1  **Genomic and geographic footprints of differential introgression between two highly**

2  **divergent fish species**

3

4  Ahmed Souissi[1,2,3], François Bonhomme[1,,2], Manuel Manchado[4], Lilia Bahri-Sfar[3], Pierre-

5  Alexandre Gagnaire[1,2]

6

7  [1] Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier, France

8  [2] CNRS - Institut des Sciences de l'Evolution, UMR5554 UM-CNRS-IRD-EPHE, Station

9  Méditerranéenne de l'Environnement Littoral, 34200 Sète, France

10  [3] Université de Tunis El Manar, Faculté des Sciences de Tunis UR11ES08 Biologie

11  intégrative et écologie évolutive et fonctionnelle des milieux aquatiques, 2092, Tunis, Tunisie

12  [4]IFAPA Centro El Toruño, Junta de Andalucía, Camino Tiro Pichón s/n, 11500 El Puerto de

13  Santa María, Cádiz

14

15

16

17

18 **Abstract**

19 Investigating variation in gene flow across the genome between closely related species is
20 important to understand how reproductive isolation builds up during the speciation process.
21 An efficient way to characterize differential gene flow is to study how the genetic interactions
22 that take place in hybrid zones selectively filter gene exchange between species, leading to
23 heterogeneous genome divergence. In the present study, genome-wide divergence and
24 introgression patterns were investigated between two sole species, *Solea senegalensis* and
25 *Solea aegyptiaca*, using a restriction-associated DNA sequencing (RAD-Seq) approach to
26 analyze samples taken from a transect spanning the hybrid zone. An integrative approach
27 combining geographic and genomic clines methods with an analysis of individual locus
28 introgression taking into account the demographic history of divergence inferred from the
29 joint allele frequency spectrum was conducted. Our results showed that only a minor fraction
30 of the genome can still substantially introgress between the two species due to genome-wide
31 congealing. We found multiple evidence for a preferential direction of introgression in the *S.*
32 *aegyptiaca* genetic background, indicating a possible recent or ongoing movement of the
33 hybrid zone. Deviant introgression signals found in the opposite direction supported that the
34 Mediterranean populations of *S. senegalensis* could have benefited from adaptive
35 introgression. Our study thus illustrates the varied outcomes of genetic interactions between
36 divergent gene pools that recently met after a long history of divergence.

37

38 ***Keywords***

41

42

43      **INTRODUCTION**

44

45      Reproductive isolation and hybridisation are antinomic but related processes that are tuned by

46      the same evolutionary forces. Natural hybridisation between two divergent lineages occurs as

47      long as reproductive isolation is not totally established (Dobzhansky, 1937; Mayr, 1942;

48      Coyne and Orr, 2004). The resulting exchanges of genetic material beyond the first hybrid

49      generation occur through the production of backcross and recombinant genotypes that are

50      exposed to selection, allowing neutral and advantageous alleles to spread among divergent

51      lineages but preventing gene flow at the loci involved in reproductive isolation (Barton,

52      1979). Hybrid zones, the geographic areas where hybridisation occurs, therefore act as

53      selective filters which allow us to unravel the strength of reproductive barriers and their

54      impact on the patterns and dynamics of gene exchanges at various stage of speciation (Barton

55      and Hewitt, 1985; Barton and Bengtsson, 1986; Hewitt, 1988; Martinsen *et al.*, 2001).

56      The semipermeable nature of species boundaries was first evidenced from differential

57      introgression patterns in hybrid zones (Harrison, 1990), that is, variation among loci in the

58      level of incorporation of alleles from one lineage into the other (Payseur, 2010; Harrison and

59      Larson, 2016). Introgression at a given neutral marker depends on the antagonistic effects

60      between counter-selection on a nearby selected locus and recombination with that barrier

61      locus (Barton and Bengtsson, 1986). Therefore, differential gene flow across the genome

62      reflects both recombinational distance to a barrier locus and selection acting on that locus.

63      Different methods have been developed to detect individual loci with introgression behaviours

64      departing from the genome-wide average. This includes the analysis of spatial allele

65      frequency patterns in a transect spanning the hybrid zone, or variation in ancestry proportion

66      at individual loci compared to genome-wide expectations in admixed genotypes (see Payseur

67      2010 for a review).

