Spatial gradients of introgressed ancestry reveal cryptic connectivity patterns in a high gene flow marine fish

Tony Robinet, Valérie Roussel, Karine Cheze, Pierre-Alexandre Gagnaire

Spatial gradients of introgressed ancestry reveal cryptic connectivity patterns in a high gene flow marine fish

Tony Robinet (1)*, Valérie Roussel (2), Karine Cheze ([1]) and Pierre-Alexandre Gagnaire (3)

ABSTRACT

Assessing genetic connectivity among populations in high gene flow species is sometimes insufficient to evaluate demographic connectivity. Genetic differentiation quickly becomes zero as soon as a few dozen migrants are exchanged per generation. This provides little information to determine whether migration can ensure demographic coupling. The resulting difficulties in delineating conservation units for the management of commercially exploited marine fish species are well illustrated in the case of the European sea bass (*Dicentrarchus labrax*). Previous attempts to assess connectivity patterns in the northeast Atlantic have been hampered by a lack of spatial genetic structure. In contrast, mark-recapture data suggested low migration rates between regional spawning areas. Here, we show how a spatial gradient of introgressed Mediterranean ancestry across the northeast Atlantic reflects cryptic

1 Biologie des Organismes et Ecosystèmes Aquatiques (BOREA) ; Muséum National d'Histoire Naturelle, CNRS, IRD, SU, UCN, UA ; Station marine de Concarneau, Quai de la Croix, 29900 Concarneau, France
2 Université de Bretagne Occidentale, Institut GéoArchi EA7462, 6 avenue Le Gorgeu, 29100 Brest, France
3 ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.

* corresponding author : robinet@mnhn.fr

patterns of genetic and demographic connectivity. Using a 1K SNP chip dataset in 827 individuals sampled from Portugal to the North Sea, we found null overall genetic differentiation across the northeast Atlantic. We however detected a subtle latitudinal admixture gradient originating at the edge of the contact zone with the Mediterranean sea bass lineage. Two significant breaks in the ancestry gradient at the tip of Galicia and northern Brittany indicated barriers to effective dispersal between demographically distinct units. Moreover, a northward expansion signal in Irish and North Seas was revealed by the surfing of rare Mediterranean alleles at the edge of the species range. Our results show that introgressed ancestry gradients offer a powerful alternative to assess genetic and demographic connectivity when the neutral migration-drift balance is not informative. (250 words)

Keywords: admixture, expansion, genetic and demographic connectivity, introgression, spatial structure, stock delineation

1 | INTRODUCTION

Assessment of genetic connectivity among populations is widely used in conservation biology and species management to evaluate effective dispersal and the consequences of spatially-dependent evolutionary processes on species persistence, resilience and adaptability (Cayuela et al. 2018; Gagnaire 2020). Using polymorphism data from individuals sampled across the landscape, the genetic approach can detect local barriers to gene flow separating distinct populations that exchange measurable amounts of migrants per generation. Demographic independence between such populations can be assumed when the fraction of exchanged migrants is too small to ensure population persistence in case of negative intrinsic growth rate (Pulliam, 1988; Lowe & Allendorf, 2010). Measuring migratory exchanges using genetic data may, however, be a difficult task.

The power of the population genetics approach depends on the relative intensity of two opposed forces, migration (m) and genetic drift ($1/N_e$). When migration overwhelms the effect of genetic drift (i.e. when the number of migrants exchanged per generation ($N_e m$) exceeds a few dozen), genetic differentiation between populations is close to zero. Null genetic differentiation can result from a variety of equilibrium scenarios ranging from high connectivity among small populations to nearly complete demographic independence among large populations. Consequently, assessing demographic connectivity in species with both large populations and high migration rates has been a long-standing challenge to population genetic approaches (Waples, 1998). Such combinations of biological parameters are frequently encountered in marine species with a dispersive larval stage (Hedgecock et al., 2007). Therefore, a persistent gap between evolutionary and ecological scales often prevents managers to use genetic connectivity information for

fisheries management (Palumbi, 2003; Waples & Gaggiotti, 2006; Waples, Punt & Cope, 2008).

Molecular markers influenced by selection can compensate the lack of signal at neutral loci when the migration-drift equilibrium is not informative. The use of adaptive differentiation signals has been proposed as a solution to improve the delineation of management units (Funk et al., 2012). This approach is now greatly facilitated by the availability of large polymorphism datasets to scan genomes for outlier markers influenced by selection (Stapley et al., 2010; Savolainen et al., 2013; Narum et al., 2013). Several studies in marine species that typically show weak to no genetic differentiation at neutral markers have found stronger signals of spatial structure at different scales using outlier loci (e.g. Bekkevold et al., 2016; Benestan et al., 2016; Van Wyngaarden, 2017). Although these could help in delineating cryptic evolutionary management units, the variation patterns displayed by outlier loci often remain challenging to translate into quantitative assessments of demographic connectivity. One of the main reasons is that observed genetic variation patterns can be generally attributed to a variety of possible selective mechanisms (Bierne et al. 2013), potentially involving complex relationships between local adaptation and gene flow (Tigano & Friesen, 2016). Contextualizing the eco-evolutionary history of the studied species thus helps to discern the mechanisms underlying the patterns of differentiation displayed by outlier loci, toward a better understanding of connectivity (Liggins et al. 2020).

Genetic admixture between differentiated lineages is a particularly informative evolutionary context to learn about connectivity. When two divergent taxa come into secondary contact and exchange genes, the spatial diffusion of foreign genetic material can be used to reveal the population genetic connectivity within each introgressed lineage (Sedghifar et al., 2015; Bertl et al., 2018; Duranton et al., 2019). If

the recipient lineage is not genetically structured, the spatial homogenization of foreign allele frequencies is a quick process. However, local barriers to gene flow may slow down the spread of foreign alleles and generate steps in the admixture and introgression gradients within the recipient lineage (Gagnaire et al., 2015). Although allele frequencies equilibrate quickly at neutral loci following secondary contact, barrier loci involved in partial reproductive isolation have a reduced effective migration rate between lineages and therefore retain their signal of differentiation for longer (Sedghifar et al., 2016). Semi-permeable species boundaries characterized by heterogeneous rates of introgression among loci thus provide favorable conditions to reveal cryptic barriers to gene flow within introgressed lineages, even after several thousands of generations of introgression (Gagnaire et al., 2015). Post-glacial lineages that are nowadays in contact and exchange genes through natural hybridization represent suitable cases for using spatial introgression gradients to study connectivity.

The European sea bass (*Dicentrarchus labrax* L.) is a marine fish species that ranges from northwestern Africa to southern Norway in the Atlantic, and throughout the Mediterranean and Black Seas (Vandeputte et al. 2019). The species is genetically subdivided into two distinct lineages, one Atlantic (ATL) and one Mediterranean (MED), which naturally hybridize at the Atlantic-Mediterranean transition zone (Naciri et al., 1999; Lemaire et al., 2005; Souche et al., 2015). Population genomic studies based on RAD-sequencing (Tine et al., 2014) and whole-genome resequencing (Duranton et al., 2018) have revealed several important aspects of the of the European sea bass evolutionary history: *(i)* genetic divergence between ATL and MED lineages is the result of about 270 kyrs of allopatric divergence, *(ii)* secondary contact has started after the last glacial retreat about 11.5 kyrs ago, and since then *(iii)* gene flow occurs at variable rates across the genome due to partial

reproductive isolation between the two lineages. *(iv)* For a still unknown reason, postglacial gene flow has been more pronounced from the Atlantic into the Mediterranean than in the opposite direction. As a result, contemporary ATL genomes are made of ~5% of MED genetic material, while western and eastern MED genomes contain ~31% and 13% of ATL ancestry, respectively (Duranton et al., 2018). Such levels of introgression are sufficiently high to provide information on genetic connectivity within both sea bass lineages. In the Mediterranean, Duranton et al. (2019) analyzed the neutral decay of introgressed haplotype length as a function of distance from the contact zone to estimate a mean per-generation dispersal distance of 5 to 50 km. This quantitative approach, based on a small number of whole-genome sequences, did not have the spatial resolution required to provide a detailed map of connectivity. Here, we used a different strategy based on moderate genome coverage but extensive spatial sampling to identify local barriers to gene flow within the Atlantic lineage. Our objective was to study both the neutral and non-neutral diffusion of MED alleles across Atlantic sea bass populations to evaluate their fine-scale spatial genetic structure.

Atlantic sea bass populations are of high economic importance to European fisheries, but overfishing probably combined with other factors have led to a decline in landings since 2009-2010, particularly north of the 48th parallel (ICES, 2019). Sea bass stock assessment by scientific authorities (ICES: International Council for the Exploration of the Sea) in the northeastern Atlantic currently relies on four presumably distinct stocks: northern Atlantic, southern Ireland/western Scotland, Biscay and Iberia. However, stock delineation remains poorly understood, making it difficult to properly assess population connectivity and implement effective management programs. Population genetic studies generally found non-significant genetic structure across the northeastern Atlantic (Frisch et al., 2007; Coscia &

Mariani, 2011; Souche et al., 2015), possibly due to insufficient spatial and genomic coverage. In parallel, mark-recapture data indicated restricted individual movement between the Bay of Biscay and the English Channel-Celtic Sea region (Fritsch et al., 2007), a result partly explained by the fidelity to winter spawning areas recently evidenced with tagging data (de Pontual et al., 2019). A small number of effective migrants per generation may account for the absence of genetic differentiation at neutral markers despite high fidelity to spawning grounds. Therefore, it is possible that independent demographic units have remained undetected by previous population genetic studies. Here, we investigate fine-scale connectivity patterns among Atlantic sea bass populations using the spatial diffusion of Mediterranean alleles from southern Portugal toward the northern part of the species range. We document the existence of two significant steps in the Mediterranean ancestry gradient along the Atlantic coast, which most likely delineate demographically independent populations that slightly differ from current management units. We also find molecular signatures of a recent northward expansion, consistent with recently expanding sea bass fisheries in the northern part of the species range.

2 | MATERIALS AND METHODS

2.1 | Sampling

Fin clips from 846 individuals were sampled from southern Portugal to the Irish Sea and North Sea in the period Oct. 2012 to May 2015 (Supplementary Table S1). Samples were mostly collected in fish markets, only from fresh fish with a known geographical origin indicated by its ICES statistical rectangle (precision 0.5° in latitude and 1° in longitude) and a known fishing day. The quality of the spatial

fishing information of each individual was checked with the name of the fishing vessel by tracking its position history on http://www.marinetraffic.com.

