



Genomic and geographic footprints of differential introgression between two divergent fish species (*Solea* spp.)

Ahmed Souissi^{1,2,3} · François Bonhomme^{1,2} · Manuel Machado⁴ · Lilia Bahri-Sfar³ · Pierre-Alexandre Gagnaire^{1,2}

Received: 3 November 2017 / Revised: 12 February 2018 / Accepted: 10 March 2018 / Published online: 1 May 2018
© The Genetics Society 2018

Abstract

Investigating gene flow between closely related species and its variation across the genome is important to understand how reproductive barriers shape genome divergence before speciation is complete. An efficient way to characterize differential gene flow is to study how the genetic interactions that take place in hybrid zones selectively filter gene exchange between species, leading to heterogeneous genome divergence. In the present study, genome-wide divergence and introgression patterns were investigated between two sole species, *Solea senegalensis* and *Solea aegyptiaca*, using restriction-associated DNA sequencing (RAD-Seq) to analyze samples taken from a transect spanning the hybrid zone. An integrative approach combining geographic and genomic clines methods with an analysis of individual locus introgression accounting for the demographic history of divergence was conducted. Our results showed that the two sole species have come into secondary contact postglacially, after experiencing a prolonged period (*ca.* 1.1 to 1.8 Myrs) of allopatric separation. Secondary contact resulted in the formation of a tension zone characterized by strong reproductive isolation, which only allowed introgression in a limited fraction of the genome. We found multiple evidence for a preferential direction of introgression in the *S. aegyptiaca* genetic background, indicating a possible recent or ongoing movement of the hybrid zone. Deviant introgression signals found in the opposite direction suggested that *S. senegalensis* could have possibly undergone adaptive introgression that has not yet spread throughout the entire species range. Our study thus illustrates the varied outcomes of genetic interactions between divergent gene pools that recently met after a long history of divergence.

Introduction

Reproductive isolation and hybridization are antinomic but related processes that are tuned by the same evolutionary

forces. Natural hybridization between two divergent lineages occurs as long as reproductive isolation is not totally established (Dobzhansky 1937; Mayr 1942; Coyne and Orr 2004). The resulting exchange of genetic material beyond the first hybrid generation occurs through the production of backcross and recombinant genotypes that are exposed to selection, allowing neutral and advantageous alleles to spread among divergent lineages but preventing gene flow at the loci involved in reproductive isolation (Barton 1979). Hybrid zones, the geographic areas where hybridization occurs, therefore act as selective filters which allow us to unravel the strength of reproductive barriers and their impact on the patterns and dynamics of gene exchanges at various stage of speciation (Barton and Hewitt 1985; Barton and Bengtsson 1986; Hewitt 1988; Martinsen et al. 2001).

The semipermeable nature of species boundaries was first evidenced from differential introgression patterns in hybrid zones (Harrison 1990), that is, variation among loci in the level of incorporation of alleles from one lineage into the other (Payseur 2010; Harrison and Larson 2016). Introgression at a given neutral marker depends on the

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41437-018-0079-9>) contains supplementary material, which is available to authorized users.

✉ Ahmed Souissi
ahmed.souissi@gmail.com

- ¹ Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier, France
- ² CNRS—Institut des Sciences de l'Evolution, UMR5554 UM-CNRS-IRD-EPHE, Station Méditerranéenne de l'Environnement Littoral, 34200 Sète, France
- ³ Faculté des Sciences de Tunis UR11ES08 Biologie intégrative et écologie évolutive et fonctionnelle des milieux aquatiques, Université de Tunis El Manar, 2092 Tunis, Tunisia
- ⁴ IFAPA Centro El Toruño, Junta de Andalucía, Camino Tiro Pichón s/n, 11500 El Puerto de Santa María, Cádiz, Spain

antagonistic effects between counter-selection on a nearby selected locus and recombination with that barrier locus (Barton and Bengtsson 1986). Therefore, differential gene flow across the genome reflects both recombinational distance to a barrier locus and selection acting on that locus. Different methods have been developed to detect individual loci with introgression behaviors departing from the genome-wide average. This includes the analysis of spatial allele frequency patterns in a transect spanning the hybrid zone, or variation in ancestry proportion at individual loci compared to genome-wide expectations in admixed genotypes (see Payseur 2010 for a review).

The geographic cline method is a powerful approach to analyze the relationship between allele frequency and geographic distance to the center of a hybrid zone (Endler 1977; Barton and Hewitt 1985; Barton and Gale 1993; Porter et al. 1997; Teeter et al. 2008). Fitting a cline model to observed allele frequency patterns allows inferring fundamental parameters such as cline center, dispersal, and the strength of selection. Therefore, important information concerning the balance between migration and selection can be obtained using the geographic cline method, such as the identification of which loci tend to introgress neutrally and which do not. The genomic cline method is a related approach developed to deal with hybrid zones that are not clinally structured, such as mosaic hybrid zones (Harrison and Rand 1989; Bierne et al. 2003; Larson et al. 2013). The change in allele frequency at individual loci is analyzed along a gradient of genomic admixture instead of a spatial gradient, enabling a comparison of individual locus introgression patterns with regard to the genome-wide average pattern of admixture (Gompert and Buerkle 2009, 2011; Fitzpatrick 2013). As for geographic clines, this genomic cline method also provides a quantitative assessment of the excess of ancestry to a given parental species, as well as the rate of change in allele frequency from one species to the other along the admixture gradient. Although both approaches are useful to document the effect of selection and recombination on differential introgression in admixed genotypes, they do not specifically account for the demographic history of the parental species (Payseur and Rieseberg 2016).

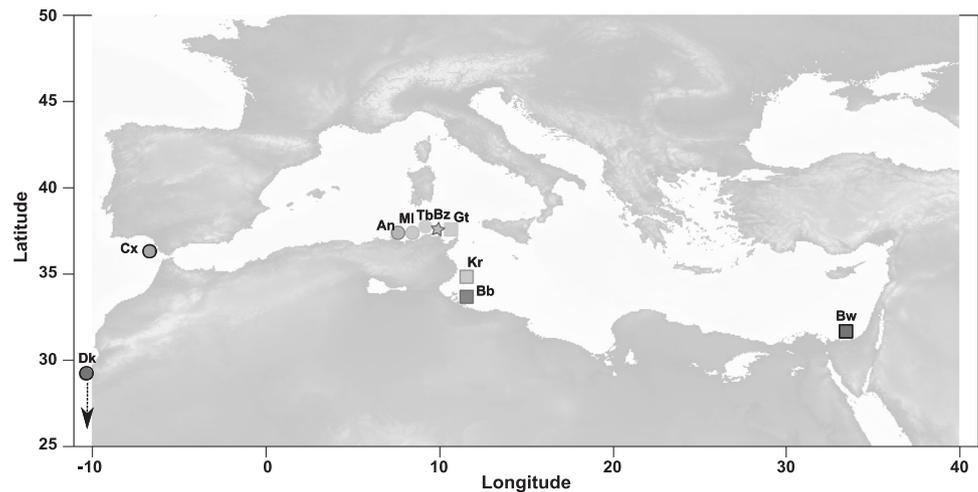
An alternative approach that partly overcomes this limitation is the use of demo-genetic models that account for varying rates of introgression among loci during the divergence history. Such methods for reconstructing the history of gene flow between semi-isolated lineages have been developed within Bayesian (Sousa et al. 2013), approximate Bayesian computation (Roux et al. 2013, 2016), or approximate likelihood frameworks (Tine et al. 2014; Le Moan et al. 2016; Rougeux et al. 2017). Using summary statistics like the joint allelic frequency spectrum,

which depicts correlations in allele frequencies between lineages outside the hybrid zone, they capture variable behaviors among loci and allow quantifying the degree of semi-permeability reflecting the overall balance between gene flow and selection. These methods have the advantage of specifically accounting for the history of gene flow during divergence, using contrasted speciation scenarios such as primary differentiation, ancient migration (AM), or secondary contact (SC). However, they do not provide information about individual locus behavior as the cline methods do. Here, we pushed these inference methods one step further in order to assess the probability for a given locus to belong to one of two categories: (i) loci with a reduced effective migration rate due to selection and linkage, and (ii) loci which can readily introgress.

The flatfish species *Solea senegalensis* and *Solea aegyptiaca* are two economically important species in the Mediterranean basin that hybridize along the northern Tunisian coasts, where they form a hybrid zone (She et al. 1987). Mitochondrial divergence between them is high (~2%) relative to other vertebrate species pairs that also hybridize in nature, such as the house mice *Mus m. musculus* and *M. m. domesticus* (0.3%, based on sequences retrieved from Genbank) or Atlantic and Mediterranean lineages of *Dicentrarchus labrax* (0.7%, Tine et al. 2014). Based on an analysis of spatial allele frequency patterns at a dozen of allozymic and intronic loci, Ouanes et al. (2011) proposed that the hybrid zone between *S. senegalensis* and *S. aegyptiaca* was centered in Bizerte lagoon, acting as a non-stable unimodal tension zone stemming from a SC. They also suggested that the zone could have undergone a recent expansion. Recently, Souissi et al. (2017) showed the existence of morphological transgressions within the contact zone, possibly indicating a reduced fitness of recombined compared to parental genotypes. However, the low number of markers in these latter studies could not provide a clear description of the genetic exchanges across the genome of these two species, nor quantifying the strength of the forces maintaining the incomplete reproductive isolation between them.