68      The geographic cline method is a powerful approach to analyse the relationship between

69      allele frequency and geographic distance to the center of a hybrid zone (Endler, 1977; Barton

70      and Hewitt, 1985; Barton and Gale, 1993; Porter *et al.*, 1997; Teeter *et al.*, 2008). Fitting a

71      cline model to observed allele frequency patterns allows inferring fundamental parameters

72      such as cline center, dispersal and the strength of selection. Therefore, important information

73      concerning the balance between migration and selection can be obtained using the geographic

74      cline method, such as the identification of which loci tend to introgress neutrally and which

75      do not. The genomic cline method is a related approach developed to deal with hybrid zones

76      such as mosaic hybrid zones that, unlike clinal hybrid zones, are not spatially structured,

77 (Harrison and Rand, 1989; Bierne *et al.*, 2003; Larson *et al.*, 2013). The change in allele

78 frequency at individual loci is analysed along a gradient of genomic admixture instead of a

79 spatial gradient, enabling a comparison of individual locus introgression patterns with regards

80 to the genome-wide average pattern of admixture (Gompert and Buerkle, 2009, 2011;

81 Fitzpatrick, 2013). As for geographic clines, this genomic cline method also provides a

82 quantitative assessment of the excess of ancestry to a given parental species, as well as the

83 rate of change in allele frequency from one species to the other along the admixture gradient.

84 Although both approaches are useful to document the effect of selection and recombination on

85 differential introgression in admixed genotypes, they do not specifically account for the

86 demographic history of the parental species (Payseur and Rieseberg, 2016).

87    An alternative approach that partly overcomes this limitation is the use of demo-genetic

88 models that account for varying rates of introgression among loci during the divergence

89 history. Such methods for reconstructing the history of gene flow between semi-isolated

90 lineages have been developed within Bayesian (Sousa *et al.*, 2013), approximate Bayesian

91 computation (Roux *et al.*, 2013, 2016), or approximate likelihood frameworks (Tine *et al.*,

92 2014 ; Le Moan *et al.*, 2016; Rougeux *et al.*, 2017). Using summary statistics like the joint

93 allelic frequency spectrum, which depicts correlations in allele frequencies between lineages

94 outside the hybrid zone, they capture variable behaviours among loci and allow quantifying

95 the degree of semi-permeability reflecting the overall balance between gene flow and

96 selection. These methods have the advantage of specifically accounting for the history of gene

97 flow during divergence, using contrasted speciation scenarios such as primary differentiation,

98 ancient migration or secondary contact. However, they do not provide information about

99 individual locus behaviour as the cline methods do. Here, we pushed these inference methods

100 one step further in order to assess the probability for a given locus to belong to one of two

101 categories: (*i*) loci with a reduced effective migration rate due to selection and linkage, and

102 (*ii*) loci which can readily introgress.

103    The flatfish species *Solea senegalensis* and *Solea aegyptiaca* are two economically

104 important species in the Mediterranean basin that hybridise along the northern Tunisian

105 coasts, where they form a hybrid zone (She *et al.*, 1987). Mitochondrial divergence between

106 them is high ($\sim 2\%$) if compared to other vertebrate species-pairs that also hybridise in nature,

107 such as the house mice *Mus m. musculus* and *M. m. domesticus* (0.3%, based on sequences

108 retrieved from Genebank) or Atlantic and Mediterranean lineages of *Dicentrarchus labrax*

109 (0.7%, Tine *et al.,* 2014). Based on an analysis of spatial allele frequency patterns at a dozen

110 of allozymes and intronic loci, Ouanes *et al.*, (2011) proposed that the hybrid zone between *S.*

111  *senegalensis* and *S. aegyptiaca* was centred in Bizerte lagoon acting as a non-stable unimodal

112  tension zone stemming from a secondary contact. They also suggested that the zone could

113  have undergone a recent expansion. Recently, Souissi *et al.*, (2017) showed the existence of

114  morphological transgressions within the contact zone, possibly indicating a reduced fitness of

115  recombined compared to parental genotypes. However, the low number of markers in these

116  latter studies could not provide a clear description of the genetic exchanges across the genome

117  of these two species, nor quantifying the strength of the forces maintaining the incomplete

118  reproductive isolation between them.

119  In the present work, high-throughput genotyping using RAD sequencing (Baird *et al.*,

120  2008) was carried out in individuals sampled at both sides of the hybrid zone, with a

121  particular effort in the contact zone itself. By combining three different methods (genomic

122  and geographic cline analysis and historical demographic inference) exploiting different

123  aspects of the data, we provide new insights on the history of divergence between *S.*

124  *senegalensis* and *S. aegyptiaca,* and a genome-wide description of varied patterns of

125  introgression attributed to recent or ongoing movement of the hybrid zone.