Tissues were preserved in 85 % ethanol at -20 °C. Individual genomic DNA was extracted following cellular lysis and proteinase K digestion using the Qiagen DNeasy Blood & Tissue kit, and then conserved in TE. DNA extraction quality was controlled using a NanoDrop spectrophotometer keeping samples with an A260/A280 ratio between 1.8 and 2.0. Concentration of double-stranded DNA was then measured with the Qbit dsDNA Broad Range kit with a standard benchmark (0-100 ng/μl) on a qPCR BioRAD. Individual dsDNA were finally normalized at 25 ng/μl and randomly distributed across nine 96-well plates before genotyping.

2.2 | Genotyping and data quality control

Genotyping was performed by the *Labogena* platform (Jouy-en-Josas, France) using an iSelect Custom Infinium Illumina array specifically developed in the European sea bass (Faggion et al., 2019). The chip contains 1531 validated SNPs covering the whole genome while being homogeneously distributed along the recombination map, and presents no ascertainment bias between Atlantic and Mediterranean sea bass lineages. Individual genotypes and hybridization intensities generated by Illumina's BeadStudio software were reanalyzed using the R package *Argyle* (Morgan et al., 2016) to perform SNP quality control. Markers were filtered based on their rate of no-call genotypes (per locus missingness < 0.1), rate of heterozygosity (Het < 0.55), and minor-allele frequency within the whole Atlantic dataset (MAF > 0.01), in order to include even variants that are frequent or fixed within the Mediterranean, but rare in the Atlantic. Individuals with an excess of missing genotypes were also excluded (per sample missingness < 0.1).

In order to remove a few miscalled markers which variation profile was correlated with samples arrangement in the 96-well plates, we performed a genotype-plate association analysis in *Plink* (Purcell et al., 2007, p-value exclusion threshold=1e-4).

Reference Mediterranean genotypes of 10 MED individuals (four from eastern MED and six from western MED) were merged to our final dataset to compare allele frequencies between ATL and MED populations. These reference genotypes were extracted from high-quality whole-genome resequencing data without any missing genotype at the retained loci (Duranton et al., 2018).

2.3 | Spatial genetic variation on different datasets

The 827 quality-filtered specimens were stratified into three different datasets to reach variable levels of precision in spatial genetics analyses: (*i*) a 'regional dataset' of 827 specimens distributed into 7 regions; (*ii*) a 'main dataset' in which the 827 specimens were assigned to 21 localities corresponding to single ICES rectangles or groups of adjacent rectangles, with a minimum precision of 1.5° in latitude and 3° in longitude, and a number of individuals per locality ranging from 11 to 111 (mean=40, Figure 1, Supplementary Figure S1 and Table S1 for more details); (*iii*) a 'refined dataset', more precise in space, consisting of 761 specimens distributed into 31 localities, with a minimum precision of 30' in latitude and 1° in longitude, and a minimal number of 8 individuals per locality (from 8 to 43, mean = 24, Supplementary Figure S2). The main dataset was used for all analyses, unless stated otherwise.

The proportion of polymorphic markers among the 1012 retained SNPs was calculated for each of the 21 localities. A regression model of the proportion of polymorphic markers as a function of sample size per location was fitted using a

nonlinear least squares analysis with two parameters in order to account for the effect of sample size. We then tested for a linear relationship between the localities' residuals to the fitted model and the latitude of localities. We also tested the existence of a linear correlation between the mean observed heterozygosity per sample locality and latitude. Sample groupings from the regional dataset were then used to test for deviations from expectations under Hardy-Weinberg Equilibrium within each of the seven regions defined above with the *Pegas* R package (Paradis, 2010).

A redundancy analysis (RDA) was performed using the R package *Vegan* (Oksanen et al., 2019) in order to evaluate the extent to which SNP variation is influenced by geographic and temporal factors. We first fitted the model Y (individual genotype) ~ (Latitude + Longitude + Month + Year of capture) and assessed the significance of each explanatory factor using 1000 permutations of genotypic data before refitting the model.

Overall FST was calculated among all of the 21 Atlantic locations using all loci and pairwise FST were then calculated for each pair of sampling locations including the two reference MED populations, as well as for pairs of regions using the 'regional dataset'. We used the *Hierfstat* R package (Goudet, 2005) to calculate FST values and assess their significance using 5000 random permutations of genotypes among localities and among regions. A sequential Bonferroni correction (Hommel, 1989) was applied to empirical p-values with the function *p.adjust()* of the R-base package *Stats*. In order to assess the presence of an isolation-by-distance pattern among ATL locations, we tested the linear correlation between [FST / (1 - FST)] and geographic distance between pairs of localities. We also performed a Mantel test between matrices of pairwise genetic and geographic distances with 10.000 permutations using ADEGENET (Jombart 2008). Pairwise geographic distances were calculated as

the shortest path by the continental plateau between two localities (i.e. bathymetry < 200m) using the R package *Marmap* (Pante & Simon-Bouhet, 2013).

We used the directionality index statistics Ψ (Peter & Slatkin, 2013) to test whether genetic differentiation among pairs of localities was due to isolation-by-distance at equilibrium or to a recent range expansion in the Atlantic distribution area of *D. labrax*. Briefly, the directionality index uses comparisons of two-dimensional site frequency spectra to capture deviations from symmetric migration, such as those generated during range expansions. We used the R package *rangeExpansion* (https://github.com/BenjaminPeter/rangeexpansion) to calculate Ψ for all pairs of Atlantic locations, and used pairwise geographic distances between locations to infer the most likely spatial origin of a possible range expansion.

2.4 | MED introgression in ATL genomes

In order to estimate the fraction of MED alleles present in ATL individuals, we used the program *Admixture* (Alexander, Novembre & Lange, 2009), which infers individual ancestry proportions from K ancestral populations (here K was set to 2). Our unsupervised ancestry inference included 10 MED reference individuals in addition to the 827 ATL samples, so that the two ancestral populations correspond to the ATL and MED sea bass lineages. We set the termination criterion for the optimization algorithm to 100 iterations to ensure convergence and used 1000 bootstrap replicates to estimate the standard error of individual ancestry proportions.

The effects of sampling date and geographic distance to the southernmost Atlantic location (Sines in southern Portugal, code "SINE") on the extent of Mediterranean introgression were tested using a Global Linear Model (GLM, family=Gaussian), with the percentage of MED ancestry as the dependent variable.

The date of capture (using the 'as.date()' class in R) and the distance to the SINE locality (calculated as the shortest path by the continental plateau) were used as explanatory factors.

The effect of geographic distance to the southernmost ATL location (SINE) on the extent of MED ancestry was then tested alone using a linear regression model. The distance to the SINE location was calculated as the shortest path by the continental plateau. In order to test for the existence of potential breaks in the admixture gradient, we then performed a piecewise regression of MED ancestry as a function of distance to SINE. Breaks separating different linear regression models were introduced one by one and, and at each step, every possible break position was examined by calculating the residual standard error of the piecewise regression model. The significance of the reduction in the residual sum of squares of each piecewise regression model compared to the simple regression model was tested using an ANOVA, taking into account the number of additional parameters in the piecewise regression model.

2.5 | Genome scans

We then evaluated the impact of variable rates of introgression between ATL and MED sea bass lineages on allele frequency gradients within the Atlantic. We more specifically tested the theoretical prediction that the genomic regions with reduced rates of introgression between lineages tend to exhibit the strongest spatial allele frequency patterns within the ATL range due to propitious ratios of between- to within-lineage gene flow intensities (Gagnaire et al., 2015). We thus expected that the loci showing the strongest genetic differentiation within the Atlantic would generally also be F_{ST} outliers between ATL and MED lineages. Given the relatively low density of markers in our data set (about 1 SNP per cM), the candidate loci detected by

genomes scans for differentiation are most likely to be indirectly influenced by selection through linkage with a nearby selected locus. A previous quantification indicated that the genomic island regions under the influence of the strongest reproductive isolation barriers between Atlantic and Mediterranean lineages occupy about 4% of the sea bass genome (Duranton et al. 2018). Therefore, our SNP panel is expected to contain at least 40 SNPs in strong linkage with reproductive isolation barriers between the two sea bass lineages, plus variants showing intermediate degrees of linkage with them. To identify those loci, we used two different genome scan methods to detect FST outlier SNPs. Both methods were applied to the analysis of genetic differentiation at two different scales: (*i*) within the ATL range among the 21 localities, and (*ii*) between the ATL and MED sea bass lineages.

The first method used was *Lositan* (Beaumont & Nichols, 1996; Antao et al., 2008), which performs coalescent simulations under the symmetric island model to generate a neutral distribution of FST conditioned on the mean expected heterozygosity among populations. The mean FST calculated across all markers was used as a target value to simulate 1 million SNPs, from which we empirically determined the 99.5th percentile of the FST distribution to identify candidate outlier loci showing an excess of genetic differentiation among locations.

We then used *BayeScan* (Foll & Gaggiotti, 2008), a Bayesian outlier detection method that relies on the multinomial-Dirichlet model. The difference in allele frequency between every sampling location and a common theoretical gene pool was measured by a subpopulation-specific FST coefficient in order to account for differences in effective sizes and migration rates among subpopulations. Selection was introduced by decomposing for each locus the subpopulation-specific FST coefficient into a subpopulation-specific component shared by all loci and a locus-specific component shared by all subpopulations. Deviation from neutrality for a

given locus was assumed when the locus-specific component was needed to explain the observed diversity pattern. The posterior probability that a locus is influenced by selection after accounting for multiple testing was determined by the Posterior Odds (PO), which is the ratio of posterior probabilities between the selection and neutral models. The minimum false discovery rate (FDR) at which a given locus reached significance was determined using the q-value of each locus. A q-value threshold of 0.05 was used for outlier detection in most analyses, corresponding to an FDR threshold of 5%. *BayeScan* chain parameters were set to 5.000 outputted iterations with 5.000 steps for pilot runs and 50.000 burn-in steps.

2.6 | Analysis of introgression gradients around barrier loci

Simulations of secondary contact between two genetically subdivided lineages exhibiting partial reproductive isolation show that neutral markers linked to barrier loci can display substantial allele frequency steps within the introgressed lineage (Gagnaire et al., 2015). These steps appear at the place where local barriers to gene flow slow down the diffusion of introgressed alleles. They are dynamically maintained by differences in flow intensity, and are amplified when gene flow between lineages is slightly higher than the rate of homogenization within the introgressed lineage. Depending on the age of the secondary contact, the steps in allele frequencies magnified by foreign introgression may appear at variable recombination distances from the barrier loci.

To explore that effect, we first focused on the subset of SNPs that were the most strongly associated to barrier loci in our dataset, using only outlier loci that were detected in the *BayeScan* MED-ATL genome-scan with a stringent FDR threshold of 1%. Since genetic differentiation is strong between ATL and MED within genomic islands, the minor frequency allele found at each of these loci in the ATL most often

corresponds to the major frequency allele in the MED. Therefore, gradients of introgression of Mediterranean alleles at those loci are expected to generate minor allele frequency (MAF) clines within the Atlantic distribution range. In order to test for a cryptic population structure within the Atlantic, we thus compared the mean MAF of MED-ATL outlier loci to that of neutral SNPs as a function of the distance to the southernmost ATL location (SINE).