In the present work, high-throughput genotyping using restriction-associated DNA (RAD) sequencing (Baird et al. 2008) was carried out in individuals sampled at both sides of the hybrid zone, with a particular effort in the contact zone itself. By combining three different methods (genomic and geographic cline analysis and historical demographic inference) exploiting different aspects of the data, we provide new insights on the history of divergence between *S. senegalensis* and *S. aegyptiaca*, and a genome-wide description of varied patterns of introgression attributed to recent or ongoing movement of the hybrid zone.

Fig. 1 Map of the sampling locations used in the present study. *Solea senegalensis* side: Dakar (Dk), Cadiz (Cx), Annaba (An), Mellah lagoon (MI), Tabarka (Tb). Hybrid zone: Bizerte lagoon (Bz). *Solea aegyptiaca* side: Gulf of Tunis (Gt), Kerkennah Islands (Kr), El Biban lagoon (Bb), Bardawil lagoon (Bw)



Materials and methods

Sampling

A total of 161 samples were collected from 10 locations spanning the distribution range of *Solea senegalensis* and *S. aegyptiaca* from Senegal to Egypt (Fig. 1). This sampling strategy aimed at covering the geographical distribution of both species, including a detailed transect of their natural hybrid zone in Tunisia. Five sampling locations were collected throughout the *S. senegalensis* parental zone, two in the Atlantic Ocean (10 individuals from Dakar in Senegal and 16 from the Gulf of Cadiz in Spain), and three in the Western Mediterranean Sea (15 individuals from Annaba and 8 from Mellah Lagoon in Algeria, and 10 from Tabarka in Tunisia). Three locations were sampled across the geographical range of *S. aegyptiaca* in the Eastern Mediterranean Sea (13 individuals from Kerkennah Island and 15 from El Biban lagoon in Tunisia, and 20 samples from Bardawil lagoon in Egypt). Sampling size was locally increased in the Tunisian region where both species coexist and hybridization has been reported (Bizerte lagoon $n = 29$ and Gulf of Tunis $n = 25$). Finally, a total of seven individuals belonging to the closely related species *S. solea* were sampled in Tunisia to provide an outgroup species for the orientation of ancestral and derived alleles in *S. senegalensis* and *S. aegyptiaca*.

RAD library preparation and sequencing

Whole-genomic DNA was extracted from fin clips using the DNeasy Blood & Tissue kit (Qiagen). The presence of high molecular weight DNA was checked on a 1% agarose gel, and double-stranded DNA concentration was quantified using Qubit 2.0 and standardized to 25 ng per μl . RAD library construction followed a modified version of the

original single-end RAD-Seq protocol (Baird et al. 2008). Briefly, 1 μg of genomic DNA from each individual was digested using the restriction enzyme *Sbf*I-HF (NEB), and ligated to one of 32 unique molecular barcodes of 5–6 bp. Ligated products were then combined in equimolar proportions into six RAD libraries, each made of a multiplex of 32 individuals originating from various localities. Each library was finally sequenced in 101 bp single read mode on a separate lane of an Illumina HiSeq2500 sequencer, at the sequencing platform “Génomique Intégrative et Modélisation des Maladies Métaboliques” (UMR 8199, Lille, France).

Bioinformatic analyses

Raw reads were de-multiplexed based on individual barcode information and subsequently end-clipped to 95 bp to homogenize read length after removing barcodes of different lengths. Read quality filtering was performed using the sliding window approach implemented in the module *process_radtags* from the *stacks* pipeline (Catchen et al. 2013). This allowed us to exclude reads in which the average quality of 15 adjacent bases fell below a raw phred score of 10. Retained reads were then aligned to a draft assembly of the *S. senegalensis* genome (98,590 scaffolds, total length 740 Mb, N50 contig length 10,767 bp, Manchado et al. 2016) using *bowtie 2v.2.1.0* (Langmead and Salzberg 2012) with the *--very-sensitive* option in *--end-to-end* mode. In order to take into account both the level of divergence among species and the possibility of introgression while accounting for hidden paralogy, we used a subset of individuals to empirically determine the optimal maximum number of mismatches allowed between aligned reads and the reference genome, following a procedure described previously (Le Moan et al. 2016; Rougemont et al. 2017; Rougeux et al. 2017). We found that a maximum of seven

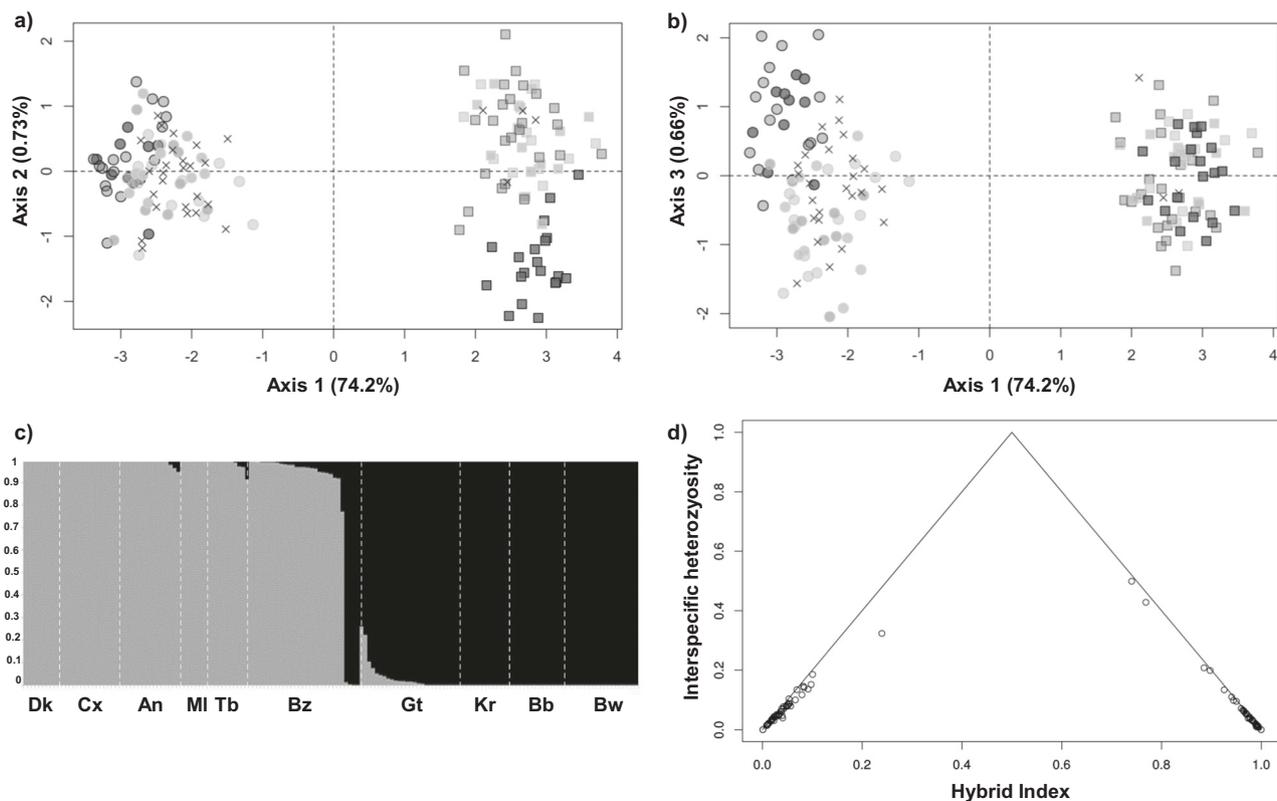


Fig. 2 Genetic structure of *S. senegalensis* and *S. aegyptiaca* analyzed using 10,758 SNPs, showing introgressive hybridization between species. Principal component analysis (PCA) based on 161 individuals representing **a** individual coordinates along PC1 and PC2 axes and **b** along PC1 and PC3 axes. Legend: ● Dk; ○ Cx; ● An; ● MI; ● Tb; X Bz; X Gt; ● Kr; ● Bb; ● Bw. **c** Triangle plot showing individual interspecific heterozygosity against hybrid index. **d** Result of *fastStructure* analysis performed for $K = 2$ genetic clusters

mismatches for a 95 bp RAD tag offered the best compromise to correctly align both species and the outgroup to the *S. senegalensis* genome. This was set to *bowtie 2* options using a minimum alignment score of $--score-min = -42$ needed for an alignment to be considered as valid, fixing the penalty of mismatch bases to $MX = 6$ and using default gap penalties. We then used *pstacks* to call variable positions under the bounded SNP model, setting the upper sequencing error rate to 2.5% and a minimum sequencing depth to $5\times$ per stack. Homologous loci across samples were merged based on their genomic position within scaffolds using *cstacks* to construct a catalog of loci. Individual stacks were then matched against the catalog of loci with *sstacks* to determine genotypes. The module *rxstacks* was ran with the $--prune_haplo$ option to exclude poor-quality loci with multiple sequencing errors using a log-likelihood threshold of -300 (determined empirically, Fig. S1). Individual genotypes were finally exported in the VCF format using the module *populations*.