126

## MATERIALS AND METHODS

### Sampling

A total of 153 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 Bardawill lagoon in Egypt). Sampling size was locally increased in the Tunisian region where both species coexist and hybridisation has been reported (Bizerte lagoon $n = 29$ and Gulf of Tunis $n = 25$). Finally, a total of 7 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. Restriction-Associated DNA (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 to 6 bp. Ligated products were then combined in equimolar proportions into five 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, end-clipped to 95 pb and quality filtered using the program *process_radtags* from the *stacks* pipeline with default settings (Catchen *et al.*, 2013). Retained read 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 2 v.2.1.0 (Langmead and Salzberg, 2012) with the --very-sensitive option. In order to take into account both the level of divergence among species and the possibility of introgression, we used a subset of individuals to empirically determine the optimal maximum number of mismatches allowed between aligned reads and the reference genome (m = 7). 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 5X per stack. Homologous loci across samples were merged based on their genomic position within scaffolds using *cstacks* to construct a catalogue of loci. Individual stacks were then matched against the catalogue of loci with *sstacks* to determine genotypes. The module *rxstacks* was ran with the --prune_haplo option to exclude poor quality loci using a log-likelihood threshold of -300 (determined empirically). Individual genotypes were finally exported in 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 *S. senegalensis* and *S. aegyptiaca* 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. Over the 116,385 remaining SNPs, we randomly selected one single SNP for each pair of RAD loci associated with the same restriction site, 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 SNPs.

**Genetic structure and hybridisation**

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). We then estimated individual admixture proportions from two differentiated genetic clusters using *fastStructure* (Raj *et al.*, 2014) with $10^8$ iterations, in order to distinguish parental and admixed genotypes.

194     **Demo-genetic inference of the divergence history**

195     We used a modified version of the $\delta a \delta i$ program (Gutenkunst *et al.*, 2009) to infer the

196     divergence history between *S. senegalensis* and *S. aegyptiaca*. This method uses the joint

197     allele frequency spectrum (JAFS) between two populations as summary statistics to

198     characterize divergence. $\delta a \delta i$ fits demographic divergence models to the observed data using

199     a diffusion approximation of the JAFS, enabling to compare different alternative divergence

200     models in a composite likelihood framework. We used seven divergence models developed in

201     a previous study (Tine *et al.*, 2014) to determine whether and how gene flow has shaped

202     genome divergence between *S. senegalensis* and *S. aegyptiaca*. The simplest model of strict

203     isolation (*SI*) corresponds to an allopatric divergence scenario in which the ancestral

204     population of effective size $N_A$ splits into two derived populations of size $N_1$ (*S. senegalensis*)

205     and $N_2$ (*S. aegyptiaca*) that evolve without exchanging genes for $T_s$ generations. We then

206     considered three models of divergence including gene flow, either during the entire

207     divergence period (isolation with migration, *IM*), the beginning of divergence (ancient

208     migration, *AM*), or the most recent part of divergence (secondary contact, SC). In these

209     models, migration occurs with potentially asymmetric rates ($m_{12}$ and $m_{21}$) that are shared

210     across all loci in the genome. We also considered simple extensions of the *IM*, *AM* and *SC*

211     models that capture the effect of selection by accounting for heterogeneous migration rates

212     among loci (*IM2m*, *AM2m* and *SC2m*). These semi-permeability models consider that two

213     categories of loci, experiencing different effective migration rates, occur in proportions $P$ and

214     1-$P$ in the genome.

215     The JAFS was obtained by pooling the least introgressed populations from *S. senegalensis*

216     (Dakar, Cadix, Annaba, and Mellah) for species 1 and *S. aegyptiaca* (Kerkennah, L. Bibane,

217     and L. Bardawill) for species 2 (Fig. 2). We used *S. solea* samples as an outgroup species to

218     determine the most parsimonious ancestral allelic state for each SNP in order to generate an

219     unfolded JAFS (Fig. 3a). The size of the JAFS was projected down to 40 sampled

220     chromosomes per species to account for missing data. For each model, we estimated the

221     parameter values that maximize likelihood using two successive simulated annealing (SA)

222     procedures before quasi-Newton (BFGS) optimization (Tine *et al.*, 2014). Comparisons

223     among models was made using the Akaike information criterion (AIC) to account for

224     variation in the number of parameters among models. A total of 20 independent runs were

225     used for the optimization of each model. Parameters uncertainties were estimated from non-

226     parametric bootstrapped data using the Godambe information matrix as implemented in $\delta a \delta i$.

227    We used 1000 bootstrapped datasets to estimate confidence intervals as the maximum

228    likelihood parameter value ± 1.96×SE.