Finally, we evaluated how different degrees of linkage to barrier loci affect MAF gradients within the Atlantic. We assigned SNPs to different categories defined by their level of allele frequency difference (Δp) between ATL and MED. Our rationale was that stronger Δp values between ATL and MED are associated to stronger linkage to barrier loci, which was supported from earlier works (Duranton et al., 2018).

3 | RESULTS

3.1 | SNP genotyping and quality filtering

After excluding non-variable markers corresponding to polymorphisms private to the Mediterranean sea bass lineage as well as marker genotyping failures, 1361 SNPs were successfully genotyped in 846 specimens over the 1531 markers present on the array, representing an overall call-rate of 90.7%. SNPs with a minor allele frequency (MAF) lower than 0.01 within the Atlantic were subsequently excluded, as were individuals with more than 10% of non-scored genotypes. A few additional SNPs showing variation profiles significantly correlated with the samples' mapping within the 96-wells plates used for genotyping were also discarded to avoid experimental

effects. After these steps of quality-filtering, 1012 bi-allelic SNP markers that were genotyped in 827 individuals were retained in the final main dataset (Supplementary Table S1), resulting in an overall genotyping call rate of 96.87%.

3.2 | Genetic diversity gradient and spatial structure

The average fraction of polymorphic markers per sampling location calculated across the 1012 SNPs was negatively correlated to latitude after standardizing for the number of specimens genotyped by locality (adjusted $R^2 = 0.77$; $p < 0.001$; Figure 2a), providing evidence of a latitudinal decrease in genetic diversity across the ATL sea bass range. Similarly, the mean observed heterozygosity per sample location was negatively correlated to latitude (adj. r2= 0.66; $p < 0.001$; Supplementary Figure S3), decreasing by nearly 10 % from Southern Portugal to northeastern UK (Supplementary Table S2).

The proportion of the total genotypic variance explained by the redundancy analysis constrained by (latitude + longitude + month + year of capture) was low (0.64 %) but highly significant (p-value < 0.001). However, constrained ordination only revealed significant marginal effects on SNP variation for the two spatial variables (latitude: p-value < 0.001; longitude: p-value = 0.037). The RDA1 axis received a large contribution of latitude, as illustrated by the gradient in individual coordinates distributions among the 7 regions from Portugal to the North Sea (Figure 2b).

The overall F_{ST} calculated over all loci among the 21 Atlantic locations was not significant ($F_{ST} = 0.0002$, $p > 0.05$). Pairwise F_{ST} values calculated between pairs of ATL locations or regions were consistently low and most often non-significant (Supplementary Tables S3-S5), and only few pairwise F_{ST} values reached significance among the 7 ATL regions considered (Supplementary Table S6). By contrast, genetic

differentiation was much higher between ATL and MED sea bass lineages (average F_{ST} = 0.1474) using the same marker dataset, consistently with previous studies (Tine et al. 2014; Duranton et al. 2018). A weak but significant positive correlation was found between geographic distance separating pairs of ATL localities and genetic distance estimated by F_{ST} / (1- F_{ST}) (adj. r2 = 0.078; p < 0.001; Figure 2c). This isolation-by-distance (IBD) pattern was remarkably stronger when the analysis was restricted to the northern part of the Atlantic range, that is, excluding Portugal, Biscay and southwestern Channel locations (adj. r2 = 0.236; p-value < 0.001; Supplementary Figure S4). This was also confirmed by the Mantel test (Supplementary Table S7).

The analysis of directionality index Ψ based on the method by Peter and Slatkin (2013) detected a strongly significant deviation from an isolation-by-distance at equilibrium model. Instead, a spatial expansion scenario was supported with an inferred origin located in the northwestern part of the Celtic Sea (p-value < 10^{-10}, Figure 2d). The gradient in the relative fit as expansion origin showed a steep decrease in likelihood location towards the southern Celtic Sea and northern Biscay, and a smoother gradient towards the northeastern part of the range (Irish Sea, English Channel and North Sea). Together with the decreasing latitudinal gradient in heterozygosity, these results support a recent northeastern range expansion from a region located in the northwestern Celtic Sea.

3.3 | Spatial gradient of MED ancestry in the ATL

Our inferences of individual MED ancestry proportions showed a clear northward-decreasing gradient (Figure 3). The southernmost location in Portugal (SINE), which is the closest to the ATL-MED transition zone, showed a significantly higher average

level of MED ancestry (11.2 %) compared to all other locations and a stronger variance in MED ancestry among individuals.

The GLM analysis showed that the explanatory variable "distance to SINE" had a very significant negative effect ($p < 2e\text{-}16$) on the percentage of MED ancestry, whereas no significant effect could be detected for the date of sampling ($p=0.135$).

The spatial gradient in MED ancestry was even more obvious when individual ancestry proportions were averaged by sampling location (Figure 4a). The mean fraction of MED ancestry per location was negatively correlated with the distance to SINE in the simple linear regression model (adj. $r2 = 0.6026$, p-value < 0.001), and the correlation was strengthened when the SINE location (which is highly admixed) was removed from the analysis (adj. $r2 = 0.7405$; p-value < 0.001).

The piecewise regression approach detected two significant break points in the spatial admixture gradient. Taking into account the presence of two additional parameters, the model with three different linear regression displayed a significantly better fit to the data than the simple linear regression model (ANOVA for model comparison: p-value = 3.094e-9; best model with 3 regression lines: adj. $r2 = 0.93$, p-value = 1.3e-9). The strongest break occurred in the southwestern English Channel near the GONB locality in the Gulf of Saint-Malo, more than 2000 km northward to SINE by the plateau (Figure 4b). The second break was detected at the tip of Galicia between CORU and ASTU sampling locations (Figure 4c). Consistent with the dilution of MED ancestry toward the north, the average fraction of MED alleles decreased from 5.7 % [0.037, 0.076] in the Portugal region, to 2.6 % [0.013, 0.041] in the Bay of Biscay and southwestern English Channel, and 1.2 % [0.004, 0.022] in the north of Brittany from Celtic Sea to Irish and North Seas (Figure 3). Each of these two steps was characterized by a more than two-fold reduction in the mean MED ancestry from south to north (Figure 4d). Although we found negative regression

447 slopes indicating a decreasing mean MED ancestry with increasing distance to SINE
448 in both the Portugal and Biscay-southwestern Channel regions, the slope was
449 significantly positive to the north of the Gulf of Saint-Malo breakpoint. This striking
450 result, indicating an inversed latitudinal gradient in MED ancestry in the northern
451 part of the range, remained visible separately in a western Channel-Irish Sea transect
452 and an eastern Channel-North Sea transect using the more spatially 'refined dataset'
453 containing 31 localities (Supplementary Figure S5a).

454

455 **3.4 | Within-ATL outliers are enriched for between-lineages outliers**

456 Genome scans for highly differentiated SNPs were performed both at the within-ATL
457 scale to detect outlier loci within the northeastern Atlantic and between ATL and
458 MED samples to detect outlier loci showing an excess of differentiation between the
459 two European sea bass lineages.

460 *Lositan* detected 32 outliers exceeding the 99.5th percentile of the neutral
461 distribution of FST at the within-ATL scale (Supplementary Figure S6a) and 183
462 outliers between ATL and MED lineages (Supplementary Figure S6b). Within-ATL
463 outliers were highly significantly enriched for between-lineages outliers (21 out of 32
464 within-ATL outliers belong to the 183 ATL-MED outliers, hypergeometric test p-
465 value = 1.9e-9).

466 Comparatively, the *BayeScan* analysis was more stringent in detecting outlier loci
467 showing q-values < 0.05, with only 7 candidate outliers being found at the within-
468 ATL scale (Supplementary Figure S6c) and 74 outliers detected between ATL and
469 MED lineages (Supplementary Figure S6d). All of the outliers identified with
470 BAYESCAN were also detected by LOSITAN. Here again, the outliers detected at the
471 within-ATL scale were highly significantly enriched for between-lineages outliers (5

out of 7 within-ATL outliers belong to the 74 ATL-MED outliers, hypergeometric test p-value = 3.4e-5).

3.5 | Minor allele frequency steps at genomic islands associated loci

We used a stringent subset of 52 outlier loci that were detected in the BAYESCAN MED-ATL genome-scan with an FDR < 0.01 to characterize the gradient in mean MAF as a function of the distance to SINE. Major steps were observed directly northward to CORU and GONB in these mean MAF profiles, which were not detected using non-outlier loci (Figure 5). In order to refine the spatial positions of these steps, the same analysis was made based on the spatially refined dataset (31 localities). Mean MAF were represented separately along two transects: one from Portugal toward the Irish Sea and one toward the North Sea (Supplementary Figure S5b). Both profiles revealed low MAF values in the western Channel area (after PBRE-CRNW and PBRE-GONB), with the lowest values being found in IRIS, CRNW2 and 3, ISWI2 and BSEI1. Northward to these regions, the mean MAF profiles steadily increased towards WALE-SGEO (Irish Sea) and DOVE-ANGL-NEUK (North Sea), on both sides of Britain.

Finally, the evolution of mean MAF profiles was evaluated as a function of the degree of marker linkage to MED-ATL islands of divergence assessed with difference in allele frequencies between ATL and MED (Δp). Progressively flattening MAF profiles were found for groups of loci showing from high to low Δp values (Supplementary Figure S7).

4 | DISCUSSION

497

498 Genetic connectivity assessment in high gene flow species generally provides little
499 information on the degree of demographic connectivity among populations (Waples,
500 1998; Lowe & Allendorf, 2010). This shortcoming typically arises when genetic
501 differentiation at neutral markers is null or close to zero, so that no real signal can be
502 used to determine whether migration rates are sufficiently high to ensure
503 demographic coupling among populations. Here, we found non-significant overall
504 genetic differentiation across the vast majority of the European sea bass northeastern
505 Atlantic range. The sea bass thus represents one of the many borderline cases where
506 neutral markers may remain silent to conservation and management issues.

507 As for many other species, however, the spatial distribution of genetic diversity
508 in sea bass has retained signals of historical and contemporary admixture. Two sea
509 bass lineages, one Atlantic and one Mediterranean, have diverged in allopatry during
510 ca. 300,000 years and have subsequently experienced a postglacial secondary contact
511 which started less than 15,000 years BP (Tine et al., 2014; Duranton et al., 2018). Since
512 that time, gene flow has been strongly asymmetrical from the Atlantic into the
513 Mediterranean lineage. However, gene flow has been also occurring in the opposite
514 direction, accounting for less than 5% of introgressed Mediterranean alleles within
515 Atlantic sea bass genomes (Duranton et al., 2018). Here, we show that the northward
516 diffusion of Mediterranean alleles within the Atlantic sea bass populations can be
517 used to reveal two dispersal barriers separating demographically independent units,
518 and an ongoing expansion in the northern part of the species range. These findings
519 could provide useful new elements to improve the delineation of European sea bass
520 stocks in fisheries management programs.