We then applied population-specific filters with *VCFtools* (Danecek et al. 2011) to remove SNPs showing significant deviation to Hardy–Weinberg equilibrium within at least one of the *S. senegalensis* or *S. aegyptiaca* locality

samples located outside the hybrid zone, using a P -value threshold of 0.01. Next, we excluded loci displaying more than 20% of missing genotypes in at least one locality. Over the 116,385 remaining SNPs, we randomly selected one single SNP for each pair of RAD loci associated with the same restriction site (i.e., within a distance less than 200 bp), in order to limit the impact of linkage disequilibrium. Finally, we only retained loci with available sequence data in the outgroup species *S. solea*, which had to be fixed for the polymorphic site found in *S. senegalensis* and *S. aegyptiaca*. This resulted in a final dataset containing 10,758 independent SNPs that were specifically filtered to comply with the *δadi* analysis requirements. The same SNPs were used in all subsequent analyses unless specified otherwise.

Genetic structure and hybridization

Genetic variation within and among *S. senegalensis* and *S. aegyptiaca* was characterized with a principal component analysis (PCA) performed with the R package *Adegenet* (Jombart 2008; Jombart and Ahmed 2011). In order to document the possible existence of population structure

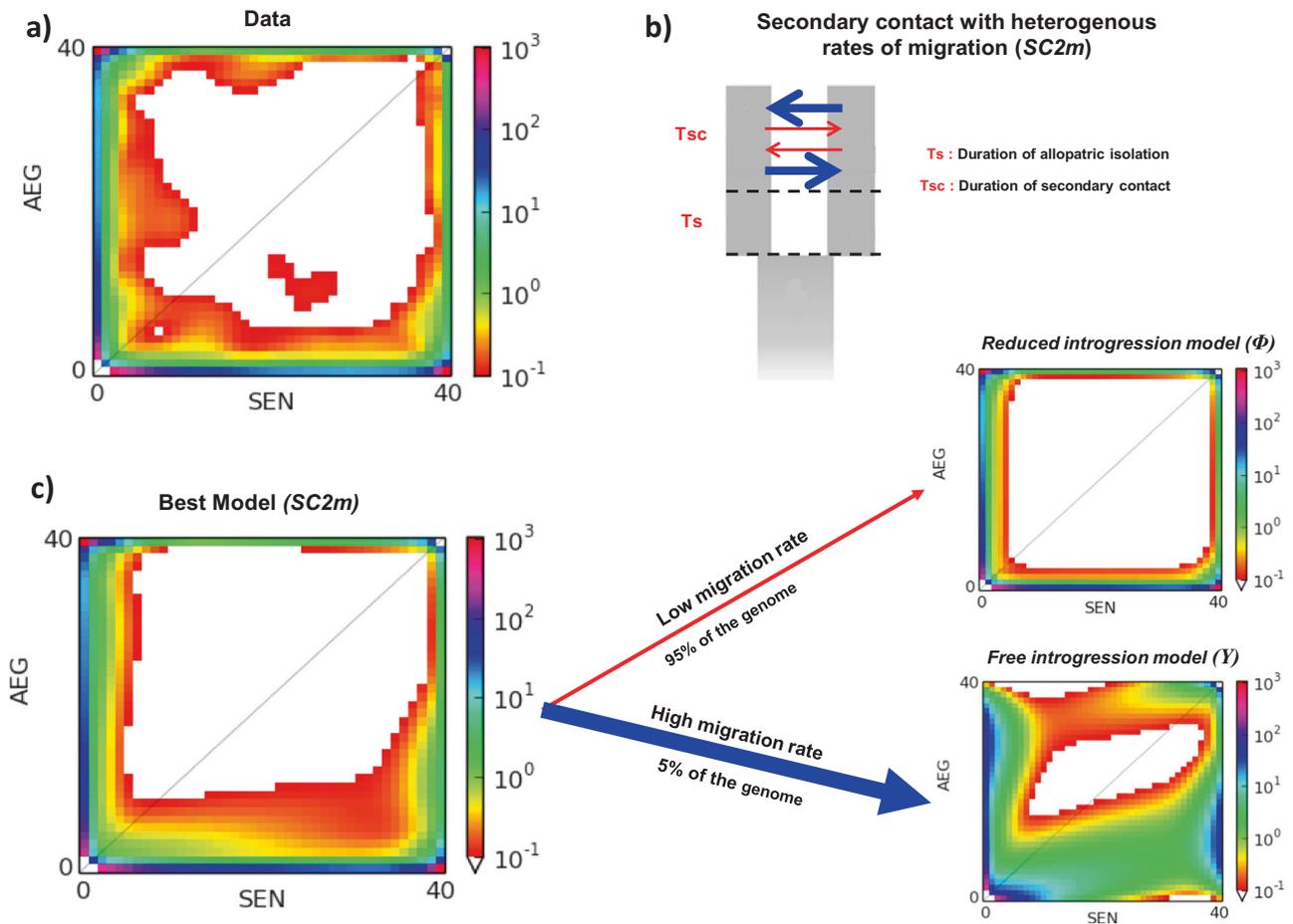


Fig. 3 Detection of introgression based on the inferred demographic divergence history of *S. senegalensis* and *S. aegyptiaca*. **a** The observed joint allele frequency spectrum (JAFA) for *S. aegyptiaca* (AEG, y-axis) and *S. senegalensis* (SEN, x-axis) showing the number of SNPs (colored scale on the right side of the spectrum) per bin of derived allele counts using 20 individuals per species. **b** Schematic representation of the secondary contact model with heterogeneous

introgression rates among loci (SC2m model). **c** The maximum-likelihood JAFA obtained under the best-fit SC2m model, which can be decomposed into a *reduced introgression model* (Φ) explaining 95% of the genome and a *free introgression model* (Y) accounting for 5% of the genome. The JAFA of the two models Φ and Y obtained from *msms* simulations, illustrating which bins are likely to be attained preferentially under the *free* vs. *reduced introgression* model

within each species and distinguish parental and admixed genotypes, we estimated individual admixture proportions from two to four differentiated genetic clusters that were inferred from the data with no a priori information on individual membership, using *fastStructure* (Raj et al. 2014) with 10^8 iterations. In addition, we estimated the proportion of alleles with *S. aegyptiaca* ancestry for each individual, using the R package *Introgess* (Gompert and Buerkle 2010) to estimate a Hybrid Index. We then searched specifically the presence of early-generation hybrids using *NewHybrids* (Anderson and Thompson 2002) with a subset of 296 diagnostic SNPs showing fixed differences between species. For each individual, we estimated the membership probability to each of six categories of pedigree including pure parental species (P_{sen} and P_{aeg}), first and second-generation hybrids (F1 and F2), and first-generation

backcrosses in each direction (BC_{sen} and BC_{aeg}), using a total of 10,000 burn-in steps and 50,000 iterations.

Demo-genetic inference of the divergence history

We used a modified version of the *delta di* program (Gutenkunst et al. 2009) that improves likelihood optimization using a simulated annealing (SA) procedure (Tine et al. 2014) in order to infer the divergence history between *S. senegalensis* and *S. aegyptiaca*. This method uses the joint allele frequency spectrum (JAFA) between two populations as summary statistics to characterize divergence. The program *delta di* fits demographic divergence models to the observed data using a diffusion approximation of the JAFA, enabling a comparison of different alternative divergence models in a composite likelihood framework. We used

seven divergence models developed in a previous study (Tine et al. 2014) to determine whether and how gene flow has shaped genome divergence between *S. senegalensis* and *S. aegyptiaca*. The simplest model of strict isolation (SI) corresponds to an allopatric divergence scenario in which the ancestral population of effective size N_A splits into two derived populations of size N_1 (*S. senegalensis*) and N_2 (*S. aegyptiaca*) that evolve without exchanging genes for T_S generations. We then considered three models of divergence including gene flow, either during the entire divergence period (isolation with migration, IM), the beginning of divergence (AM), or the most recent part of divergence (SC). In these models, migration occurs with potentially asymmetric rates (m_{12} and m_{21}) that are shared across all loci in the genome. We also considered simple extensions of the IM, AM, and SC models that capture the effect of selection by accounting for heterogeneous migration rates among loci (IM2m, AM2m, and SC2m). These semi-permeability models consider that two categories of loci, experiencing different effective migration rates, occur in proportions P and $1-P$ in the genome.

The JAFS was obtained by pooling the least introgressed populations from *S. senegalensis* (Dakar, Cadiz, Annaba, and Mellah) for species 1 and *S. aegyptiaca* (Kerkennah, El Biban, and Bardawil) for species 2 (Fig. 2), in order to avoid including admixed genotypes causing departures to the underlying population model. We also tested the robustness of our inferences with respect to the existence of a subtle genetic structure within each species (Fig. 2a, b) by considering only Annaba and Mellah for *S. senegalensis* (species 1) and Kerkennah and El Biban for *S. aegyptiaca* (species 2).