229 **Inferring the probability of locus introgression under the best divergence model**

230 Because the JAFS-based inference of the divergence history does not provide an assessment

231 of introgression probability for each locus separately, we developed an approach to estimate

232 the relative probability of each individual locus to be assigned to one of these two categories:

233 (*i*) loci which can readily introgress between species, and (*ii*) loci experiencing a highly

234 reduced introgression rate due to selection against foreign alleles and linkage. This probability

235 was estimated using the best-fit model identified in the previous section, which was a

236 secondary contact model with variable introgression rates among loci (SC*2m*) (Fig. 3b). The

237 SC*2m* model can be decomposed as a linear combination of two simple models describing

238 gene flow in two different compartments of the genome. The best-fit SC*2m* model estimated

239 that only 5% of the loci can still introgress between species with effective migration rates $m_{1-2}$

240 and $m_{2-1}$, whereas the remaining 95% of loci experience a highly reduced introgression rate

241 with effective migration rates $m'_{1-2}$ and $m'_{2-1}$. We used estimated model parameters to

242 perform coalescent simulations with MSMS (Ewing and Hermisson, 2010) under the

243 secondary contact model (assuming theta=747.202; $N_1$=0,818; $N_2$=1,136; $T_S$=4,803;

244 $T_{SC}$=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*

246 *introgression* rate dataset. One thousand JAFS with 40 sampled chromosomes per species

247 were produced under each model. We then averaged the number of derived alleles within

248 entries across replicates to obtain a single JAFS for both the *free introgression* ($\Upsilon$) (Fig. 3c)

249 and the *reduced introgression* ($\Phi$) (Fig. 3d) model. Finally, we used the average number of

250 SNPs per entry (*i,j*) in each JAFS to estimate the probability $P_{i,j}$ that a SNP with a derived

251 allele count *i* in species 1 and *j* in species 2 can be obtained under the *free introgression*

252 model:

253

$$P_{i,j} = 0.05 \times \Upsilon_{i,j} / \left(0.05 \times \Upsilon_{i,j} + 0.95 \times \Phi_{i,j}\right) \tag{1}$$

254

255 Where $\Upsilon_{i,j}$ and $\Phi_{i,j}$ are the average number of SNP predicted in the JAFS entry (*i,j*) under the

256 *free introgression* and *reduced introgression* model, respectively. Every SNP from the real

257 dataset was finally associated to the introgression probability $P_{i,j}$ given by equation (1) based

258 on its corresponding entry in the JAFS.

259

**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 $\alpha$ and $\beta$, which describe locus-specific introgression based on ancestry. Parameter $\alpha$ 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 $\alpha$ reflects an increase in the probability of ancestry from species one (introgression into *S. aegyptiaca*), whereas a negative value of $\alpha$ reflects an increase in the probability of ancestry from species two (introgression into *S. senegalensis*). Parameter $\beta$ describes the rate of transition in the probability of ancestry from parental population one to parental population two as a function of hybrid index, that is, the amount of introgression. Positives $\beta$ values thus denote a restricted amount of introgression whereas negative $\beta$ values indicate a greater introgression rate compared to the genome-wide average.

We designated outlier loci with respect to parameters $\alpha$ and $\beta$ based on the assumption that the genome-wide distributions of locus-specific cline parameters are both centred on zero. Therefore, a locus was considered as an $\alpha$ outlier if its posterior estimate of $\alpha$ was not contained in the interval bounded by the 0.025 and 0.975 quantiles of $\mathcal{N}(0, \tau_{\alpha})$ (likewise for $\beta$) (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. We used a model that fits cline center, width, and independent introgression tails using estimated values for minimum ($p_{min}$) and maximum ($p_{max}$) allele frequencies.

**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±1,878,685). Individual genotype

294     calling in Stacks produced a raw VCF file containing 174,490 SNP, from which we retained

295     10,758 unlinked and oriented SNP that were used for downstream analyses.

296

**Genetic structure and hybridisation**

298     The Principal Component Analysis (PCA) clearly separated *S. senegalensis* from *S.*

299     *aegyptiaca* samples along the first PC axis (PC1), which explained 74.2% of the total

300     genotypic variance (Fig. 2a). *S. senegalensis* samples were organised along that axis

301     according to their geographical proximity from the contact zone, following a gradient of

302     genetic similarity to *S. aegyptiaca* increasing from Dakar to Bizerte Lagoon. The two species

303     were found to coexist only in Bizerte Lagoon, with four *S. aegyptiaca* individuals being found

304     among a majority of *S. senegalensis* genotypes. The second principal component (PC2)

305     revealed a weak signal of differentiation (0.73% of associated variance) separating Bardawil

306     lagoon from other samples within *S. aegyptiaca*. The third principal component (PC3, 0.66%

307     of genotypic variance) mostly differentiated Atlantic samples from Dakar and Cadix from the

308     Mediterranean *S. senegalensis* samples (Fig. 2b). The FastStructure analysis (Fig. 2c)

309     confirmed the separation of the two species into different genetic clusters and their

310     coexistence in Bizerte lagoon, as detected from the PCA. The finding of intermediate

311     admixture proportions in some individuals revealed signs of introgressive hybridization

312     between *S. senegalensis* and *S. aegyptiaca* around the contact zone.