521

522 **4.1 | Cryptic dispersal barriers revealed by introgression clines**

Introgression between differentiated lineages can be used as an alternate way to assess within-species genetic and demographic connectivity (Gagnaire et al., 2011; Gagnaire et al., 2015; Duranton et al., 2019). Several marine species display large spatial scale allele frequency patterns that can be attributed to ongoing gene flow between divergent units (e.g. Milano et al., 2014; Dahle et al., 2018; Lehnert et al., 2018; Wennevik et al., 2019). These admixture gradients originate at the edge of contact zones and sometimes attenuate over large distances beyond the contact zones. They are dynamically maintained by continuing gene flow between lineages, combined with a delayed rate of homogenization within the introgressed lineages due to spatially limited dispersal. Selection against introgressed fragments may also contribute to maintaining these admixture gradients (Sedghifar et al., 2016). Spatial discontinuities in admixture gradients indicate local reductions in the spread of introgressed alleles (Gagnaire et al., 2015). Therefore, they can be used to identify cryptic dispersal barriers that are not strong enough to produce genetic differentiation at migration-drift equilibrium.

Here, the analysis of individual ancestry proportions within the Atlantic revealed the existence of a wide admixture gradient extending from southern Portugal to the British Isles. Our closest sampling location from the contact zone with the Mediterranean sea bass lineage (Sines, South Portugal) displayed the highest mean and variance in individual Mediterranean ancestry (mean=0.11, s.d.=0.09). This was mainly driven by the presence of admixed genotypes with significantly higher proportions of Mediterranean ancestry above the background population level (e.g. 6 individuals between 0.15 and 0.40 in Sines), which most likely represent late-generation backcrosses. Therefore, our results support that ongoing gene flow between the two sea bass lineages contributes to maintain the admixture gradient.

548 Northward to Sines (SINE), the level of Mediterranean ancestry mostly remained
549 below 5%, displaying a general decreasing trend with increasing latitude. A detailed
550 examination of this admixture gradient revealed the existence of two significant
551 breakpoints located at the tip of Galicia (CORU) and near the Gulf of Saint-Malo
552 (GONB). Two-fold reductions in mean Mediterranean ancestry were observed
553 between sampling sites directly located in the southern vs. northern sides of these
554 two breaks. Such a magnitude of change cannot be easily explained by insufficient
555 sampling coverage generating spurious breaks in an otherwise continuous gradient.
556 Instead, it indicates the presence of cryptic barriers that significantly reduce gene
557 flow over relatively short distances, on a scale of a hundred kilometers or less (i.e. the
558 distance between two consecutive sampling sites).

559 The observed discontinuities in admixture proportions are predicted under
560 spatial admixture models, and can be maintained as long as differences in allele
561 frequencies exist between hybridizing lineages (Gagnaire et al., 2015). For marine
562 species with large population sizes and a planktonic larval stage, it is often assumed
563 that a rapid erosion of allele frequency differences occurs upon secondary contact.
564 However, the existence of genetic barriers to gene flow between Atlantic and
565 Mediterranean sea bass lineages have potentially delayed the dissipation of allele-
566 frequency gradients for a substantial amount of time (Tine et al., 2014; Duranton et
567 al., 2018). Our results support this interpretation since most outlier loci found within
568 the Atlantic were likely explained by the introgression of Mediterranean alleles
569 (Figure 5). Moreover, loci showing delayed rates of introgression between lineages
570 also displayed the largest steps in allele frequencies within the Atlantic, as predicted
571 by theory (Gagnaire et al., 2015; Supplementary Figure S7).

572 Our results suggest that migration rates between populations on either side of
573 the identified breaks are too low to ensure demographic coupling, otherwise the

574 breaks would not have been observed. However, quantifying the rate of effective
575 migration across these breaks would require additional data on the intensity of
576 selection acting at the barrier loci and their recombination rate with the neutral
577 markers we used here. This remains challenging, and therefore calls for the use of
578 complementary information on individual movement, as inferred for instance from
579 archival tags (de Pontual et al., 2019) or scale microchemistry (Cambiè et al., 2016).
580 Such a coupling between life-history migration and genetic data has the potential to
581 refine our understanding of sea bass demographic connectivity in the northeastern
582 Atlantic.

583

584 **4.2 | Surfing of introgressed alleles owing to northward range expansion**

585 Northward to the break identified near the Gulf of Saint-Malo, the mean
586 Mediterranean ancestry surprisingly increased towards the northernmost localities of
587 WALE-SGEO (Irish Sea) and DOVE-ANGL-NEUK (North Sea) on both sides of
588 Britain (Supplementary Figure S5). This result was explained by an increased
589 variance in Mediterranean ancestry among individuals (Figure 3). The mean
590 Mediterranean ancestry in the English Channel is close to 1%, and therefore the
591 increased variance in ancestry in the Irish Sea and North Sea results in increased
592 average ancestry due to border effect (i.e. individual ancestries can deviate more
593 strongly when they lie above versus below the 1% average ancestry).

594 A possible mechanism underlying the increased variance in Mediterranean
595 ancestry is a recent range expansion in the northern part of the species distribution
596 area. This spatial expansion hypothesis was supported both by the decreasing
597 latitudinal gradient in heterozygosity and the directionality index analysis. A
598 number of direct observations also support the recent increase in sea bass abundance
599 around Britain (Wright et al. 2020). This is exemplified by the rapid increase of

Eastern Channel, North and Irish Seas populations attributed to increased seawater winter temperatures (Henderson, 2007; Pawson et al., 2007), and the new or expanding sea bass fisheries in UK and Irish seas (Cheung et al., 2012).

In a spatially expanding population, the intensity of genetic drift is amplified at the wave front, which leads to increased variance in allele frequencies compared to the core population (Edmonds et al., 2004; Klopfstein et al., 2006). Putting this in the sea bass context may help explain how range expansion affects Mediterranean ancestry in the northern part of the species range. Alleles of Mediterranean origin segregate at low frequencies at the place where spatial expansion occurs. Therefore, most of these low-frequency Mediterranean alleles are likely to be lost by random genetic drift or to remain at low frequency during range expansion. Some alleles of Mediterranean origin, however, can successfully surf on the wave of advance and reach significantly higher frequencies. These two opposite effects of genetic drift (i.e. small decrease in frequency for most Mediterranean alleles but successful surfing for part of them) is expected to increase the variance in ancestry among individuals living in the expansion zone. Thus, our results are consistent with the surfing of a fraction of the rare Mediterranean alleles, as a consequence of ongoing northward expansion of sea bass populations on both sides of Britain. A more thorough examination of this hypothesis would require denser genome-wide polymorphism data to test the surfing of local ancestry blocks of Mediterranean origin.

A possible alternative explanation for the observed ancestry patterns is selection against weakly deleterious introgressed fragments of Mediterranean origin. Since the rate of introgression is particularly reduced in low-recombining regions of the sea bass genome, selection against introgression is thought to rely on many loci of small individual effects (Duranton et al., 2018). In the large populations of Portugal, Biscay and western Channel, selection against introgressed tracts of Mediterranean origin

could be sufficiently efficient to progressively decrease Mediterranean ancestry as the distance to the contact zone increases. By contrast, lower effective population sizes in the northern margins of the species range could reduce the efficacy of selection against introgression, resulting in slightly increased mean Mediterranean ancestry in these regions. The two proposed mechanisms (i.e. surfing of introgressed alleles and reduced efficacy of selection against introgression) are not mutually exclusive and may collectively explain the spatial ancestry patterns documented here.

4.3 | Applications to conservation and management

Our results mirror previous findings that support the existence of dispersal barriers in the two identified regions of Galicia and Brittany for other marine taxa.

The northwestern coast of the Iberian Peninsula is characterized by complex seasonal oceanographic circulation patterns that may act as potential barriers to dispersal (Varela et al., 2005). Mixed evidence from genetic studies point out that, on the one hand, most marine species lack genetic structure across northwestern Iberia (reviewed in Gomes-Gesteira et al., 2011). On the other hand, some exceptions have been found such as in *Littorina saxatilis* (Piñeira et al., 2008) or in *Fucus ceranoides* (Neiva et al., 2012), which both lack planktonic dispersive stages. In these species, however, genetic breaks may also be attributed to past secondary contacts.

Brittany represents a less controversial biogeographic transition zone between marine biotas. In some species like *Fucus vesiculosus*, the genetic break occurs near the Ushant thermal front that lies off the coast of western Brittany (e.g. Almedia et al., 2017), and separates cold inshore surface waters from warm stratified offshore

651 waters (Chevallier et al., 2014). Another important transition zone, visible in species
652 assemblages of benthic macrofauna (Gaudin et al., 2018), and contact zones between
653 divergence lineages (Bierne et al., 2003; Jolly et al., 2005), corresponds to the Gulf of
654 Saint-Malo that separates western from eastern English Channel.

655 Most interestingly, a genetic connectivity study in the marine trematode parasite
656 *Bucephalus minimus*, whose final and most vagile host is *Dicentrarchus labrax*, revealed
657 significant population structure along the northeastern Atlantic coastline (Feis et al.,
658 2015). Population samples of parasites taken from both sides of the two dispersal
659 barriers evidenced here exhibited genetic differentiation, which in turn suggested
660 some level of isolation among their host sea bass populations. By contrast,
661 population genetic studies conducted in sea bass generally failed to detect any
662 significant genetic structure between populations of the Bay of Biscay and English
663 Channel-Celtic Sea (Frisch et al., 2007; Coscia & Mariani, 2011; Souche et al., 2015).
664 On the other hand, mark-recapture data indicated restricted movement between
665 these two populations, thus supporting their demographic independence (Fritsch et
666 al., 2007). An important component in the maintenance of independent demographic
667 units is the fidelity to winter spawning areas, recently evidenced with tagging data
668 (de Pontual et al., 2019). Limited effective dispersal of larvae across barriers in Galicia
669 and Brittany is also implicitly supported by our findings, and could be explained by
670 the location of spawning areas in relation to the main oceanic current features.

