how the parallel pattern of differentiation observed at these six loci would be highly unlikely under the inferred demographic history of populations.

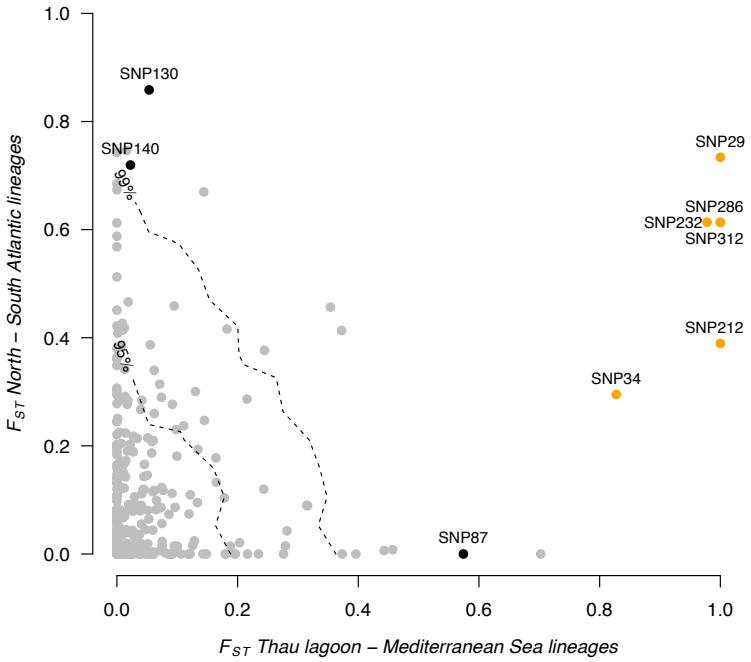


Figure 6 Genome scan of infra-specific differentiation in *H. guttulatus*. Dashed lines represent the 95% and 99% quantiles of the neutral envelope of F_{ST} obtained following Fraïsse et al. (2014). Loci identified by all methods as outliers are colored in orange – the six outliers that displayed parallel 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. 6 and Supporting Information SI4) 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 chromosome.