313

**Demo-genetic history of divergence**

315     The $\delta a \delta i$ analysis showed that the secondary contact model with varying introgression rates

316     along the genome best explained the observed JAFS (SC2*m*, delta AIC > 25). In comparison,

317     the six other models had a significantly lower performance for different reasons. While the

318     strict isolation model (SI) could explain SNPs occupying the outer frame of the JAFS, it could

319     not predict at the same time the presence of loci in the more inner part of the spectrum. The

320     IM, AM and SC models assuming genome-wide homogeneous migration rates better

321     predicted loci occupying the central part of the JASF. However, they underestimated the

322     density of private and highly differentiated SNPs. Finally, the three models including

323     heterogeneous migration rates along the genome (IM2*m*, AM2*m* and SC2*m*) yielded

324     significantly improved fits. However, only the SC2*m* model provided a good prediction for

325     both a high density of highly differentiated SNPs between the two species and a lack of SNPs

326     towards the central part of the spectrum.

327     Using the best-fit obtained for the SC2$m$ model over 20 independent runs to get estimates

328     of model parameters and confidence intervals, we found that the duration of the isolation

329     period between *S. senegalensis* and *S. aegyptiaca* was about 60 times longer than the duration

330     of the secondary contact. A large proportion of the genome (~95%) was associated with

331     relatively small effective migration rates and limited gene flow corresponding to less than one

332     effective migrant per generation in both directions ($N_1m'_{12} = 0.047$, $N_2m'_{21} = 0.175$). By

333     contrast, the effective migration rate was more elevated in the remaining small fraction of the

334     genome (~5%), especially in the direction from *S. senegalensis* to *S. aegyptiaca* where

335     introgression was found 80 times higher than elsewhere in the genome. Introgression was

336     found 12 times lower in direction of *S. senegalensis*, although still more than twice higher

337     than in the remainder (95%) of the genome (Table 1).

338

### Genomic Clines

340     The genomic cline analysis performed with the BGC program revealed different behaviours

341     among the 10,756 SNPs with respect to the cline parameters $\alpha$ and $\beta$. Considering the locus-

342     wise ancestry shift parameter $\alpha$, we observed that 48% of the loci have an excess of *S.*

343     *senegalensis* ancestry (negative $\alpha$) whereas only 29% of them have an excess of *S. aegyptiaca*

344     ancestry (positive $\alpha$). The quantile method for outlier detection confirmed that a higher

345     proportion of loci was characterized by an excess of *S. senegalensis* ancestry (43% of

346     negative $\alpha$ outliers) compared to *S. aegyptiaca* ancestry (17% of positive $\alpha$ outliers). By

347     contrast, the locus-wise slope parameter $\beta$ depicting the rate of transition between species was

348     symmetrically distributed between negative and positive values, with equal proportions of loci

349     showing a decreased (48% of positive $\beta$) and increased (51% of negative $\beta$) introgression

350     rates. The proportion of loci showing significant deviation compared to the genomic average

351     amounted to 52%, and were equally distributed among negative (26%) and positive (26%) $\beta$

352     outliers.

353

### Geographic clines

355     The mean position of cline center ($C$) calculated over all fitted clines was located 12 km to the

356     east of Bizerte lagoon (Fig. 4). Most individual locus clines tended to co-localize at this

357     position, especially for the steepest clines whose centres localised essentially 10 Km or less

358     around the center of the hybrid zone. In general, loci whose individual centres did not

359     coincide with the central part of the hybrid zone harboured less steep clines (Fig. 4)

360

**Links between different approaches**

Correlation tests were used to investigate whether different approaches 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 joint distribution of the genomic cline parameter $\alpha$ and geographic cline center $c$ (Fig. 5), showed a significant positive correlation on both sides of the hybrid zone. 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 ($R^2_{c-\alpha} = 0.023$, $P < 10^{-10}$).