671 The results presented here are thus broadly consistent with previous findings.
672 They confirm that (*i*) Atlantic sea bass populations exhibit strong genetic connectivity
673 at neutral markers, which implies that (*ii*) gene flow is high enough to ensure the
674 spread of adaptive variants among populations. (*iii*) However, the number of
675 migrants per generation among populations separated by barriers does not seem
676 sufficiently large to homogenize the latitudinal gradient in Mediterranean ancestry

677 across the whole genome. This supports the existence of already suspected
678 demographically independent units, maintained by behavioral processes and
679 spatially limited larval dispersal. Finally, (*iv*) northern Atlantic sea bass populations
680 seem to exhibit genetic signatures of ongoing northward expansion on both sides of
681 Britain. The demographic independence of these expanding populations with respect
682 to neighboring populations from Celtic Sea and English Channel could not be
683 specifically addressed here and will need further examination. However, the finding
684 of weak but significant genetic differentiation between Irish-Celtic Seas and North
685 Sea populations, indicates demographic independence over long distances.

686 These conclusions potentially have direct implications for the management of
687 European sea bass populations in the northeastern Atlantic. The most recent report of
688 the ICES advisory committee for sea bass populations (ICES, 2019) states that
689 spawning-stock biomass has been declining since 2009-2010 and is now just above or
690 below the maximum sustainability yield, depending on the region considered. The
691 need for precise stock assessment is therefore more crucial than ever for a proper
692 management of fisheries. To this end, the spatial delineation of at least three main
693 stocks would need to be revised to account for these new findings. Namely, the
694 northern management unit comprising populations from the North Sea, Irish Sea,
695 English Channel, Bristol Channel, and Celtic Sea includes areas that most likely
696 belong to the northern part of the Bay of Biscay population. Moreover, the Bay of
697 Biscay south, which is currently grouped with the Atlantic Iberian Waters, most
698 likely belongs to the Bay of Biscay population.

699 To conclude, the approach presented in this study supports that introgression
700 gradients consisting of less than 5% of introgressed ancestry are sufficient to detect
701 cryptic contemporary barriers to gene flow, even using moderate genome coverage.
702 Estimation of migration rates across these barriers and identification of ideal

703 conditions for quantitative assessments will require further development. However,
704 we believe that these will be encouraged by the frequent discovery of admixture in
705 many species, thus expanding the possibilities of studying genetic connectivity in
706 other taxa.

707

708 **DATA ACCESSIBILITY STATEMENT**

709 - Raw data, genotypes per site, variables associated with sampling sites have
710 been deposited here : https://doi.org/10.5281/zenodo.3899247

711 - All scripts used have been deposited here :
712 https://github.com/tonyrobinet/introgression

713

714

715 **ACKNOWLEDGEMENTS**

716 This study was funded by France Filière Pêche (FFP, convention PH/2013/11)
717 and the Muséum National d'Histoire Naturelle (MNHN, SJ 833-13). Authors thank
718 Morgane Ramonet (FFP) and Thierry Guigues (Pêcheurs de Bretagne) for supporting
719 this work from the project (2012) to the final report (2018). Thanks to producer's
720 organizations for their collaboration (Coopérative Maritime Étaploise, Cobrenord,
721 Pêcheurs de Bretagne, Pêcheurs d'Aquitaine, Plymouth Trawler Agents), and to the
722 34 fish markets that kindly opened their doors, nights and days. Special thanks to:
723 Mickaël Drogou and the Ifremer LBH lab in Brest for sharing tissues from Biscay;
724 Ifremer sea campain in the Bay of Biscay (N/O Thalassa, EVOHE 2012-14); the
725 fishing vessel Les Calanques (South Ireland); Jasmin Hilien and Filip Volckaert
726 (Leuven Univ.) for sharing some tissues from UK; David Righton, Stephen Shaw
727 (CEFAS Lowestoft), John Ashworth (Newlyn) for collecting fin-clips in UK waters (in

fish markets and on N/O Endeavour); Alison M. Pessell (Plymouth Trawler Agents); Eva Maria Velasco Gil (Instituto Espanol de Oceanografia, Gijon) and Eneko Aierbe Sarasola (Gijon) for collecting fin-clips in Asturian fish markets; Alexandre Alonso-Fernandez, Fran Saborido-Rey and David Villegas-Rios (Instituto de Investigaciones Marinas, CSIC, Vigo-Pontevedra) for their kind collaboration in Galicia; Joana Robalo and Carla Sousa Santos (ISPA MARE, Lisboa) for collecting fin-clips in Portuguese fish markets. Thanks to Daniel Sellos (MNHN Concarneau) for his kind help in the lab. The genotyping chipset was designed by both François Allal (Ifremer Palavas) and PAG. The last special-special-thanks go to the numerous little hands from the marine station MNHN of Concarneau, for their much appreciated effort (Claudie Guenguant, Thomas Barreau, Bernard Bourlès, Marie Cadou, Guillaume Dirberg, Lou Frotté, Samuel Iglésias, Pierre-Yves Lebon, Sébastien Mayot, Julien Ringelstein, Héloïse You!). The manuscrit was improved by three anonymous reviewers, that we would like to acknowledge for their time, positive assessment and helpful comments.

AUTHORS CONTRIBUTIONS

TR and VR wrote the proposal, raised funds; TR collected specimens with many others; PAG conducted and supervised the analysis, with the help of TR and VR; TR and KC managed collections and prepared DNA for genotyping; TR and PAG analyzed and interpreted the results; TR and PAG wrote the paper, with the help of VR.

752 **REFERENCES**

753 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of
754 ancestry in unrelated individuals. *Genome research*, 19(9), 1655-1664.

755 Almeida, S. C., Nicastro, K. R., Zardi, G. I., Pearson, G. A., Valero, M., & Serrão, E. A.
756 (2017). Reproductive strategies and population genetic structure of *Fucus* spp.
757 across a northeast Atlantic biogeographic transition. *Aquatic Living Resources*, *30*,
758 16.

759 Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: a
760 workbench to detect molecular adaptation based on a FST-outlier method. *BMC*
761 *bioinformatics*, 9(1), 323.

762 Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of
763 population structure. *Proceedings of the Royal Society of London, Series B: Biological*
764 *Sciences*, 263, 1619– 1626.

765 Bekkevold, D., Gross, R., Arula, T., Helyar, S. J., & Ojaveer, H. (2016). Outlier loci
766 detect intraspecific biodiversity amongst spring and autumn spawning herring
767 across local scales. *PloS one*, *11*(4).

768 Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., ...
769 & Bernatchez, L. (2016). Seascape genomics provides evidence for thermal
770 adaptation and current−mediated population structure in American lobster
771 (*Homarus americanus*). *Molecular Ecology*, *25*(20), 5073-5092.

772 Bertl, J., Ringbauer, H., & Blum, M. G. (2018). Can secondary contact following range
773 expansion be distinguished from barriers to gene flow?. *PeerJ*, *6*, e5325.

774 Bierne, N., Borsa, P., Daguin, C., Jollivet, D., Viard, F., Bonhomme, F., & David, P.
775 (2003). Introgression patterns in the mosaic hybrid zone between *Mytilus edulis*
776 and *M. galloprovincialis*. *Molecular Ecology*, *12*(2), 447-461.

777 Bierne, N., Roze, D., & Welch, J. J. (2013). Pervasive selection or is it…? why are FST
778 outliers sometimes so frequent? *Molecular ecology*, *22*(8), 2061-2064.

779 Cambiè, G., Kaiser, M. J., Marriott, A. L., Fox, J., Lambert, G., Hiddink, J. G., ... &
780 McCarthy, I. D. (2016). Stable isotope signatures reveal small-scale spatial
781 separation in populations of European sea bass. *Marine Ecology Progress Series*, *546*,
782 213-223.

783 Cayuela, H., Rougemont, Q., Prunier, J. G., Moore, J. S., Clobert, J., Besnard, A., &
784 Bernatchez, L. (2018). Demographic and genetic approaches to study dispersal in
785 wild animal populations: A methodological review. *Molecular Ecology*, *27*(20),
786 3976-4010.

787 Cheung, W. W., Pinnegar, J., Merino, G., Jones, M. C., & Barange, M. (2012). Review
788 of climate change impacts on marine fisheries in the UK and Ireland. *Aquatic*
789 *Conservation: Marine and Freshwater Ecosystems*, *22*(3), 368-388.

790 Chevallier, C., Herbette, S., Marié, L., Le Borgne, P., Marsouin, A., Péré, S., ... &
791 Reason, C. (2014). Observations of the Ushant front displacements with
792 MSG/SEVIRI derived sea surface temperature data. *Remote sensing of environment*,
793 *146*, 3-10.

794 Coscia, I., & Mariani, S. (2011). Phylogeography and population structure of
795 European sea bass in the north-east Atlantic. *Biological Journal of the Linnean Society*,
796 104(2), 364-377.

797 Dahle, G., Quintela, M., Johansen, T., Westgaard, J. I., Besnier, F., Aglen, A., ... &
798 Glover, K. A. (2018). Analysis of coastal cod (*Gadus morhua* L.) sampled on
799 spawning sites reveals a genetic gradient throughout Norway's coastline. *BMC*
800 *genetics*, *19*(1), 42.

801 Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P. A.
802 (2018). The origin and remolding of genomic islands of differentiation in the
803 European sea bass. *Nature communications*, 9(1), 2518.

804 Duranton, M., Bonhomme, F., & Gagnaire, P. A. (2019). The spatial scale of dispersal
805 revealed by admixture tracts. *Evolutionary Applications*. 12(9) 1743-1756.

806 Edmonds, C. A., Lillie, A. S., & Cavalli-Sforza, L. L. (2004). Mutations arising in the
807 wave front of an expanding population. *Proceedings of the National Academy of*
808 *Sciences*, *101*(4), 975-979.

809 Faggion, S., Vandeputte, M., Chatain, B., Gagnaire, P. A., & Allal, F. (2019).
810 Population-specific variations of the genetic architecture of sex determination in
811 wild European sea bass Dicentrarchus labrax L. *Heredity*, *122*(5), 612.

812 Feis, M. E., Thieltges, D. W., Olsen, J. L., de Montaudouin, X., Jensen, K. T., Bazaïri,
813 H., ... & Luttikhuizen, P. C. (2015). The most vagile host as the main determinant
814 of population connectivity in marine macroparasites. *Marine Ecology Progress*
815 *Series*, *520*, 85-99.

816 Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci
817 appropriate for both dominant and codominant markers: a Bayesian perspective.
818 *Genetics* 180:977-93.

819 Fritsch, M., Morizur, Y., Lambert, E., Bonhomme, F., & Guinand, B. (2007).
820 Assessment of sea bass (*Dicentrarchus labrax*, L.) stock delimitation in the Bay of
821 Biscay and the English Channel based on mark-recapture and genetic data.
822 *Fisheries Research*, 83(2-3), 123-132.

823 Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing
824 genomics for delineating conservation units. *Trends in ecology & evolution*, *27*(9),
825 489-496.

826 Gagnaire, P. A., Minegishi, Y., Zenboudji, S., Valade, P., Aoyama, J., & Berrebi, P.
827 (2011). Within-population structure highlighted by differential introgression
828 across semipermeable barriers to gene flow in *Anguilla marmorata. Evolution:*
829 *International Journal of Organic Evolution*, 65(12), 3413-3427.