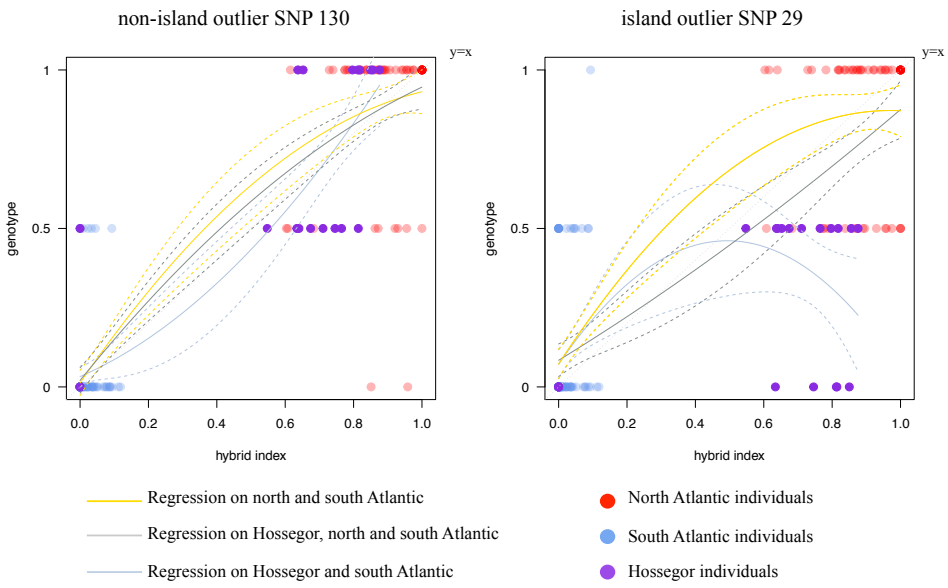


Figure 7 Genomic cline plots for two illustrative markers showing extreme level of differentiation (SNPs 29 and 130). Genomic clines were conducted on North and South Atlantic lineages following polynomial regressions on all Atlantic individuals without Hossegor (in yellow), on all Atlantic individuals (in grey) and on all Atlantic individuals, without North Atlantic (in blue). Dotted lines represent the 95% confidence intervals. Circles indicate the raw genotype data (ancestral homozygotes on the top line –the "1" genotype, heterozygotes in the middle –the "0.5" genotype, and derived homozygotes on the bottom line –the "0" genotype).

Figure 7 shows the genomic cline analysis obtained with SNP 29, an outlier that clustered on the unique chromosome (Fig. 5 and SI5), and SNP 130, an outlier between the northern and the southern Atlantic lineage but that mapped to another chromosome and 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. Introgression did not occur at the same rate for the loci mapping to the unique chromosome, with a clear discordance between the clustered outlier loci and the rest of the genome in the Hossegor sample, in which North Atlantic-like seahorses have the South Atlantic allele at the genomic island outliers.

Note that of all the 168 fish sexed (out of 292 fishes), comprising individuals from each of the five lineage and contact zones in a balanced sex ratio, loci with extreme level of differentiation could not be related to any sex differences. 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 and Vincent 2006), each lineage showed remarkable genetic homogeneity over very large distances, with genetic panmixia observed within each lineage. This spatial structure, with strong and sharp genetic subdivisions, is not expected if genetic drift was predominantly responsible for the genetic differentiation between these populations.

The spatial organization of the different genetic backgrounds proved to be an unusual combination of geographic subdivision in the Atlantic Ocean, and genetic structure related to the sea-lagoon ecological contrast in the Mediterranean Sea. Surprisingly, these two subdivisions partly relied on the same genetic architecture. We observed genetic parallelism at some markers showing extreme levels of differentiation between habitats in the Mediterranean Sea, but also between geographic lineages in the Atlantic Ocean. Intriguingly, all the loci showing convergent allele frequency patterns mapped to a single chromosome. This tends either to suggest genetic differences do not accumulate evenly across the genome, or that these genomic regions are easier to characterize with low-density genome scans.

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

1- Genome scans in hybrid zones

Our SNP panel allowed us to define five discrete panmictic genetic clusters, two in the Atlantic Ocean, two in the Mediterranean Sea, and one in the Black Sea, which cannot be morphologically distinguished with reliability so far (*i.e.* cryptic lineages). The average genetic differentiation between clusters and its associated variance were strong ($0.09 \pm 0.02 < F_{ST} \pm \text{sd} < 0.26 \pm 0.04$). In this context, identifying outlier loci is a complex task with problems of false positives (Lotterhos and Whitlock 2014, 2015). It is also increasingly being acknowledged that a signal of local adaptation is not easily captured without a broad sampling of the genome in a standard infra-specific low-linkage disequilibrium context (Hoban et al. 2016). Finding easily nine well-supported outlier loci with a limited number of loci suggests strong variance in differentiation levels associated with the existence of cryptic genetic barriers involving many selected loci (Bierne et al. 2011). Extensive linkage disequilibrium is also maintained in this complex of semi-isolated genetic backgrounds when compared to a standard infra-specific context. However, this is not expected to be a rare situation as genomic studies have provided accumulating evidence that semi-permeable barriers to gene flow are widespread (Roux et al. 2016) and affect a substantial proportion of genomes (Harrison and Larson 2016).