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 $\beta$ outliers detected with BGC, and compared $\beta$ 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, 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 centres are located outside of the contact zone, with a spatial shift more pronounced in the *S. aegyptiaca* direction (Fig. 7a). Similarly, the distribution of the genomic cline parameter $\alpha$ 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

395    hybrid zone near Bizerte lagoon (She *et al.*, 1987 ; Ouanes *et al.*, 2011; Souissi *et al.*, 2017).

396    However, the proportion of hybrid genotypes detected here in the contact zone (i.e. 3 among

397    64 individuals) is much lower than established in previous studies. Although this could be due

398    to large temporal variations in the rate of hybridisation, the most plausible hypothesis is that

399    real hybrids (e.g. F1, F2 and first generations backcrosses) are more easily distinguished from

400    introgressed genotypes using a high number of diagnostic markers (e.g. Pujolar *et al.*, 2014;

401    Jeffery *et al.*, 2017). Our results thus support the existence of an abrupt geographic transition

402    between the two species across their contact zone, where predominantly parental genotypes

403    occur in sympatry with an overall low frequency of hybrids. These observations are consistent

404    with the tension zone model (Barton and Hewitt, 1985), in which the hybrid zone is

405    maintained by a balance between the influx of parental genotypes from outside the zone and

406    the counter-selection of hybrid genotypes inside the zone. This interpretation is also in

407    keeping with the finding of transgressive body shape variation attributed to a reduced

408    condition index of admixed genotypes from the contact zone (Souissi *et al.*, 2017).

409        The regions located outside the contact zone, which contain only parental genotypes,

410    provided relevant information to infer the demo-genetic history of divergence between these

411    two species. Gene exchange between *S. senegalensis* and *S. aegyptiaca* is best explained by a

412    secondary contact model with heterogeneous rates of introgression along the genome. Similar

413    divergence histories have already been found between geographical lineages or ecotypes of

414    the same species, like in the European sea bass and anchovy (Tine *et al.*, 2014; Le Moan *et

415    al.*, 2016). Using the same approach as implemented here, these studies could determine that a

416    minor but significant fraction of the genome (i.e. 20 to 35%) does not neutrally introgress

417    between hybridizing lineages/ecotypes, thus providing a quantitative assessment of the degree

418    of semi-permeability to gene flow. In the present case, by contrast, we found that the great

419    majority of the genome (approximately 95%) experiences a highly reduced effective

420    migration rate (i.e. < 1 migrant per generation) between species. Moreover, the inferred

421    divergence period estimated using a mutation rate of $10^{-8}$ per bp per year and a generation

422    time of 3 to 5 years corresponds to a separation time of *ca*. 1.1 to 1.8 Myrs, which together

423    with the mitochondrial sequence divergence of 2%, indicates a relatively ancient speciation

424    event. This timing of divergence is much older than the approximately 300 Kyrs that were

425    inferred between glacial lineages in anchovy and sea bass (Tine *et al.*, 2014; Le Moan *et al.*,

426    2016). Therefore, our results are consistent with the prediction that more anciently diverged

427    species should display stronger reproductive isolation due to the establishment of more

428　　　genetic barriers (Orr, 1995; Moyle and Nakazato, 2010; Roux *et al.*, 2016), which is reflected

429　　　here by a lower level of permeability to gene flow compared to sea bass and anchovy.

430　　　　　Another aspect of this low miscibility of the two genomes was captured by individual

431　　　locus cline parameters. Consistent with the view that stepped clines generate a strong barrier

432　　　to gene flow (Szymura and Barton, 1986), we found a high proportion of loci combining

433　　　extremely low introgression rates (i.e. positive $\beta$ outliers) and steep geographic clines (i.e.

434　　　cline width below the average) with centres colocalized in the central part of the hybrid zone.

435　　　Because this type of situation generates strong linkage disequilibrium, each locus is expected

436　　　to receive, in addition to its own selective coefficient, the effect of selection on other loci

437　　　(Barton, 1983; Kruuk *et al.*, 1999). This makes the genome acting almost as a single

438　　　underdominant locus causing a global reduction in gene flow, which sometimes is referred to

439　　　as a "congealed genome" (Turner, 1967; Bierne *et al.*, 2011; Gompert *et al.*, 2012b). In

440　　　contrast to this genome-wide reduction in gene flow, we also found support for a small

441　　　proportion of genomic regions with higher-than-average introgression rates between species.

442

### Multiple approaches to detect increased introgression rates

444　　Differential introgression among loci was evidenced by combining several approaches that

445　　capture different but complementary aspects of the data. For instance, genomic and

446　　geographic cline methods both rely on the sampling of admixed genotypes from the hybrid

447　　zone, whereas the modelling approach based on the JAFS relies of the joint distribution of

448　　allele frequencies between parental populations located outside of the hybrid zone. By

449　　combining these methods, we found that the 5% of loci with higher-than-average

450　　introgression rates identified with the JASF approach generally display shifts in both

451　　geographic and genomic cline centres. Therefore, our method to detect increased introgression

452　　while accounting for the demographic divergence history identifies genomic regions that also

453　　depart from the genome average in those places where the two species meet and admix.