830 Gagnaire, P.A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F.,
831 Arnaud-Haond, S., & Bierne, N. (2015). Using neutral, selected, and hitchhiker loci
832 to assess connectivity of marine populations in the genomic era. *Evolutionary*
833 *Applications*, 8 (8), 769–786.

834 Gagnaire, P. A. (2020). Comparative genomics approach to evolutionary process
835 connectivity. *Evolutionary Applications*, 13(6), 1320-1334.

836 Gaudin, F., Desroy, N., Dubois, S. F., Broudin, C., Cabioch, L., Fournier, J., ... &
837 Thiébaut, É. (2018). Marine sublittoral benthos fails to track temperature in
838 response to climate change in a biogeographical transition zone. *ICES Journal of*
839 *Marine Science*, *75*(6), 1894-1907.

840 Gomez-Gesteira, M., Beiras, R., Presa, P., & Vilas, F. (2011). Coastal processes in
841 northwestern Iberia, Spain. *Continental Shelf Research*, *31*(5), 367-375.

842 Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-
843 statistics. *Molecular Ecology Notes*, 5(1), 184-186.

844 Hedgecock, D., Barber, P. H., & Edmands, S. (2007). Genetic approaches to
845 measuring connectivity. *Oceanography*, *20*(3), 70-79.

846 Henderson, P. A. (2007). Discrete and continuous change in the fish community of
847 the Bristol Channel in response to climate change. *Journal of the Marine Biological*
848 *Association of the United Kingdom*, *87*(2), 589-598.

849 Hommel, G. (1989). A comparison of two modified Bonferroni procedures.
850 *Biometrika*, 76(3), 624-625.

ICES. 2019. Sea bass (*Dicentrarchus labrax*) in Divisions 4.b–c, 7.a, and 7.d–h (central and southern North Sea, Irish Sea, English Channel, Bristol Channel, and Celtic Sea). In Report of the ICES Advisory Committee, 2019, bss.27.4bc7ad-h, https://doi.org/10.17895/ices.advice.4779

Jolly, M. T., Jollivet, D., Gentil, F., Thiébaut, É., & Viard, F. (2005). Sharp genetic break between Atlantic and English Channel populations of the polychaete *Pectinaria koreni*, along the North coast of France. *Heredity*, *94*(1), 23.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular biology and evolution*, *23*(3), 482-490.

Lehnert, S. J., DiBacco, C., Jeffery, N. W., Blakeslee, A. M., Isaksson, J., Roman, J., ... & Hamilton, L. C. (2018). Temporal dynamics of genetic clines of invasive European green crab (*Carcinus maenas*) in eastern North America. *Evolutionary Applications*, 11(9), 1656-1670.

Liggins, L., Treml, E. A., & Riginos, C. (2020). Seascape genomics: contextualizing adaptive and neutral genomic variation in the ocean environment. *Population Genomics: Marine Organisms*, 171-218.

Lowe, W.H., & Allendorf, F.W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology* 19, 3038-3051.

Lemaire, C., Versini, J. J., & Bonhomme, F. (2005). Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal of Evolutionary Biology*, 18(1), 70-80.

874 Milano, I., Babbucci, M., Cariani, A., Atanassova, M., … & Bargelloni, L. (2014).
875 Outlier SNP markers reveal fine−scale genetic structuring across European hake
876 populations (*Merluccius merluccius*). *Molecular Ecology 23*(1), 118-135.

877 Morgan, A. P. (2016). argyle: an R package for analysis of Illumina genotyping
878 arrays. *G3: Genes, Genomes, Genetics*, *6*(2), 281-286.

879 Naciri, M., Lemaire, C., Borsa, P., & Bonhomme, F. (1999). Genetic study of the
880 Atlantic/Mediterranean transition in sea bass (*Dicentrarchus labrax*). *Journal of*
881 *Heredity*, 90(6), 591-596.

882 Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013).
883 Genotyping−by−sequencing in ecological and conservation genomics. *Molecular*
884 *ecology*, *22*(11), 2841-2847.

885 Neiva, J., Pearson, G. A., Valero, M., & Serrão, E. A. (2012). Fine-scale genetic breaks
886 driven by historical range dynamics and ongoing density-barrier effects in the
887 estuarine seaweed *Fucus ceranoides* L. *BMC Evolutionary Biology*, *12*(1), 78.

888 Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., O'hara, R. B., Simpson, G. L., ... &
889 Wagner, H. (2019). Vegan: community ecology package. R package version 2.5-6.
890 https://cran.r-project.org/web/packages/vegan/ (accessed 2020.03.22)

891 Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design
892 of marine reserves. *Ecological applications*, *13*(sp1), 146-158.

893 Pante, E., & Simon-Bouhet, B. (2013). marmap: a package for importing, plotting and
894 analyzing bathymetric and topographic data in R. *PLoS One*, 8(9), e73051.

895 Paradis, E. (2010). pegas: an R package for population genetics with an integrated–
896 modular approach. *Bioinformatics*, 26(3), 419-420.

897 Pawson, M. G., Kupschus, S., & Pickett, G. D. (2007). The status of sea bass
898 (*Dicentrarchus labrax*) stocks around England and Wales, derived using a separable

899 catch-at-age model, and implications for fisheries management. *ICES Journal of*
900 *Marine Science*, *64*(2), 346-356.

901 Peter, B. M., & Slatkin, M. (2013). Detecting range expansions from genetic data.
902 *Evolution*, *67*(11), 3274-3289.

903 Piñeira, J., Quesada, H., Rolán-Alvarez, E., & Caballero, A. (2008). Genetic
904 discontinuity associated with an environmentally induced barrier to gene
905 exchange in the marine snail *Littorina saxatilis*. *Marine Ecology Progress Series*, *357*,
906 175-184.

907 Pontual (de), H., Lalire, M., Fablet, R., Laspougeas, C., Garren, F., Martin, S., Drogou,
908 M., & Woilez, M. (2019). New insights into behavioural ecology of European
909 seabass off the West Coast of France: implications at local and population
910 scales. *Ices Journal of Marine Science*, 76(2), 501-515.

911 Pulliam, H. R. (1988). Sources, sinks, and population regulation. *The American*
912 *Naturalist*, *132*(5), 652-661.

913 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... &
914 Sham, P. C. (2007). PLINK: a tool set for whole-genome association and
915 population-based linkage analyses. *The American journal of human genetics*, 81(3),
916 559-575.

917 Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local
918 adaptation. *Nature Reviews Genetics*, *14*(11), 807-820.

919 Sedghifar, A., Brandvain, Y., Ralph, P., & Coop, G. (2015). The spatial mixing of
920 genomes in secondary contact zones. *Genetics*, *201*(1), 243-261.

921 Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and
922 haplotype blocks in hybrid zones. *Molecular ecology*, *25*(11), 2559-2576.

923 Souche, E. L., Hellemans, B., Babbucci, M., MacAoidh, E., Guinand, B., Bargelloni,
924 L., ... & Volckaert, F. A. (2015). Range-wide population structure of European sea
925 bass *Dicentrarchus labrax*. *Biological Journal of the Linnean Society*, 116(1), 86-105.

926 Stapley, J., Reger, J., Feulner, P. G., Smadja, C., Galindo, J., Ekblom, R., ... & Slate, J.
927 (2010). Adaptation genomics: the next generation. *Trends in ecology & evolution*,
928 *25*(12), 705-712.

929 Tigano, A., & Friesen, V.L. (2016). Genomics of local adaptation with gene flow.
930 *Molecular Ecology*, 25, 2144–64.

931 Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S., ... &
932 Dieterich, R. (2014). European sea bass genome and its variation provide insights
933 into adaptation to euryhalinity and speciation. *Nature communications*, 5, 5770.

934 Vandeputte, M., Gagnaire, P. A., & Allal, F. (2019). The European sea bass: a key
935 marine fish model in the wild and in aquaculture. *Animal genetics*, *50*(3), 195-206.

936 Van Wyngaarden, M., Snelgrove, P. V., DiBacco, C., Hamilton, L. C., Rodríguez‐
937 Ezpeleta, N., Jeffery, N. W., ... & Bradbury, I. R. (2017). Identifying patterns of
938 dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*,
939 using RAD seq‐derived SNP s. *Evolutionary Applications*, *10*(1), 102-117.

940 Varela, R. A., Rosón, G., Herrera, J. L., Torres-López, S., & Fernández-Romero, A.
941 (2005). A general view of the hydrographic and dynamical patterns of the Rías
942 Baixas adjacent sea area. *Journal of Marine Systems*, *54*(1-4), 97-113.

943 Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic
944 differentiation in high gene flow species. *Journal of Heredity*, 89, 438–450.

945 Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation
946 of some genetic methods for identifying the number of gene pools and their
947 degree of connectivity. Molecular Ecology, 15(6), 1419–1439.

948 Waples, R.S., Punt, A.E., & Cope J.M. (2008). Integrating genetic data into
949 management of marine resources: how can we do it better? *Fish and Fisheries*
950 2008(9), 423-449.

951 Wennevik, V., Quintela, M., Skaala, Ø., Verspoor, E., Prusov, S., & Glover, K. A.
952 (2019). Population genetic analysis reveals a geographically limited transition zone
953 between two genetically distinct Atlantic salmon lineages in Norway. *Ecology and*
954 *Evolution* , 9(12), 6901-6921.

955 Wright, P. J., Pinnegar, J. K., & Fox, C. (2020). Impacts of climate change on fish,
956 relevant to the coastal and marine environment around the UK. *MCCIP Science*
957 *Review 2020*, 354-381.

958

959

960 **FIGURES CAPTIONS**

961

962 Figure 1. Sampling map of the 827 Atlantic sea bass individuals analyzed in this
963 study. In the 'main dataset', the 827 specimens were assigned to 21 localities
964 corresponding to single ICES rectangles or groups of adjacent rectangles
965 represented with the same color. In the 'regional dataset', specimens were
966 grouped into 7 regions represented by colored polygons. Total number of samples
967 is indicated below each region name. Details about sample locations and the
968 spatially 'refined dataset' are provided in Fig. S1 and Table S1.