We used *S. solea* samples as an outgroup species to determine the most parsimonious ancestral allelic state for each SNP in order to generate an unfolded JAFS oriented with the derived allele (Fig. 3a). The size of the JAFS was projected down to 40 sampled chromosomes per species to account for missing data. For each model, we estimated the parameter values that maximize likelihood using two successive SA procedures before quasi-Newton (BFGS) optimization (Tine et al. 2014). Comparisons among models was made using the Akaike information criterion (AIC) to account for variation in the number of parameters among models. A total of 20 independent runs were used for the optimization of each model. Parameter uncertainties were estimated from non-parametric bootstrapped data using the Godambe information matrix as implemented in *deltai*. We used 1000 bootstrapped datasets to estimate confidence intervals as the maximum likelihood parameter value $\pm 1.96 \times \text{SE}$. Estimated parameter values were converted into biologically meaningful quantities by taking into account our SNP filtering procedures following the method in Rougeux et al. (2017).

Inferring the probability of locus introgression under the best divergence model

Because the JAFS-based inference of the divergence history does not provide an assessment of introgression probability for each locus separately, we developed an approach to estimate the relative probability of each individual locus to be assigned to one of these two categories: (i) loci which can readily introgress between species, and (ii) loci experiencing a highly reduced introgression rate due to selection against foreign alleles and linkage. This probability was estimated using the best-fit model identified in the previous section, which was a SC model with variable introgression rates among loci (SC2m) (Fig. 3b). The SC2m model can be decomposed as a linear combination of two simple models describing gene flow in two different compartments of the genome (Fig. 3c). The best-fit SC2m model estimated that only 5% of the loci can still introgress between species with effective migration rates m_{1-2} and m_{2-1} , whereas the remaining 95% of loci experience a highly reduced introgression rate with effective migration rates m'_{1-2} and m'_{2-1} . We used estimated model parameters to perform coalescent simulations with *msms* (Ewing and Hermisson 2010) under the SC model (assuming $\theta = 747.202$; $N_1 = 0.818$; $N_2 = 1.136$; $T_S = 4.803$; $T_{SC} = 0.081$) using two different conditions for gene flow (assuming either $m_{1-2} = 4.607$ and $m_{2-1} = 0.381$, or $m'_{1-2} = 0.057$ and $m'_{2-1} = 0.153$) to generate a *free introgression* and a *reduced introgression* rate dataset. One thousand JAFS with 40 sampled chromosomes per species were produced under each model. We then averaged the number of derived alleles within entries across replicates to obtain a single JAFS for both the *free introgression* (Y) and the *reduced introgression* (Φ) model (Fig. 3c). Finally, we used the average number of SNPs per entry (i, j) in each JAFS to estimate the probability $P_{i,j}$ that a SNP with a derived allele count i in species 1 and j in species 2 can be obtained under the *free introgression* model:

$$P_{i,j} = 0.05 \times Y_{i,j} / (0.05 \times Y_{i,j} + 0.95 \times \Phi_{i,j}), \quad (1)$$

where $Y_{i,j}$ and $\Phi_{i,j}$ are the average number of SNPs predicted in the JAFS entry (i, j) under the *free introgression* and *reduced introgression* model, respectively. Every SNP from the real dataset was finally associated to the introgression probability $P_{i,j}$ given by Eq. (1) based on its corresponding entry in the JAFS.

Genomic clines

The Bayesian Genomic Cline program *BGC* (Gompert and Buerkle 2012) was used to quantify individual locus introgression relative to genome-wide introgression. *BGC* describes the probability of locus-specific ancestry from one

parental species given the genome-wide hybrid index. The *BGC* model considers two principal parameters, called α and β , which describe locus-specific introgression based on ancestry. Parameter α quantifies the change in probability of ancestry relative to a null expectation based on genome-wide hybrid index, that is, the direction of introgression. A positive value of α reflects an increase in the probability of ancestry from species 1 (introgression into *S. aegyptiaca*), whereas a negative value of α reflects an increase in the probability of ancestry from species 2 (introgression into *S. senegalensis*). Parameter β describes the rate of transition in the probability of ancestry from parental population 1 to parental population 2 as a function of hybrid index, that is, the amount of introgression. Positive β values thus denote a restricted amount of introgression, whereas negative β values indicate a greater introgression rate compared to the genome-wide average.

We ran *BGC* under the genotype-uncertainty model, assuming a sequence error probability of 0.0001 and using information from the data to initialize ancestry and hybrid index. Two MCMC chains each made of 150,000 steps were ran, recording every 20th value. Cline parameter quantiles were calculated to designate outlier loci with respect to parameters α and β based on the assumption that the genome-wide distributions of locus-specific cline parameters are both centered on zero. Therefore, outlier loci are markers with extreme patterns of introgression relative to the remainder of the genome. A locus was considered as an α outlier if its posterior estimate of α was not contained in the interval bounded by the 0.025 and 0.975 quantiles of $\mathcal{N}(0, \tau_\alpha)$ (likewise for β) (Gompert and Buerkle 2011; Gompert et al. 2013).

Geographic clines

The geographic cline analysis was carried out in order to link allele frequencies at individual loci with geographic position along a transect spanning the hybrid zone. We used the R package *HZAR* (Derryberry et al. 2014) that fits allele frequency data to classic equilibrium models of geographic clines (Szymura and Barton 1986) using the MCMC algorithm. Instead of searching for the best model separately for each cline, we used the full model that fits cline center, width, and independent introgression tails using estimated values for minimum (p_{\min}) and maximum (p_{\max}) allele frequencies. In this way, we avoided the potentially confounding effects of fitting different models for the comparison of cline center and slope parameters across loci.

Results

A total of 833.3 million raw reads were obtained, 89.1% of which were retained after demultiplexing and quality

filtering for reference mapping against the *S. senegalensis* draft genome (average number of reads per individual: 3,906,809, s.d.=1,878,685). Individual genotype calling in *Stacks* produced a raw VCF file containing 174,490 SNPs, from which we retained 10,758 oriented SNPs (using *S. solea* as an outgroup to identify ancestral states) after controlling for linkage and filtering for quality. The mean sequencing depth per locus per individual was superior to 100× (Fig. S2), and the mean genotype missing rate was 1.3% (s.d. = 2.6%) per individual and 1.4% (s.d. = 1.8%) per locus. The final VCF containing 10,758 SNPs was used for all downstream analyses except *NewHybrids*.

Genetic structure, hybridization, and introgression

The PCA clearly separated *S. senegalensis* from *S. aegyptiaca* samples along the first PC axis (PC1), which explained 74.2% of the total genotypic variance (Fig. 2a). *S. senegalensis* samples were organized along that axis according to their geographical proximity from the contact zone, following a gradient of genetic similarity to *S. aegyptiaca* increasing from Dakar to Bizerte lagoon. The two species were found to coexist only in Bizerte lagoon, with four *S. aegyptiaca* individuals being found among a majority of *S. senegalensis* genotypes. The second principal component revealed a weak signal of differentiation (PC2, 0.73% of associated variance) separating Bardawil lagoon from other samples within *S. aegyptiaca*. Likewise, the third principal component (PC3, 0.66% of genotypic variance) captured a subtle differentiation signal between the Atlantic samples from Dakar and Cadiz and the Mediterranean *S. senegalensis* samples (Fig. 2b). The *fastStructure* analysis (Fig. 2c) confirmed the separation of the two species into different genetic clusters and their coexistence in Bizerte lagoon, as detected from the PCA. However, it failed to detect any further genetic subdivision (Fig. S3), confirming that the signal of genetic structure within each species is at most very small.

The finding of intermediate admixture proportions in some individuals revealed signs of introgressive hybridization between *S. senegalensis* and *S. aegyptiaca* around the contact zone.

The relationship between individual hybrid index and interspecific heterozygosity represented in a triangle plot illustrated well the absence of F1 and F2 hybrids in our dataset, and the presence of a few backcrosses together with several introgressed genotypes (i.e., late-generation backcrosses) (Fig. 2d). This was confirmed by the detection of three first generation backcrosses by *NewHybrids* (detected with an assignment probability of 1), one in the direction of *S. senegalensis* (in Bizerte lagoon) and two in the opposite direction (in Gulf of Tunis), plus two unassigned individuals likely representing later-generation backcrosses (Fig.