454　　　　Discordant clines, were not symmetrically distributed between the *S. senegalensis* and *S.*

455　　*aegyptiaca* side of the hybrid zone. The majority of the highly introgressed loci were shifted

456　　into the *S. aegyptiaca* geographic range. For these loci, the excess of *S. senegalensis* ancestry

457　　(measured by the genomic cline parameter $\alpha$) was positively related to the extent of the spatial

458　　shift of the cline center into the *S. aegyptiaca* territory. This correlation was much stronger

459　　than previously observed in a bird hybrid zone study (Grossen *et al.*, 2016), possibly due to a

460　　stronger variance in cline shift in the *Solea* system. By contrast, much fewer loci were found

461　　with increased introgression rates in the opposite direction. Such loci generally displayed

462  cline centres shifted far away from the contact zone, near the entrance of the Mediterranean

463  Sea. The spatial shift in cline center was in this case only weakly positively correlated with

464  the excess of *S. aegyptiaca* ancestry. This could be expected, however, since the genomic

465  cline method has limited power to infer cline parameters outside the range of observed hybrid

466  indexes in admixed genotypes (Gompert *et al.*, 2012a), which in our case were biased in

467  favour of *S. aegyptiaca* ancestry.

468

469  **Differential introgression patterns and the dynamics of the hybrid zone**

470

471  The inferred history of divergence between *S. senegalensis* and *S. aegyptiaca* places the

472  beginning of gene flow *ca.* 18 to 30 Kyrs ago, supporting a scenario of secondary contact at

473  the end of the last glacial period. Therefore, the hybrid zone is probably sufficiently old for

474  the dynamics of introgression to be well established. This supports the idea that reproductive

475  isolation between *S. senegalensis* and *S. aegyptiaca* was strong enough to prevent genetic

476  homogenization throughout most of the genome during this period. At the same time, the

477  small fraction of introgressing loci that were still able to cross the species barrier indicate the

478  existence of genomic regions unlinked to reproductive isolation loci, either due to a local lack

479  of barrier loci and/or to a high local recombination rate (Roux *et al.*, 2013).

480      Whether pre- or postzygotic effects could explain the observed asymmetric introgression

481  under the hypothesis of a stable hybrid zone remains to be fully addressed. Alternatively, the

482  geographic pattern of asymmetrical introgression evidenced here could reveal a movement of

483  the hybrid zone (Buggs, 2007), especially if the markers concerned are randomly spread

484  across the genome (Wielstra *et al.*, 2017). In the absence of a well-assembled reference

485  genome, the genomic distribution of introgressing loci could not be addressed. However, a

486  northward movement of the hybrid zone has already been proposed based on a temporal trend

487  of increasing abundance of *S. aegyptiaca* in Bizerte lagoon (Ouanes *et al.*, 2011). Therefore,

488  our results probably reflect a recent or ongoing shift of the hybrid zone, with *S. aegyptiaca*

489  incorporating the small fraction of compatible genes from *S. senegalensis* as it moves

490  northward in its territory, leaving behind a tail of neutral introgressed alleles. Similar

491  signatures have already been reported in other sister species such as in house mice (Wang *et*

492  *al.*, 2011), rabbits (Carneiro *et al.*, 2013), salamanders (Visser *et al.*, 2017), toads (Arntzen *et*

493  *al.*, 2017), newts (Wielstra *et al.*, 2017) and chikadees (Taylor *et al.*, 2014). The possible

494  causes of hybrid zone movement include the tracking of environmental changes, fitness

495  differences among individuals from the two species, or gradients of population density

496    (Barton and Hewitt, 1985; Buggs, 2007; Taylor *et al.*, 2015; Gompert *et al.*, 2017). Future

497    studies will have to establish which of these hypotheses accounts for the asymmetric

498    introgression pattern in soles.

499        Opposite to the preferential direction of introgression, we also found a few loci within the

500    *S. senegalensis* genetic background whose geographic center was apparently shifted near the

501    entrance of the Mediterranean, far away from the contact zone. This pattern was also captured

502    by the first and third PCA axes, along which we observed a slight differentiation between the

503    *S. senegalensis* samples from Dakar, Cadiz and Annaba, which seems to be congruent with

504    increased introgression in Mediterranean compared to Atlantic *S. senegalensis* samples. Long

505    range gene flow is unlikely to explain these discordant clines that only concern a few loci. A

506    more probable hypothesis would be that such loci exhibiting a shifted center could be related

507    to an adaptive introgression of *S. aegyptiaca* alleles into the *S. senegalensis* background.