969

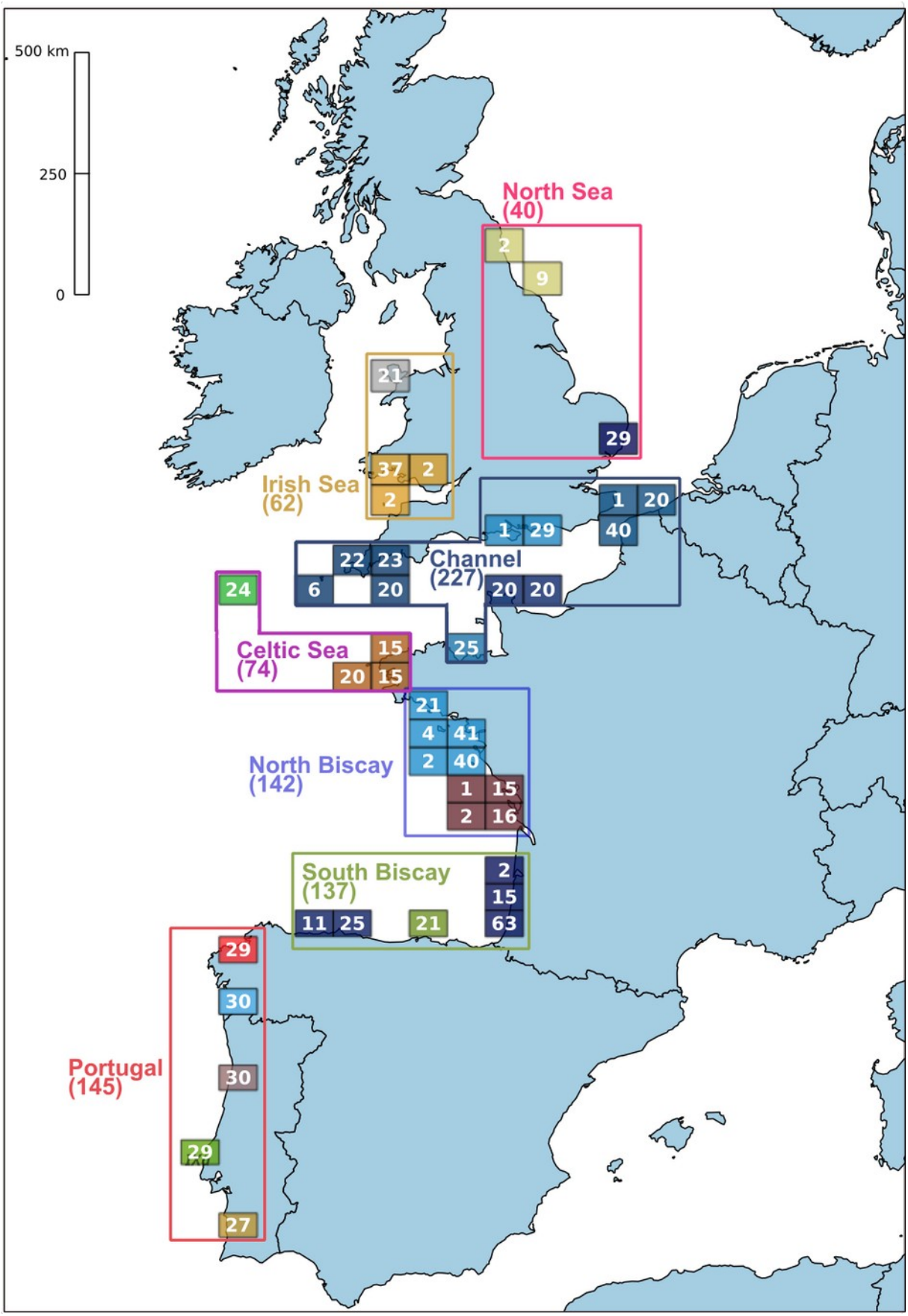
970 Figure 2. Spatial genetic variation. a: Negative correlation between standardized
971 polymorphism corrected for sampling size and latitude of the 21 sampling
972 localities. The blue line shows the linear regression (slope $p<0.001$, $r^2=0.77$) and its

0.95 confidence interval (grey shade). b: Distribution of individual coordinates on RDA1 axis as a function of latitude for each of the 7 regions: Portugal (Po), South Biscay (SB), North Biscay (NB), Celtic Sea (CS), Channel (Ch), Irish Sea (IS) and North Sea (NS). The RDA was constrained using only spatial coordinates. For each region, the boxplot horizontal line represents the mean, whiskers the standard deviation and vertical lines the 0.95 confidence interval. Dots indicate outlier individuals, and whisker width cover the latitudinal range of each region. c: Positive correlation between genetic differentiation represented by Fst/(1-Fst) and distance by the plateau (in km) between pairs of localities. The blue line shows the linear regression (slope $p<0.001$, $r^2=0.078$) and its 0.95 confidence interval (grey shade). d: Results of the directionality index analysis showing the gradient in the relative fit as expansion origin (color scale) and most likely origin of expansion in southern Ireland (blue cross). For each of the 21 sampling localities, the inner circle indicates the mean heterozygosity (grey scale) and the color of the outer circle indicates the group identified with the admixture gradient analysis (blue: Portugal, pink: Biscay, green: Celtic Sea - Channel - North Sea, see Fig. 3).

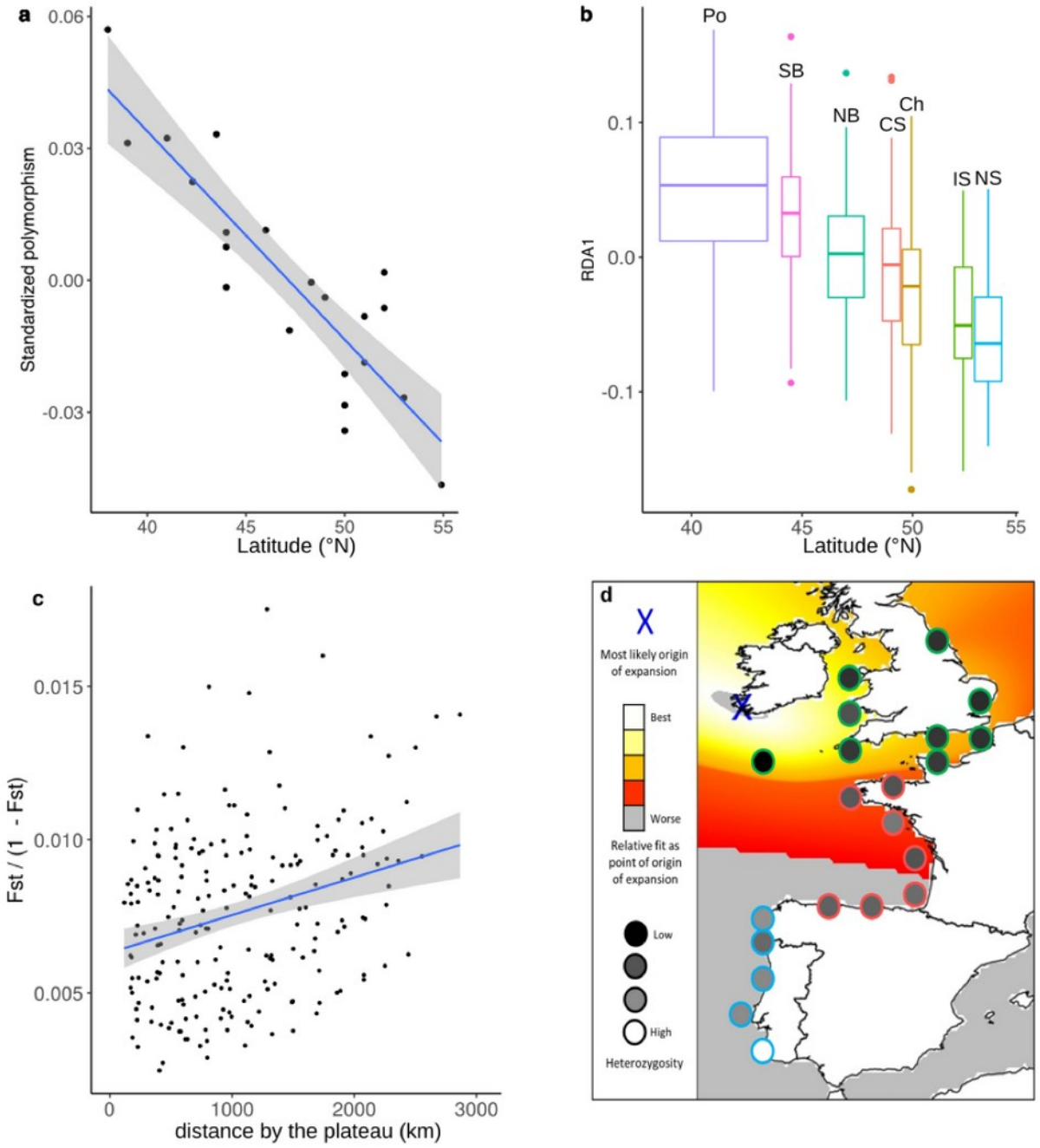
Figure 3. Individual ancestry proportions. Individual proportions of Mediterranean ancestry inferred with ADMIXTURE (point estimation and standard error bars) are displayed for the 10 reference Mediterranean samples (MED) and the 827 Atlantic individuals. Three groups of sampling locations that differ in their average ancestry proportions (shown with the means of lower and upper values) are delimited by vertical dashed bars: Portugal, Biscay, and Celtic Sea - Channel - North Sea. Individuals are ranked by their specimen code number, as referred in the dataset, so their ranking within the groups is meaningless.

Figure 4. Determination of the positions of barriers to dispersal. a: Mean proportion of MED ancestry (shown with the means of lower and upper values) inferred with ADMIXTURE for each of the 21 localities, as a function of the distance to SINE by the plateau (km). b: Plot of the residual standard error obtained using different breakpoint values (x axis) between two linear models of mean MED ancestry as a function of distance. The position with the minimal residual standard error is directly northward to the GONB locality (red line). c: Detection of a second breakpoint directly northward to the CORU locality (red line). d: Representation of the best piecewise regression model with the two break positions determined in b and c. Colored dots: mean MED ancestry for each group of locations: Portugal (blue), Biscay (pink) and Celtic Sea - Channel - North Sea (green).

Figure 5. Loci linked to genetic barriers to gene flow between ATL and MED lineages display stronger genetic differentiation within the Atlantic. a: Scaled distributions of per-locus FST values between ATL and MED lineages for non-outlier (N = 960, red density curve) and outlier SNPs (N=52, blue density curve). Candidate outliers for between-lineage divergence were identified with BayeScan using a q-value threshold of 0.01. b: Spatial profiles of mean minor allele frequency (MAF) for non-outlier (N = 960, red dots) and outlier SNPs (N=52, blue triangles), as a function of the distance to SINE by the plateau (km).