Table 1 The demographic divergence history of *S. senegalensis* and *S. aegyptiaca* is best explained by a secondary contact model with heterogeneous introgression rates among loci

Model	AIC	Log-likelihood	<i>theta</i>	<i>nuA</i>	N_1	N_2	m_{12}	m_{21}	m'_{12}	m'_{21}	T_S	T_{AM}	T_{SC}	P	O	
<i>SI</i>	4611.53	-2301.76	2569.18	1	0.226	0.287	-	-	-	-	0.571	-	-	-	0.988	
<i>IM</i>	3049.93	-1518.96	1018.56	1	0.572	0.768	0.035	0.053	-	-	3.809	-	-	-	0.981	
<i>AM</i>	3048.24	-1517.12	514.41	1	1.14	1.53	0.017	0.026	-	-	8.55	0.0002	-	-	0.977	
<i>IM2m</i>	3041.33	-1511.66	1491.59	1	0.395	0.525	0.110	6.635	0.046	0.074	2.412	-	-	0.054	0.982	
<i>AM2m</i>	3041.03	-1510.51	1207.61	1	0.477	0.649	0.846	0.308	0.033	0.058	3.10	0.0001	-	0.051	0.982	
<i>SC</i>	2564.84	-1275.42	2160.72	1	0.291	0.354	0.215	0.521	-	-	1.011	-	0.023	-	0.981	
<i>SC2m</i>	2424.46	-1202.23	747.20	1	0.818	1.137	4.607	0.381	0.057	0.153	4.803	-	0.081	0.05	0.971	
Standard deviation (SC2m model)					0.185	0.215	0.196	0.488	0.302	0.317	0.210	0.209	-	0.231	0.247	0.003

In order of appearance in the table: the model fitted, the Akaike information criterion (AIC), the maximum-likelihood estimate, and the theta parameter for the ancestral population before split obtained over 20 independent runs. Following are the inferred values for the model parameters (scaled by theta): the effective population size of *S. aegyptiaca* (N_1) and *S. senegalensis* (N_2), the effective migration rates for the two categories of loci in the genome (category 1: m_{12} and m_{21} , category 2: m'_{12} and m'_{21}), the duration of the split (T_S), of ancestral migration (T_{AM}) and secondary contact (T_{SC}) episodes, and the proportion (P) of the genome falling within category 1 (i.e., experiencing migration rates m_{12} and m_{21}). The m_{12} and m'_{12} parameters specify migration rate from population 2 (*S. senegalensis*) into 1 (*S. aegyptiaca*) and reciprocally for m_{21} and m'_{21} . The inferred proportion of correctly orientated SNPs (O) using *S. solea* as an outgroup appears in the last column. Model description: *SI* strict isolation, *IM* isolation with migration, *AM* ancient migration, *SC* secondary contact, *IM2m* isolation with heterogeneous migration, *AM2m* ancient migration with heterogeneous migration, *SC2m* secondary contact with heterogeneous migration. The last line indicates the standard deviation of estimated parameter values for the best-fit SC2m model

S4). Therefore, our results provide evidence for contemporary introgressive hybridization between the two sole species.

Demo-genetic history of divergence

The $\delta a \delta i$ analysis showed that the SC model with varying introgression rates along the genome best explained the observed JAFS (SC2m, delta AIC with the second ranked model >25). In comparison, the six other models had a significantly lower performance for different reasons. While the SI model could explain SNPs occupying the outer frame of the JAFS, it could not predict at the same time the presence of loci in the more inner part of the spectrum. The IM, AM, and SC models assuming genome-wide homogeneous migration rates better predicted loci occupying the central part of the JASF. However, they underestimated the density of private and highly differentiated SNPs. Finally, the three models including heterogeneous migration rates along the genome (IM2m, AM2m, and SC2m) yielded significantly improved fits. Among them, only the SC2m model provided a good prediction for both a high density of highly differentiated SNPs between the two species and the presence of SNPs toward the central part of the spectrum. Model selection was not sensible to the potential effect of unaccounted structure within each species (Table S1), consistent with the very weak signals of within-species differentiation detected in PCA analysis.

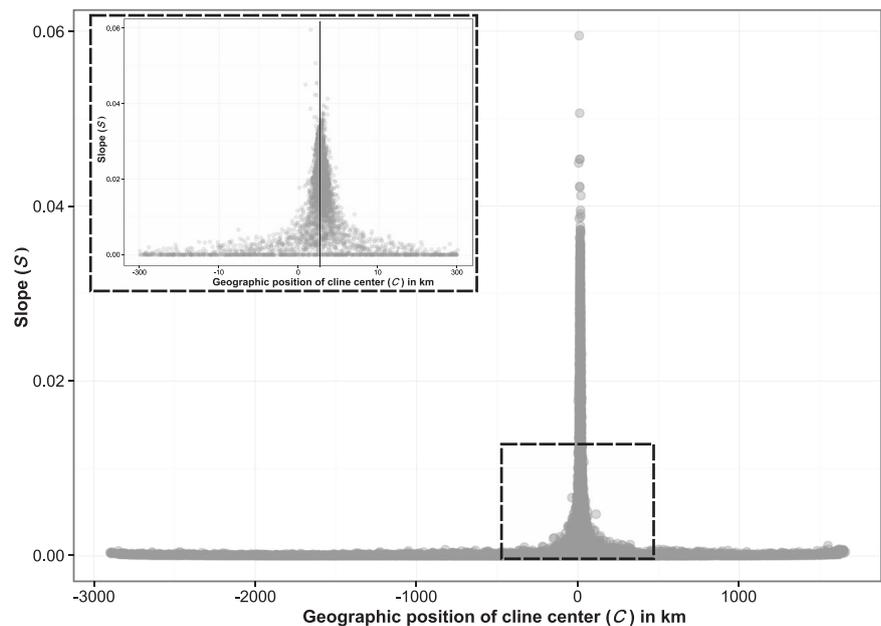
Using the best-fit obtained for the SC2m model over 20 independent runs to get estimates of model parameters and

confidence intervals, we found that the duration of the isolation period between *S. senegalensis* and *S. aegyptiaca* was about 60 times longer than the duration of the SC. A large proportion of the genome (~95%) was associated with relatively small effective migration rates and limited gene flow corresponding to less than one effective migrant per generation in both directions ($N_1 m'_{12} = 0.047$, $N_2 m'_{21} = 0.175$). By contrast, the effective migration rate was more elevated in the remaining small fraction of the genome (~5%), especially in the direction from *S. senegalensis* to *S. aegyptiaca* where introgression was found 80 times higher than elsewhere in the genome. Introgression was found 12 times lower in direction of *S. senegalensis*, although still more than twice as high than in the remainder (95%) of the genome (Table 1).

Genomic clines

The genomic cline analysis performed with the *BGC* program revealed different behaviors among the 10,758 SNPs with respect to the cline parameters α and β . Considering the locus-wise ancestry shift parameter α , we observed that 48% of the loci have an excess of *S. senegalensis* ancestry (negative α), whereas only 29% of them have an excess of *S. aegyptiaca* ancestry (positive α). The quantile method for outlier detection confirmed that a higher proportion of loci was characterized by an excess of *S. senegalensis* ancestry (43% of negative α outliers) compared to *S. aegyptiaca* ancestry (17% of positive α outliers). By contrast, the locus-wise slope parameter β depicting the rate of transition

Fig. 4 Distribution of geographic cline slope (S) for the 10,758 loci as a function of their cline center (C). The box represents a zoom in the central part of the contact zone. Cline centers falling outside this central region mostly correspond to rare variants with low information content



between species was symmetrically distributed between negative and positive values, with equal proportions of loci showing a decreased (48% of positive β) and increased (51% of negative β) introgression rates. The proportion of loci showing significant deviation compared to the genomic average amounted to 52%, and were equally distributed among negative (26%) and positive (26%) β outliers.

Geographic clines

The mean position of cline center (C) calculated over all fitted clines was located 12 km to the east of Bizerte lagoon (Fig. 4). Most individual locus clines tended to co-localize at this position, especially for the steepest clines whose centers localized essentially 10 km or less around the center of the hybrid zone. In general, loci whose individual centers did not coincide with the central part of the hybrid zone harbored less steep clines, either because they correspond to rare variants with low information content (i.e., that do not form a cline) or due to softer selection at these loci which are uncoupled from the hybrid zone (Fig. 4).

Links between different approaches

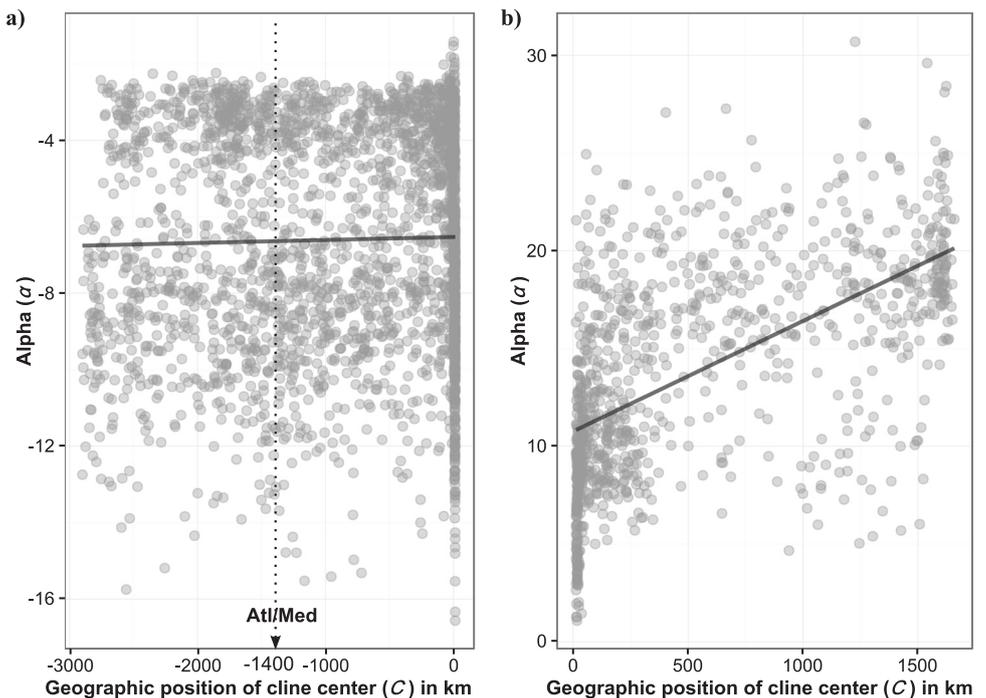
Linear regression models were used to investigate whether different analyses of individual locus introgression that use different aspects of the data tend to produce similar results. First, we evaluated the extent to which an excess of ancestry from a given species relates to a spatial shift of cline center into the other species range. The genomic cline parameter α showed a significant positive correlation with the geographic cline center C on both sides of the hybrid zone (Fig.