In this study, six of the nine outliers showed a signature of genetic parallelism (Fig. 2, Fig. SI3). These six outliers not only proved to be the most differentiated loci using a pairwise comparison between Northern and Southern lineages in the Atlantic Ocean and between the lagoon and marine ecotypes in the Mediterranean Sea (Fig. 6), but they also displayed a genetic structure that is not compatible with the sample covariance matrix inferred with the full set of loci (FLK test, Supporting Information SI6) or with simulations under the best-supported demographic model inferred from transcriptome data (Roux et al. 2016). Finally, these six outliers (3.15% of the SNPs

586 analyzed) proved to map to the same chromosome (4.5% of the genome). These six loci
587 mapped to the same chromosome in available fish genomes, although the order on the
588 chromosome was not so well conserved in distantly related species. This result suggests
589 the existence of a large genomic island of differentiation as already reported in other
590 fishes (e.g. sticklebacks: Jones et al. 2012, cod: Hemmer-Hansen et al. 2013, Berg et al.
591 2017, seabass: Duranton et al. 2018), and possibly involving a chromosomal inversion or
592 another form of recombination suppression. This chromosome-wide island of
593 differentiation also explains why we detected the signal of parallel outlying
594 differentiation with a moderate number of loci analyzed. The chromosome could have
595 been easily missed with a few microsatellites but our SNP panel was largely enough to
596 have a few SNPs on the chromosome. These preliminary results call for genome
597 sequence analysis in order to better characterize the genomic island and its structural
598 variation, and to investigate the genome localization of additional barrier loci with a
599 smaller chromosomal footprint.
600
601 *2- Parapatrically distributed semi-isolated cryptic lineages in the Atlantic Ocean*
602 In the Atlantic Ocean we confirmed the existence of two well-differentiated
603 genetic clusters previously identified by Woodall et al. (2015) with five microsatellites
604 and two mitochondrial genes. A genetic homogeneity was observed over large distances
605 within each lineage (Fig. 1), contrasting with a very strong and abrupt genetic
606 differentiation between them. Inter-lineage genetic divergence was not only captured by
607 the most differentiated SNP that were nearly fixed between the two lineages (SNP 130 in
608 SI4, F_{ST} = 0.9), but also, by many other SNPs (SI3B), suggesting isolating mechanisms
609 and loci with genome-wide effects. This mechanism is also suggested by the discovery of
610 two coexisting lineages in the same lagoon in South West of France without early
611 generation hybrids sampled (Hossegor lagoon, site 6 in Fig. 1). Such zone of co-existence
612 in sympatry contrasting with genetic homogeneity at a large spatial scale within each
613 lineage suggests strong reproductive isolation (Jiggins and Mallet 2000), involving either
614 pre-zygotic isolation or strong selection against hybrids at early life stages.
615 Recent migration of the two diverged genetic backgrounds in this lagoon does not
616 easily explain the co-existence of the two lineages with no hybridization. In addition to
617 the low probability of sampling the first-generation migrants without any genomic
618 modifications, local introgression of the North-Atlantic seahorse in this lagoon (Fig. 3)
619 attests hybridization occurred and that a mechanism of isolation must maintain the
620 divergence. The decreasing proportion of south lineage ancestry from Hossegor to the
621 English Channel also provides indirect evidence for asymmetric introgression (Fig. 2,
622 SI3A). In addition, an interesting pattern was observed at genomic island outlier loci in
623 Hossegor (site 6) that further support partial reproductive isolation. The North Atlantic
624 lineage mostly carries the southern alleles at genomic island loci (pale pink dots in Fig. 4,
625 see also SI3B). To produce such North Atlantic-type individuals with a South-Atlantic
626 island, recombination between parental backgrounds is required, suggesting hybridization
627 occurred in this case. In the genomic cline analyses (Fig. 7) the geographic information
628 was included to contrast the results obtained with or without the Hossegor sample. The
629 objective of this analysis was to infer the existence of a local discordance in this site. The
630 swamping of North Atlantic fishes by the southern allele at the genomic island suggests
631 epistatic or coupling interactions among loci implied in reproductive isolation.