1020 Figure 1.



1021 Figure 2.

1022

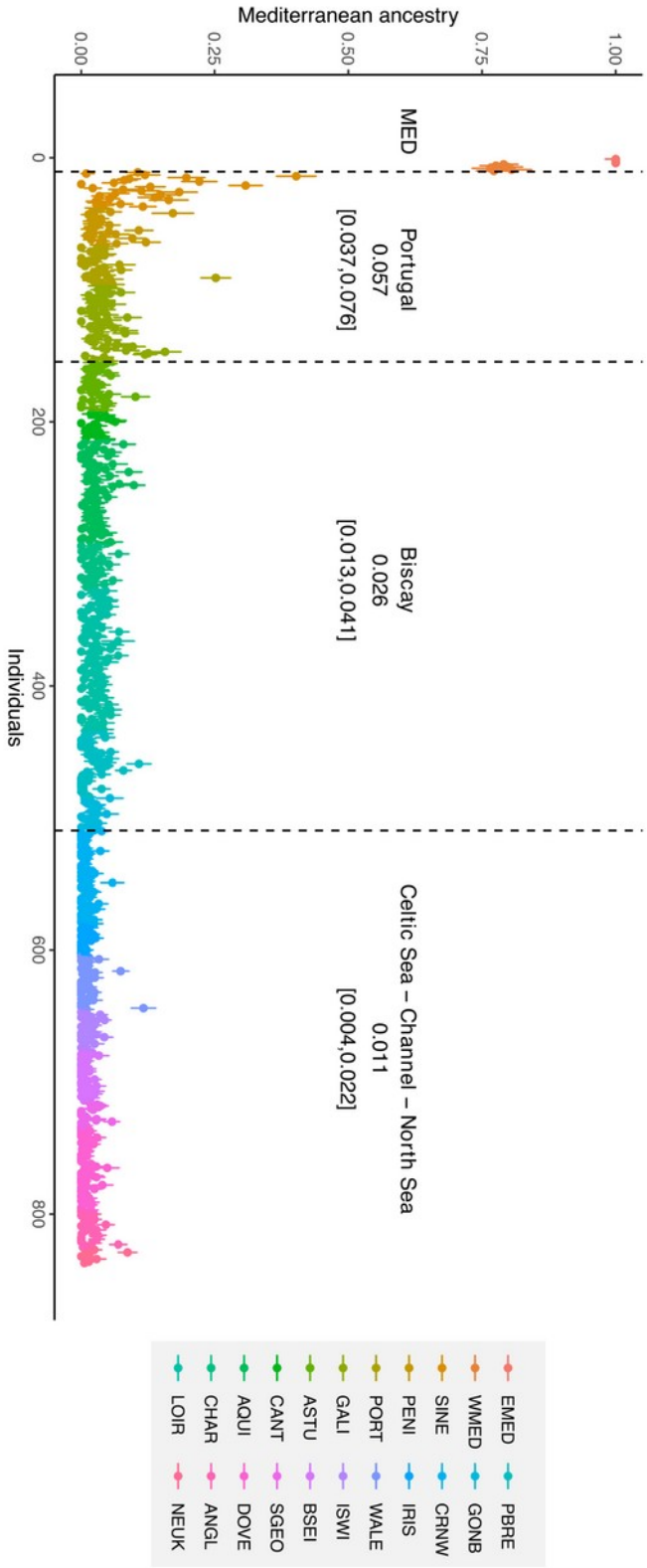
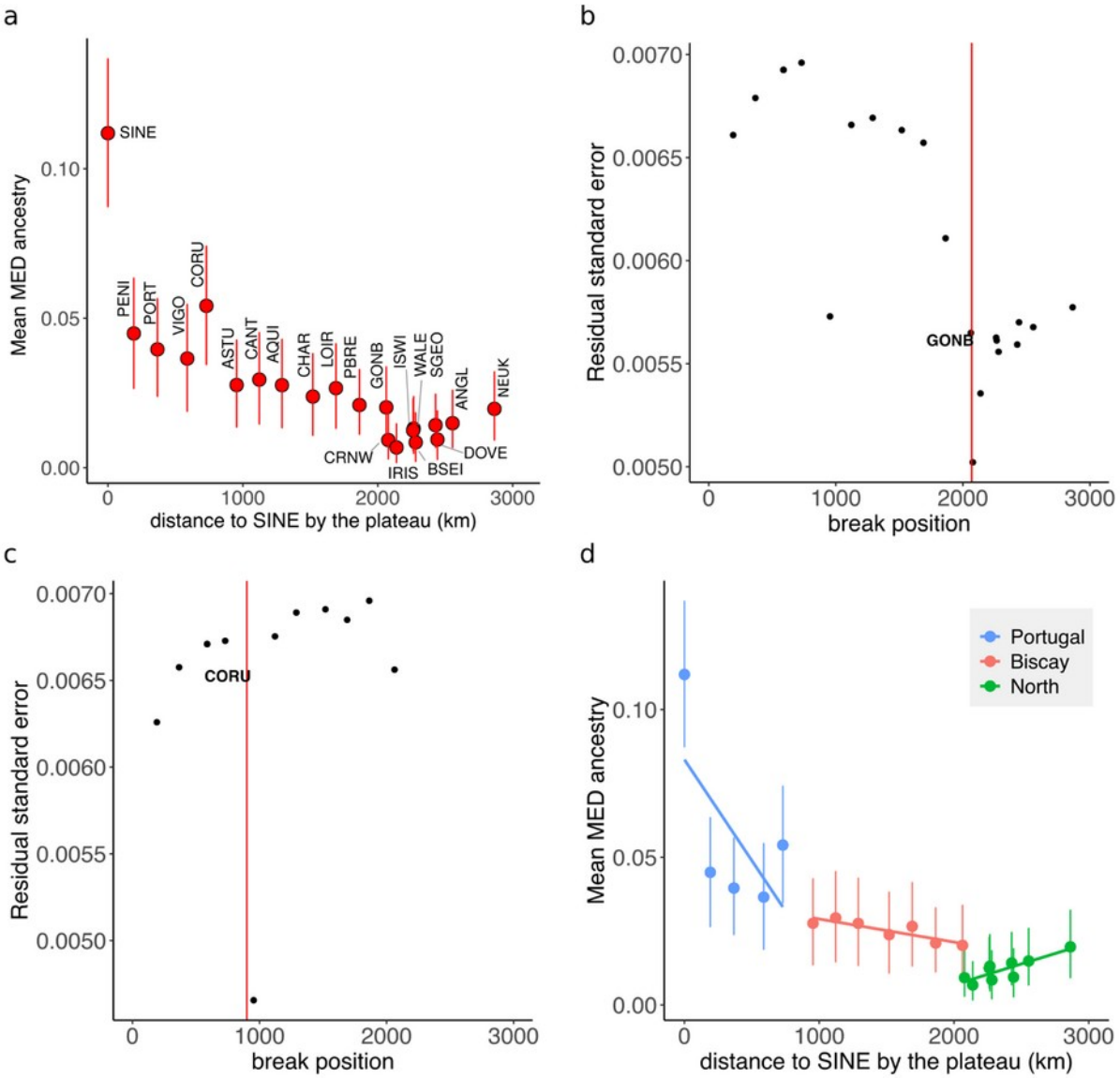

Mediterranean ancestry
0.00
0.25
0.50
0.75
1.00
MED
Portugal
0.057
[0.037,0.076]
Biscay
0.026
[0.013,0.041]
Celtic Sea – Channel – North Sea
0.011
[0.004,0.022]
Individuals
0
200
400
600
800
EMED
WMED
SINE
PENI
PORT
GALI
ASTU
CANT
AQUI
CHAR
LOIR
PBRE
GONB
CRNW
IRIS
WALE
ISWI
BSEI
SGEO
DOVE
ANGL
NEUK

Figure 3.



1046 Figure 4.

1047

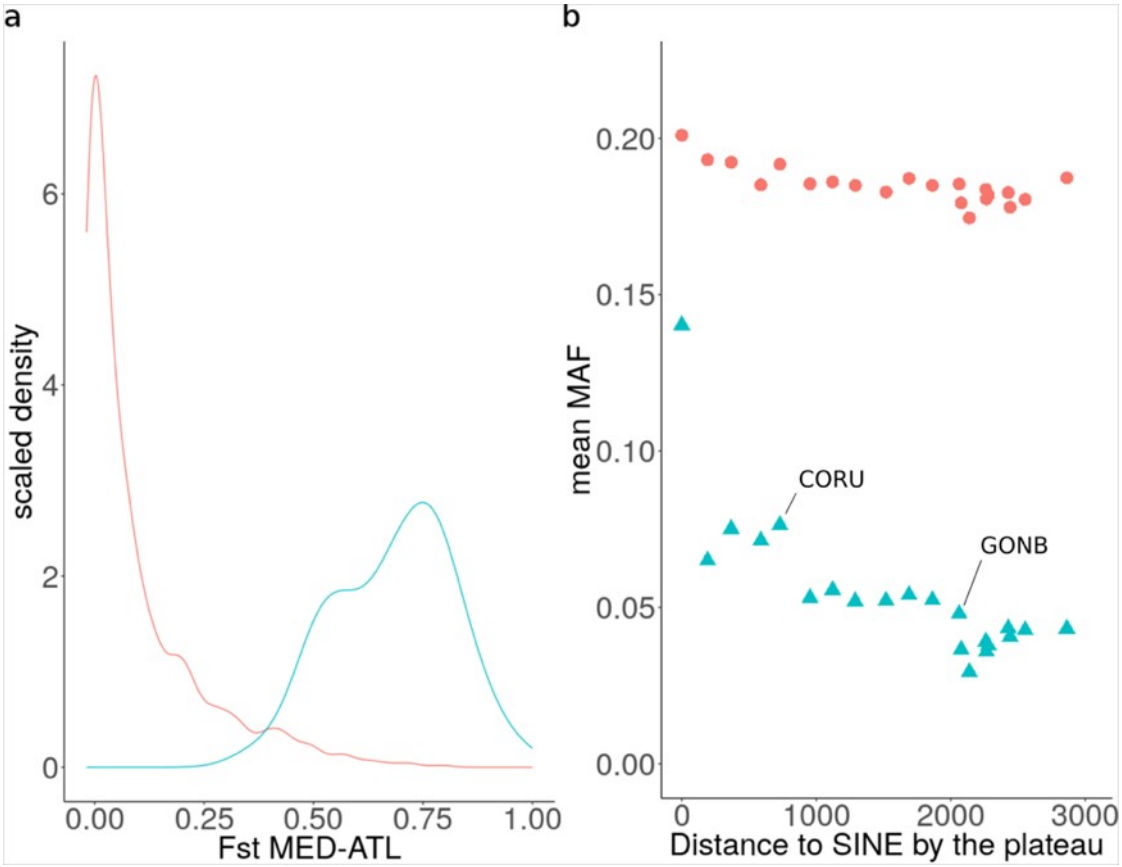

a
scaled density
0
2
4
6
0.00
0.25
0.50
0.75
1.00
Fst MED-ATL
b
mean MAF
0.00
0.05
0.10
0.15
0.20
CORU
GONB
0
1000
2000
3000
Distance to SINE by the plateau

Figure 5.