5). The correlation was, however, much stronger in the *S. aegyptiaca* side ($R^2_{C-\alpha} = 0.274$, $P < 10^{-10}$) compared to the *S. senegalensis* side where it was barely detectable ($R^2_{C-\alpha} = 0.023$, $P < 10^{-10}$).

We then tested whether the loci exhibiting the most abrupt transition in ancestry between the two species (i.e., a decreased introgression rate) also tended to display the steepest geographic clines. To validate this prediction, we focused on the 26% of positive β outliers detected with *BGC*, and compared β values with the slope parameter S of the geographic clines. We found a significantly positive correlation ($R^2_{S-\beta} = 0.27$, $P < 10^{-10}$) between the two parameters, confirming that loci with a low introgression rate also tend to display steeper geographic clines (Fig. 6).

Moreover, we tried to connect genomic and geographic cline parameters with our estimated probability that individual loci belong to the small fraction of the genome showing the highest introgression rate. As expected, geographic cline slope was significantly negatively correlated with the inferred probability of introgression ($R^2_{S-Prob_a} = 0.04$, $P < 10^{-10}$, Fig. S5). Therefore, private SNPs located on the outer frame of the JAFS, for which we inferred a small probability of introgression, were usually associated with steep cline slopes. By contrast, loci occupying the most central part of the JAFS, which were assigned the highest introgression probabilities, were also characterized by shallower geographic clines. Finally, when restricting the analysis to the 5% of loci with the highest introgression probability, we show that a majority of their geographic cline centers are located outside the contact zone, with a spatial shift more pronounced in the *S. aegyptiaca* direction (Fig. 7a). Similarly, the distribution of the genomic cline

Fig. 5 Correlation between genomic cline parameter (α) and geographic position of the cline center (C) on both sides of the hybrid zone: **a** in the *Solea senegalensis* geographic range, with the Atlantic/Mediterranean transition zone indicated by the dotted arrow, and **b** in the *Solea aegyptiaca* geographic range. Negative α values indicate an excess of *S. aegyptiaca* ancestry, whereas positive α values indicate an excess of *S. senegalensis* ancestry



parameter α also showed a deficit of values around zero, and a majority of loci with an excess of *S. senegalensis* ancestry, in keeping with the general asymmetry of the exchanges between the two species (Fig. 7b).

Discussion

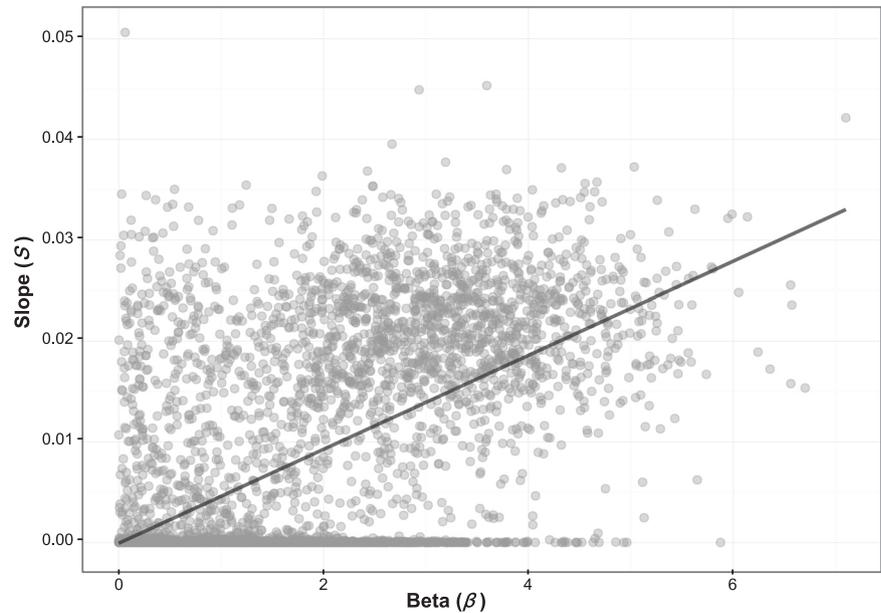
Divergence history and semi-permeability to gene flow

Our results using high-throughput genotyping confirmed previous observations, based on a limited number of nuclear markers, that the flatfish *S. senegalensis* and *S. aegyptiaca* are genetically divergent sibling species, which are still exchanging genes across their natural hybrid zone near Bizerte lagoon (She et al. 1987; Ouanes et al. 2011; Souissi et al. 2017). However, the proportion of hybrid genotypes detected here in the contact zone (i.e., 3 early generation hybrids among 54 individuals) is much lower than established in previous studies. Although this could be due to large temporal variations in the rate of hybridization, the most plausible hypothesis is that real hybrids (e.g., F1, F2, and first generations backcrosses) are more easily distinguished from introgressed genotypes using a high number of diagnostic markers (Pujolar et al. 2014; Jeffery et al. 2017), whereas earlier studies have used principally non-diagnostic markers. Our results thus support the existence of an abrupt geographic transition between the two species across their contact zone, where predominantly parental

genotypes occur in sympatry with an overall low frequency of hybrids. These observations are consistent with the tension zone model (Barton and Hewitt 1985), in which the hybrid zone is maintained by a balance between the influx of parental genotypes from outside the zone and the counter-selection of hybrid genotypes inside the zone. This interpretation is also in keeping with the finding of transgressive body shape variation attributed to a reduced condition index of admixed genotypes from the contact zone (Souissi et al. 2017).

The regions located outside the contact zone, which contain only parental genotypes, provided relevant information to infer the demo-genetic history of divergence between these two species. Gene exchange between *S. senegalensis* and *S. aegyptiaca* is best explained by a SC model with heterogeneous rates of introgression along the genome. Similar divergence histories have already been found between geographical lineages or ecotypes of the same species, like in the European sea bass and anchovy (Tine et al. 2014; Le Moan et al. 2016). Using the same approach as implemented here, these studies could determine that a minor but significant fraction of the genome (i.e., 20–35%) does not neutrally introgress between hybridizing lineages/ecotypes, thus providing a quantitative assessment of the degree of semi-permeability to gene flow. In the present case, by contrast, we found that the great majority of the genome (approximately 95%) experiences a highly reduced effective migration rate (i.e., <1 migrant per generation) between species. Moreover, the inferred divergence period estimated using a mutation rate of 10^{-8} per bp

Fig. 6 Correlation between the geographic cline slope parameter (S) and the genomic cline parameter β for positive β outlier loci, showing significantly restricted introgression rates compared to the genome-wide average



per year and a generation time of 3–5 years corresponds to a separation time of *ca.* 1.1 to 1.8 Myrs, which together with the mitochondrial sequence divergence of 2%, indicates a relatively ancient speciation event. This timing of divergence is much older than the approximately 300 Kyr that were inferred between glacial lineages in anchovy and sea bass (Tine et al. 2014; Le Moan et al. 2016). Therefore, our results are consistent with the prediction that more anciently diverged species should display stronger reproductive isolation due to the establishment of more genetic barriers (Orr 1995; Moyle and Nakazato 2010; Roux et al. 2016), which is reflected here by a lower level of permeability to gene flow compared to sea bass and anchovy.

Another aspect of this low miscibility of the two genomes was captured by individual locus cline parameters. Consistent with the view that stepped clines generate a strong barrier to gene flow (Szymura and Barton 1986), we found a high proportion of loci combining extremely low introgression rates (i.e., positive β outliers) and steep geographic clines (i.e., cline width below the average) with centers colocalized in the central part of the hybrid zone. Because this type of situation generates strong linkage disequilibrium, each locus is expected to receive, in addition to its own selective coefficient, the effect of selection on other loci (Barton 1983; Kruuk et al. 1999). This makes the genome acting almost as a single underdominant locus causing a global reduction in gene flow, which sometimes is referred to as a “congealed genome” (Turner 1967; Bierne et al. 2011; Gompert et al. 2012b) relatively to its very low capacity to incorporate foreign genes. In contrast to this genome-wide reduction in gene flow, we also found support for a small proportion of genomic regions with higher-than-average introgression rates between species.

Multiple approaches to detect increased introgression rates

Differential introgression among loci was evidenced by combining several approaches that capture different but complementary aspects of the data. For instance, genomic and geographic cline methods both rely on the sampling of admixed genotypes from the hybrid zone, whereas the modeling approach based on the JAFS relies on the joint distribution of allele frequencies between parental populations located outside of the hybrid zone. By combining these methods, we found that the 5% of loci with higher-than-average introgression rates identified with the JAFS approach generally display shifts in both geographic and genomic cline centers, as illustrated by their tendency to have non-zero cline center and α parameter values (Fig. 7). Therefore, our method to detect increased introgression while accounting for the demographic divergence history identifies genomic regions that also depart from the genome average in those places where the two species meet and admix.