632 Endogenous post-zygotic selection against hybrids does not result in a stable
633 polymorphism in a single isolated population, but instead in transient polymorphisms.
634 Also known as bistable variants, the maintenance of underdominant or epistatically
635 interacting genetic incompatibilities requires a migration-selection balance in a spatially
636 subdivided population (Barton and Turelli 2011), or frequency-dependent selection
637 (Barton and de Cara 2009). The system may have fixed one state of a bistable variant in
638 the Hossegor lagoon. Alternatively, the southern allele may confer a selective advantage
639 in this lagoon whatever the genetic background. However the latter hypothesis does not
640 seem coherent as the northern allele is always in higher frequency in lagoon-like habitats
641 elsewhere, though we may have missed more subtle ecological parameters that would
642 have driven this pattern.
643 Overall, the Atlantic contact zone possesses all the characteristics of standard
644 clinal hybrid zones that follow the tension zone model (Barton and Hewitt 1985), *i.e.* a
645 secondary contact zone maintained by a balance between migration and intrinsic
646 reproductive isolation. Exogenous selection may also contribute, although the two
647 Atlantic lineages both inhabit indifferently lagoon and sea habitats (Fig. 1), and co-exist
648 in syntopy in Hossegor, suggesting a limited role of ecological contrast in their genetics.
649 In any case, only strong intrinsic reproductive isolation can guarantee a genome-wide
650 barrier to gene flow explaining the co-existence of the two lineages in sympatry in the
651 Hossegor lagoon.
652
653 *3- Sea and lagoon ecotypes in the Mediterranean Sea*
654 Contrasting with the Atlantic hybrid zone, the two cryptic lineages identified in
655 the Mediterranean Sea were associated with lagoon/sea ecosystem variation. Our broad
656 genomic and spatial sampling revealed two crucial observations. First, while the marine
657 lineage was surprisingly homogeneous over the whole Mediterranean Sea, from Greece
658 to Spain (F_{IT} = 0.0078 n.s.), lagoon-like samples, especially the Thau lagoon, showed a
659 strong and genome-wide genetic differentiation from them. Samples from two lagoons
660 (Thau in France and Bizerte in Tunisia, sites 16 and 20 in Fig. 1) were sufficient to
661 reveal an association with the environment that was previously unseen, the Thau lagoon
662 being the only sample from Western Mediterranean basin in Woodall et al. (2015).
663 Second, fixed differences between lagoon and marine samples were observed, although
664 they were sampled only few kilometers apart (e.g. sites 14-16 in Fig. 2, 4). A single but
665 important seahorse sampled on the seaside of the Thau lagoon (site 15 in Fig. 1), plus
666 seven others sampled on the seaside of another lagoon of the region (site 14 in Fig. 1)
667 proved to belong to the marine genetic cluster, without any sign of introgression.
668 Likewise, no evidence of introgression was observed in the Thau sample (Fig. 3). Once
669 again, despite genetic homogeneity over large area, such strong and abrupt genetic
670 differentiation suggests partial reproductive isolation between these two lineages. In this
671 case there are obvious ecological drivers, *i.e.* habitat specialization. Indeed, the Bizerte
672 lagoon (site 20 in Fig. 1), which is ecologically similar to the Thau lagoon (Sakka Hlaili
673 et al. 2008), has a population genetically similar to marine Mediterranean samples at
674 most loci and only share the genetic composition of the Thau lagoon at the genomic
675 island (Fig. 4 and SI3B). In addition, a subsample of the Halkida population (Greece, site
676 24B in Fig. 4) was composed of five individuals heterozygous at the genomic island. The
677 environmental parameters at this location are hypothesized to be more lagoon-like, being

a secluded bay beyond the northern end of the Euipus Strait. This sample only provides evidence that the genomic island polymorphism and the mosaic spatial structure extends to the eastern basin without really providing further clues about the role of the environment.

Our results from the Mediterranean Sea (*i.e.* lagoon/marine system) resemble those obtained from the emblematic three-spined sticklebacks marine/freshwater system (Jones et al. 2012), or more recently from the marine-migratory/freshwater-resident lampreys (Rougemont et al. 2016) and coastal/marine ecotypes in European anchovies (Le Moan et al. 2016). When a shared divergent genomic island is observed –*i.e.* genetic parallelism, as here in long-snouted seahorse among the Mediterranean lagoons, or in the examples cited above, there could be three possible interpretations: (i) parallel gene reuse from a shared ancestral polymorphism present in the marine supposedly ancestral population (Jones et al. 2012), (ii) the spread of a locally adapted allele (*i.e.* the 'transporter hypothesis'; Schluter and Conte 2009) or (iii) secondary contact followed by spatial reassortment of the divergent lineages and extensive introgression swamping such that only selected loci and their chromosomal neighborhood retain the history of adaptation (Bierne et al. 2013). The three scenarios are difficult to discriminate as they converge toward a similar pattern (Johannesson et al. 2010, Bierne et al. 2013, Welch and Jiggins 2014). Here, as for the lampreys (Rougemont et al. 2016), the Thau lagoon provides a possible support for the secondary contact model because the differentiation, although stronger at the genomic island, is genome-wide. The Bizerte lagoon however can either be interpreted as a marine lineage introgressed by the lagoon allele at the genomic island, or as a lagoon lineage (defined by adaptive/speciation genes) massively introgressed by neutral marine alleles. Incorporating heterogeneous migration rates in demographic inference methods allowed Le Moan et al. (2016), Rougeux et al. (2016) and Rougemont et al. (2016) to identify the signal of a secondary contact history carried by islands of differentiation in lampreys, white fishes and anchovies. Unfortunately our 286-SNPs dataset does not allow performing such historical demographic reconstruction. Nonetheless the TREEMIX analysis reveals that episodes of secondary admixtures strongly improve the fit to the sample covariance matrix, but adaptive introgression or massive introgression swamping can both explain them. Anyhow, demographic reconstruction does not completely refute the 'transporter hypothesis' which stipulates lagoon alleles spread from lagoon to lagoon (or freshwater allele from river to river) and is a scenario that produces a very similar genomic pattern of differentiation to the one produced by a standard secondary contact (Bierne et al. 2013, Rougemont et al. 2016). In the case of the seahorse complex, however, we made the new observation that genetic parallelism is observed with the Atlantic populations where the structure is geographic and independent of the lagoon-sea habitats, which offers complementary arguments to the debate.

4- Genetic parallelism in two different spatial/ecological contexts

The most astonishing result of our genetic analysis was genetic parallelism between the Mediterranean lagoon ecotype and the north Atlantic lineage at a large genomic island. Parallel evolution implies distinct but repetitive ecological characteristics (e.g. Butlin et al. 2014). However, this claim comes from an inductive reasoning, validated by many observations, which a single counter-example can yet alone refute. In