Discordant clines were not symmetrically distributed between the *S. senegalensis* and *S. aegyptiaca* side of the hybrid zone. The majority of the highly introgressed loci were shifted into the *S. aegyptiaca* geographic range. For these loci, the excess of *S. senegalensis* ancestry (measured by the genomic cline parameter α) was positively related to the extent of the spatial shift of the cline center into the *S. aegyptiaca* territory. This correlation was much stronger than previously observed in a bird hybrid zone study (Grossen et al. 2016), possibly due to a stronger variance in cline shift in the *Solea* system because of hybrid zone movement (see below). By contrast, much fewer loci were

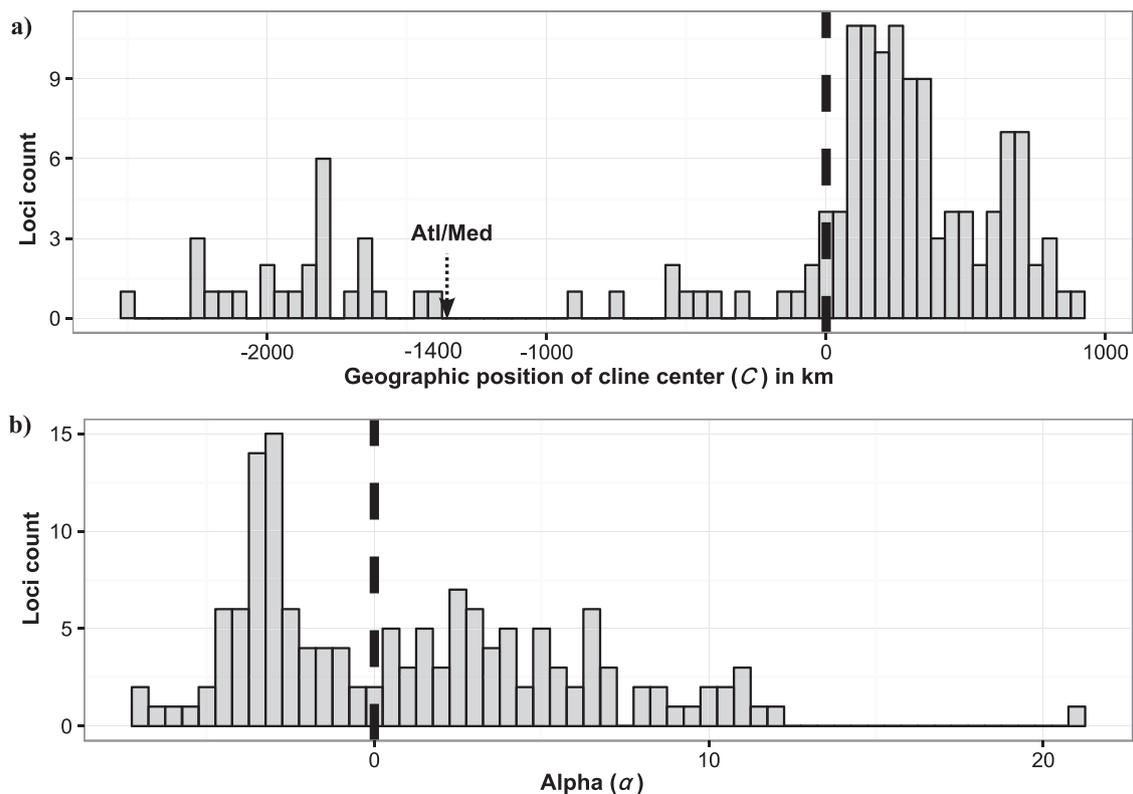


Fig. 7 Distribution of geographic and genomic cline centers for the 5% of loci showing a probability of introgression greater or equal to 95%. **a** Geographic distribution of the geographic cline center (C) with the Atlantic/Mediterranean transition zone indicated by the dotted arrow. **b**

Distribution of the genomic cline parameter α . Negative values for both cline center and α indicate an excess of *S. aegyptiaca* ancestry, whereas positive values indicate an excess of *S. senegalensis* ancestry

found with increased introgression rates in the opposite direction. Such loci generally displayed cline centers shifted away from the contact zone, probably because the Atlantic samples of *S. senegalensis* are introgressed to a lesser extent, if any, than those inside the Mediterranean as evidenced by their position along axis one of the PCA (Fig. 2b). The spatial shift in cline center was in this case only weakly positively correlated with the excess of *S. aegyptiaca* ancestry. This could be expected, however, since the genomic cline method has limited power to infer cline parameters outside the range of observed hybrid indexes in admixed genotypes (Gompert et al. 2012a), which in our case were biased in favor of *S. aegyptiaca* ancestry.

Differential introgression patterns and the dynamics of the hybrid zone

The inferred history of divergence between *S. senegalensis* and *S. aegyptiaca* places the beginning of gene flow *ca.* 18 to 30 Kyr ago, supporting a scenario of SC at the end of the last glacial period. Therefore, the hybrid zone is probably sufficiently old for the dynamics of introgression to be well established. This supports the idea that reproductive isolation between *S. senegalensis* and *S. aegyptiaca* was strong

enough to prevent genetic homogenization throughout most of the genome during this period. At the same time, the small fraction of introgressing loci that were still able to cross the species barrier indicate the existence of genomic regions unlinked to reproductive isolation loci, either due to a local lack of barrier loci and/or to a high local recombination rate (Roux et al. 2013).

Whether prezygotic or postzygotic effects could explain the observed asymmetric introgression under the hypothesis of a stable hybrid zone remains to be fully addressed. Alternatively, the geographic pattern of asymmetrical introgression evidenced here could reveal a movement of the hybrid zone (Buggs 2007), especially if the markers concerned are randomly spread across the genome (Wielstra et al. 2017). In the absence of a well-assembled reference genome, the genomic distribution of introgressing loci could not be addressed. However, a northward movement of the hybrid zone has already been proposed based on a temporal trend of increasing abundance of *S. aegyptiaca* in Bizerte lagoon (Ouanes et al. 2011). Therefore, our results probably reflect a recent or ongoing shift of the hybrid zone, with *S. aegyptiaca* incorporating the small fraction of compatible genes from *S. senegalensis* as it moves northward in its territory, leaving behind a tail of neutral or quasi-neutral

introgressed alleles. Similar signatures have already been reported in other sister species such as in house mice (Wang et al. 2011), rabbits (Carneiro et al. 2013), salamanders (Visser et al. 2017), toads (Arntzen et al. 2017), newts (Wielstra et al. 2017), and chickadees (Taylor et al. 2014). The possible causes of hybrid zone movement include the tracking of environmental changes, fitness differences among individuals from the two species, or gradients of population density (Barton and Hewitt 1985; Buggs 2007; Taylor et al. 2015; Gompert et al. 2017). Future studies will have to establish which of these hypotheses accounts for the asymmetric introgression pattern in soles.

Opposite to the preferential direction of introgression, we also found a few loci within the *S. senegalensis* genetic background whose geographic center was apparently shifted far away from the contact zone. This pattern was also captured by the first and third PCA axes, along which we observed a slight differentiation between the *S. senegalensis* samples from Dakar and Cadiz on one side and Annaba on the other, which seems to be congruent with increased introgression in Mediterranean compared to Atlantic *S. senegalensis* samples. Two non-mutually exclusive explanations can be put forward to account for this difference. Strong barriers to gene flow only have a delaying effect on the dynamics of introgression (Barton 1979). If the barrier is strong enough, many generations may be required for advantageous or simply neutral alleles to extricate themselves from the hybrid zone, depending on the balance between the selective coefficient of the focal alleles and that of the deleterious genes contained in their chromosomal vicinity. Hence, spatial allele frequency patterns may appear even well after SC, and since the entrance of the Mediterranean Sea is a well-known barrier to dispersal for many marine organisms (Patarnello et al. 2007), the wave of advance of introgression clines may be delayed as they travel toward Dakar to the south of *S. senegalensis*' range. Nevertheless, advantageous alleles are expected to race ahead neutral ones once they are freed from their background, a possibility would thus be that some of these loci exhibiting a shifted center could be related to an adaptive introgression of *S. aegyptiaca* alleles into the *S. senegalensis* background. Evidence for adaptive alleles spreading between species has been found in recent hybrid zones induced by human introductions (Fitzpatrick et al. 2010), but should be difficult to observe in historical hybrid zones where advantageous alleles have had ample time to spread (Hewitt 1988). Alternatively, these clines may correspond to alleles providing a local advantage to the *S. senegalensis* populations in the Mediterranean environment but not in the Atlantic. The hypothesis of adaptive introgression needs to be scrutinized in more details by focusing on the chromosomal signature of differentiation around the putative selected locus (Bierne 2010).