724 the present study, we found that the genomic island was associated with the sea-lagoon
725 ecological contrast in the Mediterranean Sea, while there was no such genetic
726 differentiation between lagoon and sea samples in the Atlantic Ocean. The North and
727 South lineages inhabit indifferently lagoons and seas, so that what seemed obvious in the
728 Mediterranean Sea regarding the divergence of the two lineages, *i.e.* habitat
729 specialization, was not observed along the Atlantic Ocean where the divergence seems
730 uncorrelated to the ecological contrasts that explain the two Mediterranean ecotypes.
731 Besides, abiotic parameters such as temperature or salinity, did not allow bounding the
732 North Atlantic and Mediterranean lagoon lineages genetic parallelism. Although we may
733 have missed putative ecological drivers of such genetic parallelism, parallel gene reuse
734 driven by ecological convergence seems here unlikely. Shared ancestral polymorphism
735 sieved by adaptation in a patchy environment (Bierne et al. 2013) and incipient speciation
736 (Guerrero and Hahn 2017) would better explain our data. The association with habitat in
737 the Mediterranean Sea and with space in the Atlantic Ocean could be explained by a
738 secondary evolution of locally adapted genes within the genomic island in the
739 Mediterranean Sea, benefiting from the barrier to gene flow imposed by intrinsic
740 selection (divergence hitchhiking; Via 2012, Yeaman 2013). Alternatively, the genomic
741 island could have coupled with local adaptation polymorphisms localized elsewhere in
742 the genome in the Mediterranean Sea (Bierne et al. 2011), while it would have been
743 trapped by a barrier to dispersal in the Atlantic Ocean (Barton 1979, Barton and Hewitt
744 1985). Without further data and the true genomic position of loci in the seahorse genome,
745 it is difficult to disentangle the two hypotheses.
746 One locus (SNP 87) localized in the same chromosome as parallel SNPs (Fig. 5)
747 is differentiated between the marine and lagoon ecotypes in the Mediterranean Sea, but is
748 not differentiated between the northern and southern lineages in the Atlantic Ocean. At
749 first sight, it could be interpreted as evidence for a possible secondary local sweep in
750 Mediterranean lagoons. However, according to the gene order inferred from the closest
751 species (Gulf pipefish and Tiger tail seahorse) along with a barrier to gene flow less
752 effective in the Atlantic (see SNPs 34 and 212 in Fig. 4 and 5B), it could also be
753 interpreted as being localized in the island "shoulders" (*i.e.* loci in the vicinity of the
754 regions harboring local adaptation and/or reproductive isolation loci; Gagnaire et al.
755 2015, Le Moan et al. 2016) in which a stronger introgression rate has erased the
756 differentiation faster in the Atlantic than in the Mediterranean, in favor of the alternative
757 interpretation. Importantly, whatever the explanation -divergence hitchhiking or
758 coupling- it requires invoking interaction between intrinsic and ecological selection and
759 not ecological selection alone (Bierne et al. 2011, Kulmuni and Westram 2017).
760 Furthermore, intrinsic selection has most probably evolved first in this system as no
761 genetic parallelism was observed in outliers discriminating lagoon to marine ecotypes,
762 which would contradict the predominant view that ecological selection is necessarily the
763 initial catalyzer of the chain of accumulation of barriers in ecological speciation.
764
765 **Conclusions**
766 Analyzing the population genetics of the long-snouted seahorse *Hippocampus guttulatus*
767 revealed a complex of panmictic genetic backgrounds subdivided by sharp semi-
768 permeable hybrid zones. This is now standard observation in marine species (Knowlton
769 1993, Pante et al. 2015, Sheets et al. 2018) where morphological stasis might be more

widespread than in terrestrial species. The subdivision of species by hybrid zones is a long-lasting observation in the terrestrial realm (Hewitt 1989) but arguably more readily detected by morphological differences. We also easily found outlier loci despite a moderate number of loci analyzed, and the clustering of these outlier loci in a single genomic region. This also tends to become a standard observation of the recent hybrid zone literature (e.g. in stickelbacks, Jones et al. 2012; jaera, Ribardière et al. 2017; cod, Hemmer-Hansen et al. 2013).

However, we also made two additional observations that are less commonly reported and deserve a broader interest outside the study of seahorse themselves, although are important to further inform captive breeding and *in-situ* management decisions. First, we found the two, usually opposed, standard spatial structures of the hybrid zone literature, namely 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.

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 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.

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.

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.

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

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

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

and F. Di Palma. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.

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

Butlin, R. K., M. Saura, G. Charrier, B. Jackson, C. André, A. Caballero, J. A. Coyne, J. Galindo, J. W. Grahame, J. Hollander, P. Kemppainen, M. Martínez-Fernández, M. Panova, H. Quesada, K. Johannesson, and E. Rolán-Alvarez. 2014. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. Evolution 68:935–949.

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

Conesa, A., and S. Götz. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics 2008.

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

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

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

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

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.

Gayral, P., J. Melo-Ferreira, S. Glémin, N. Bierne, M. Carneiro, B. Nabholz, J. M. Lourenco, P. C. Alves, M. Ballenghien, N. Faivre, K. Belkhir, V. Cahais, E. Loire, A. Bernard, and N. Galtier. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. PLOS Genet. 9:e1003457.

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

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

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

Harrison, R. G. 1993. Hybrid zones and the evolutionary process. Oxford University Press.

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

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

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

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

Hewitt, G. M. 1989. The subdivision of species by hybrid zones. *Speciation and its Consequences*, 85-110.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, L. K. Reed, A. Storfer, and M. C. Whitlock. 2016. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am. Nat. 188:379–397.

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

1042 Keenan, K., P. McGinnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl. 2013.
1043 diveRsity: An R package for the estimation and exploration of population genetics
1044 parameters and their associated errors. Methods Ecol. Evol. 4:782–788.
1045
1046 Knowlton, N. 1993. Sibling species in the sea. Annu. Rev. Ecol. Syst. 24:189–216.
1047
1048 Kulmuni, J., and A. M. Westram. 2017. Intrinsic incompatibilities evolving as a by-
1049 product of divergent ecological selection: Considering them in empirical studies on
1050 divergence with gene flow. Mol. Ecol. 26:3093-3103.
1051
1052 Larson, E. L., J. A. Andrés, S. M. Bogdanowicz, and R. G. Harrison. 2013. Differential
1053 introgression in a mosaic hybrid zone reveals candidate barrier genes. Evol. Int. J. Org.
1054 Evol. 67:3653–3661.
1055
1056 Larson, E. L., T. A. White, C. L. Ross, and R. G. Harrison. 2014. Gene flow and the
1057 maintenance of species boundaries. Mol. Ecol. 23:1668–1678.
1058
1059 Le Moan, A., P.-A. Gagnaire, and F. Bonhomme. 2016. Parallel genetic divergence
1060 among coastal-marine ecotype pairs of European anchovy explained by differential
1061 introgression after secondary contact. Mol. Ecol. 25:3187-3202.
1062
1063 Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-
1064 Wheeler Transform. Bioinformatics. 25:1754-60.
1065
1066 Lin, Q., S. Fan, Y. Zhang, M. Xu, H. Zhang, Y. Yang, A. P. Lee, J. M. Woltering, V.
1067 Ravi, H. M. Gunter, W. Luo, Z. Gao, Z. W. Lim, G. Qin, R. F. Schneider, X. Wang, P.
1068 Xiong, G. Li, K. Wang, J. Min, C. Zhang, Y. Qiu, J. Bai, W. He, C. Bian, X. Zhang, D.
1069 Shan, H. Qu, Y. Sun, Q. Gao, L. Huang, Q. Shi, A. Meyer, and B. Venkatesh. 2016. The
1070 seahorse genome and the evolution of its specialized morphology. Nature 540:395–399.
1071
1072 López, A., M. Vera, M. Planas, and C. Bouza. 2015. Conservation genetics of threatened
1073 *Hippocampus guttulatus* in vulnerable habitats in NW Spain: temporal and spatial
1074 stability of wild populations with flexible polygamous mating system in captivity. PLoS
1075 ONE 10:e0117538.
1076
1077 Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and
1078 neutral parameterization on the performance of FST outlier tests. Mol. Ecol. 23:2178–
1079 2192.
1080
1081 Lotterhos, K. E., and M. C. Whitlock. 2015. The relative power of genome scans to
1082 detect local adaptation depends on sampling design and statistical method. Mol. Ecol.
1083 24:1031–1046.
1084
1085 Lourie, S. A., and A. C. J. Vincent. 2004. Using biogeography to help set priorities in
1086 marine conservation. Conserv. Biol. 18:1004–1020.
1087

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.