Conclusion

To conclude, our combination of analytical approaches provides new insights into the genomic architecture and the dynamics of gene flow between two divergent but still interacting parapatric species. Despite a relatively modest geographic coverage and the scarcity of available admixed genotypes in the tension zone, the genome-wide analyses taking into account the inferred history of divergence provided an efficient way to detect loci with deviant introgression behaviors. This is complementary to classical approaches based on genomic and geographic cline analyses. Our results bring new support for the tension zone model against a simple coexistence in sympatry. We show that differential gene flow has shaped genetic divergence across the tension zone, although most loci behave as if they were sitting on a congealed genome, rendering very unlikely the future remixing of the two genes pools via the creation of a hybrid swarm. Nevertheless, a few genes seem able to escape counter-selection, possibly due to different underlying processes. The first involves a possible movement of the zone, in which a shift of species range boundary leaves a tail of neutral introgressed alleles behind the front of the invading species. The second is possibly related to the existence of a physical barrier to gene flow near Gibraltar and/or the spread of globally or locally adaptive alleles into the range of the invaded species. Our results thus provide a snapshot on the genetic outcome of evolutionary processes potentially involved when divergent gene pools come back into contact after a long period of geographical isolation.

Data archiving

Demultiplexed RAD sequencing read data (fastq files) are available on NCBI under BioProject Accession: PRJNA443143. VCF files and input files for *dadi* are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.h508g38>.

Acknowledgements This work was conducted during the Ph.D. thesis of AS with the financial support of program PHC Maghreb No. XE27961 from the French Ministère des Affaires Étrangères to FB and Ministère de l'Enseignement Supérieur et de la Recherche to LB-S. The authors are indebted to M-T Augé for skillful technical assistance and thank Nicolas Bierne for insightful discussions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160:1217–1229

- Amtzen JW, de Vries W, Canestrelli D, Martínez-Solano I (2017) Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. *Mol Ecol* 26:5663–5675
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376
- Barton NH (1979) Gene flow past a cline. *Heredity* 43:333–339
- Barton NH (1983) Multilocus clines. *Evolution* 37:454–471
- Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity* 57:357–376
- Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: Harrison RG (ed.) *Hybrid Zones and the Evolutionary Process*. Oxford University Press, pp. 13–45
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annu Rev Ecol Syst* 16:113–148
- Bierne N (2010) The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution* 64:3254–3272
- Bierne N, Borsa P, Daguin C, Jollivet D, Viard F, Bonhomme F et al. (2003) Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Mol Ecol* 12:447–461
- Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol Ecol* 20:2044–2072
- Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity* 99:301
- Carneiro M, Baird SJE, Afonso S, Ramirez E, Tarroso P, Teotonio H et al. (2013) Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. *Mol Ecol* 22:2511–2525
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Inc., Sunderland, MA
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- Derryberry EP, Derryberry GE, Maley JM, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Mol Ecol Resour* 14:652–663
- Dobzhansky T (1937) Genetic nature of species differences. *Am Nat* 71:404–420
- Endler JA (1977) *Geographic variation, speciation, and clines*. Princeton University Press
- Ewing G, Hermisson J (2010) MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26:2064–2065
- Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecol Evol* 3:1951–1966
- Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB (2010) Rapid spread of invasive genes into a threatened native species. *Proc Natl Acad Sci USA* 107:3606–3610
- Gompert Z, Buerkle C (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol Ecol* 18:1207–1224
- Gompert Z, Buerkle CA (2010) introgress: a software package for mapping components of isolation in hybrids. *Mol Ecol Resour* 10:378–384
- Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. *Mol Ecol* 20:2111–2127
- Gompert Z, Buerkle CA (2012) bgc: software for Bayesian estimation of genomic clines. *Mol Ecol Resour* 12:1168–1176
- Gompert Z, Lucas LK, Nice CC, Buerkle CA (2013) Genome divergence and the genetic architecture of barriers to gene flow between *Lycaeides Idas* and *L-Melissa*. *Evolution* 67:2498–2514
- Gompert Z, Lucas LK, Nice CC, Fordyce JA, Forister ML, Buerkle CA (2012a) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution* 66:2167–2181
- Gompert Z, Mandeville EG, Buerkle CA (2017) Analysis of population genomic data from hybrid zones. *Annu Rev Ecol Evol Syst* 48:207–229
- Gompert Z, Parchman TL, Buerkle CA (2012b) Genomics of isolation in hybrids *Philos Trans R Soc Lond B Biol Sci* 367:439–450
- Grossen C, Seneviratne SS, Croll D, Irwin DE (2016) Strong reproductive isolation and narrow genomic tracts of differentiation among three woodpecker species in secondary contact. *Mol Ecol* 25:4247–4266
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5:e1000695
- Harrison RG (1990) Hybrid zones: windows on evolutionary process. *Oxf Surv Evol Biol* 7:69–128
- Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Mol Ecol* 25:2454–2466
- Harrison RG and Rand DM (1989) Mosaic hybrid zones and the nature of species boundaries. In: Otte D and Endler JA (eds) *Speciation and its Consequences*. Sinauer, Sunderland, MA, pp. 111–133
- Hewitt GM (1988) Hybrid zones-natural laboratories for evolutionary studies. *Trends Ecol Evol* 3:158–167
- Jeffery NW, DiBacco C, Van Wyngaarden M, Hamilton LC, Stanley RRE, Bernier R et al. (2017) RAD sequencing reveals genome-wide divergence between independent invasions of the European green crab (*Carcinus maenas*) in the Northwest Atlantic. *Ecol Evol* 7:2513–2524
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070–3071
- Kruuk LE, Baird SJ, Gale KS, Barton NH (1999) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* 153:1959–1971
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359
- Larson EL, Andrés JA, Bogdanowicz SM, Harrison RG (2013) Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution* 67:3653–3661
- Manchado M, Planas JV, Cousin X, Rebordinos L, Claros MG (2016) Current status in other finfish species. In: MacKenzie S. and Jentoft S. (eds) *Genomics in Aquaculture*. Academic Press
- Le Moan A, Gagnaire P-A, Bonhomme F (2016) Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Mol Ecol* 25:3187–3202
- Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution* 55:1325–1335
- Mayr, E. (1942) *Systematics and the Origin of Species*. New York, Columbia Univ. Press
- Moyle LC, Nakazato T (2010) Hybrid incompatibility “snowballs” between *Solanum* species. *Science* 329:1521–1523
- Orr HA (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139:1805–1813
- Ouanes K, Bahri-Sfar L, Ben Hassine OK, Bonhomme F (2011) Expanding hybrid zone between *Solea aegyptiaca* and *Solea senegalensis*: genetic evidence over two decades. *Mol Ecol* 20:1717–1728

- Patarnello T, Volckaert FA, Castilho R (2007) Pillars of Hercules: is the Atlantic–Mediterranean transition a phylogeographical break? *Mol Ecol* 16:4426–4444
- Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol Ecol Resour* 10:806–820
- Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Mol Ecol* 25:2337–2360
- Porter AH, Wenger R, Geiger H, Scholl A, Shapiro AM (1997) The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. *Evolution* 51:1561–1573
- Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Magnussen E, Jónsson B et al. (2014) Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. *Hereditas* 112:627
- Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197:573–589
- Rougemont Q, Gagnaire P-A, Perrier C, Genthon C, Besnard A-L, Launey S et al. (2017) Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol Ecol* 26:142–162
- Rougeux C, Bernatchez L, Gagnaire P-A (2017) Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of Lake Whitefish species pairs (*Coregonus clupeaformis*). *Genome Biol Evol* 9:2057–2074
- Roux C, Fraisse C, Romiguier J, Anciaux Y, Galtier N, Bierne N (2016) Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol* 14:e2000234
- Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Mol Biol Evol* 30:1574–1587
- She JX, Autem M, Kotulas G, Pasteur N, Bonhomme F (1987) Multivariate-analysis of genetic exchanges between *Solea aegyptiaca* and *Solea senegalensis* (Teleosts, Soleidae). *Biol J Linn Soc* 32:357–371
- Souissi A, Gagnaire PA, Bonhomme F, Bahri-Sfar L (2017) Introgressive hybridization and morphological transgression in the contact zone between two Mediterranean *Solea* species. *Ecol Evol* 7:1394–1402
- Sousa VC, Carneiro M, Ferrand N, Hey J (2013) Identifying loci under selection against gene flow in isolation-with-migration models. *Genetics* 194:211–233
- Szymura JM, Barton NH (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
- Taylor SA, Curry RL, White TA, Ferretti V, Lovette I (2014) Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution* 68:3066–3081
- Taylor SA, Larson EL, Harrison RG (2015) Hybrid zones: windows on climate change. *Trends Ecol Evol* 30:398–406
- Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, O'Brien JE et al. (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res* 18:67–76
- Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RST et al. (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun* 5:5770
- Turner JR (1967) Why does the genotype not congeal? *Evolution* 21:645–656
- Visser M, de Leeuw M, Zuiderwijk A, Arntzen JW (2017) Stabilization of a salamander moving hybrid zone. *Ecol Evol* 7:689–696
- Wang LY, Luzynski K, Pool JE, Janousek V, Dufkova P, Vyskocilova MM et al. (2011) Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Mol Ecol* 20:2985–3000
- Wielstra B, Burke T, Butlin RK, Avci A, Üzümlü N, Bozkurt E (2017) A genomic footprint of hybrid zone movement in crested newts. *Evol Lett* 1:93–101