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

Pante, E., N. Puillandre, A. Viricel, S. Arnaud-Haond, D. Aurelle, M. Castelin, A. Chenuil, C. Destombe, D. Forcioli, M. Valero, F. Viard, and S. Samadi. 2015. Species are hypotheses: avoid connectivity assessments based on pillars of sand. Mol. Ecol. 24:525–544.

Pérez‐Figueroa, A., M. J. García‐Pereira, M. Saura, E. Rolán‐Alvarez, and A. Caballero. 2010. Comparing three different methods to detect selective loci using dominant markers. J. Evol. Biol. 23:2267–2276.

Pickrell, J. K., and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8:e1002967.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43:432–449.

Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M. a. F. Noor, B. Mehlig, and A. M. Westram. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30:1450–1477.

Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. J. Hered. 86:248–249.

Ribardière, A., C. Daguin‐Thiébaut, C. Houbin, J. Coudret, C. Broudin, O. Timsit, and T. Broquet. 2017. Geographically distinct patterns of reproductive isolation and hybridization in two sympatric species of the *Jaera albifrons* complex (marine isopods). Ecol. Evol. 7:5352–5365.

1133 Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari,
1134 R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G.
1135 Tsagkogeorga, A. a.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, and N.
1136 Galtier. 2014. Comparative population genomics in animals uncovers the determinants of
1137 genetic diversity. Nature 515:261–263.
1138
1139 Rougemont, Q., P.-A. Gagnaire, C. Perrier, C. Genthon, A.-L. Besnard, S. Launey, and
1140 G. Evanno. 2016. Inferring the demographic history underlying parallel genomic
1141 divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Mol. Ecol.
1142 26:142–162.
1143
1144 Rougeux, C., L. Bernatchez, and P.-A. Gagnaire. 2016. Modeling the multiple facets of
1145 speciation-with-gene-flow towards inferring the divergence history of lake whitefish
1146 species pairs (*Coregonus clupeaformis*). bioRxiv 068932.
1147
1148 Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software
1149 for Windows and Linux. Mol. Ecol. Resour. 8:103–106.
1150
1151 Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016.
1152 Shedding light on the grey zone of speciation along a continuum of genomic divergence.
1153 PLOS Biol. 14:e2000234.
1154
1155 Sakka Hlaili, A., B. Grami, N. Niquil, M. Gosselin, D. Hamel, M. Troussellier, and H.
1156 Hadj Mabrouk. 2008. The planktonic food web of the Bizerte lagoon (south-western
1157 Mediterranean) during summer: I. Spatial distribution under different anthropogenic
1158 pressures. Estuar. Coast. Shelf Sci. 78:61–77.
1159
1160 Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad.
1161 Sci. U. S. A. 106 Suppl 1:9955–9962.
1162
1163 Sheets, E. A., P. A. Warner, and S. R. Palumbi. 2018. Accurate population genetic
1164 measurements require cryptic species identification in corals. Coral Reefs 37:549–563.
1165
1166 Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, and İ. Birol. 2009.
1167 ABySS: A parallel assembler for short read sequence data. Genome Res. 19:1117–1123.
1168
1169 Small, C. M., S. Bassham, J. Catchen, A. Amores, A. M. Fuiten, R. S. Brown, A. G.
1170 Jones, and W. A. Cresko. 2016. The genome of the Gulf pipefish enables understanding
1171 of evolutionary innovations. Genome Biol. 17:258.
1172
1173 Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S.
1174 Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and
1175 P. Nosil. 2014. Stick insect genomes reveal natural selection's role in parallel speciation.
1176 Science 344:738–742.
1177

Storey, J. D. 2002. A direct approach to false discovery rates. J. R. Stat. Soc. Ser. B Stat. Methodol. 64:479–498.

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.

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.

Yeaman, S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc. Natl. Acad. Sci. 110:E1743–E1751.

Data archival location

SNPs data have been deposited at DRYAD. DOI: *to be completed after the manuscript is accepted for publication*