508    Evidence for adaptive alleles spreading between species has been found in recent hybrid

509    zones induced by human introductions (Fitzpatrick *et al.*, 2010), but should be difficult to

510    observe in historical hybrid zones where advantageous alleles have had ample time to spread

511    (Hewitt, 1988). Strong barriers to gene flow only have a delaying effect on the dynamics of

512    introgression of favourable alleles (Barton, 1979). If the barrier is strong enough, many

513    generations may be required for advantageous alleles to extricate themselves from the hybrid

514    zone, and therefore spatial allele frequency patterns related to adaptive introgression may

515    appear even well after secondary contact. Moreover, the entrance of the Mediterranean Sea is

516    a well-known barrier to dispersal for many marine organisms (Patarnello *et al.*, 2007), which

517    may delay the wave of advance of adaptive introgression clines as they travel towards Dakar

518    to the south of *S. senegalensis'* range. Alternatively, these clines may correspond to alleles of

519    *S. aegyptiaca* ancestry, providing a local advantage to the *S. senegalensis* populations in the

520    Mediterranean environment but not in the Atlantic. The hypothesis of adaptive introgression

521    needs to be scrutinized in more details by focusing on the chromosomal signature of

522    differentiation around the putative selected locus (Bierne, 2010).

523

524    **Conclusion**

525    To conclude, our combination of analytical approaches provides new insights into the

526    genomic architecture and the dynamics of gene flow between two highly divergent but still

527    interacting parapatric species. Despite a relatively modest geographic coverage and the

528    scarcity of available admixed genotypes in the tension zone, the genome-wide analyses taking

529    into account the inferred history of divergence provided an efficient way to detect loci with

530    deviant introgression behaviours. This is complementary to classical approaches based on

531    genomic and geographic cline analyses. Our results bring new support for the tension zone

532    model against a simple coexistence in sympatry. We show that differential gene flow has

533    shaped genetic divergence across the tension zone, although most loci behave as if they were

534    sitting on a congealed genome, rendering very unlikely the future remixing of the two genes

535    pools via the creation of a hybrid swarm. Nevertheless, a few genes seem able to escape

536    counter-selection, possibly due to different underlying processes. The first involves a possible

537    movement of the zone, in which a shift of species range boundaries leaves a tail of neutral

538    introgressed alleles behind the front of the invading species. The second is possibly related to

539    the spread of globally or locally adaptive alleles into the range of the invaded species. Our

540    results thus provide a snapshot on the genetic outcome of evolutionary processes potentially

541    involved when divergent gene pools come back into contact after a long period of

542    geographical isolation.

543

## References:

Arntzen 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*: n/a-n/a.

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. *Hybrid Zones Evol Process*: 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**: 3124–40.

Coyne JA, Orr HA (2004). *Speciation. Sunderland, MA*. Sinauer Associates, Inc.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156–8.

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.

578    Fitzpatrick BM (2013). Alternative forms for genomic clines. *Ecol Evol* **3**: 1951–66.

579    Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB (2010). Rapid spread of invasive
580            genes into a threatened native species. *Proc Natl Acad Sci* **107**: 3606–3610.

581    Gompert Z, Buerkle C (2009). A powerful regression-based method for admixture mapping of
582            isolation across the genome of hybrids. *Mol Ecol* **18**: 1207–1224.

583    Gompert Z, Buerkle CA (2011). Bayesian estimation of genomic clines. *Mol Ecol* **20**: 2111–27.

584    Gompert Z, Buerkle CA (2012). bgc: Software for Bayesian estimation of genomic clines. *Mol Ecol*
585            *Resour* **12**: 1168–1176.

586    Gompert Z, Lucas LK, Nice CC, Buerkle CA (2013). Genome Divergence and the Genetic Architecture
587            of Barriers to Gene Flow between Lycaeides Idas and L-Melissa. *Evolution* **67**: 2498–2514.

588    Gompert Z, Lucas LK, Nice CC, Fordyce JA, Forister ML, Buerkle CA (2012). Genomic regions with a
589            history of divergent selection affect fitness of hybrids between two butterfly species.
590            *Evolution* **66**: 2167–2181.

591    Gompert Z, Mandeville EG, Buerkle CA (2017). Analysis of Population Genomic Data from Hybrid
592            Zones. *Annu Rev Ecol Evol Syst* **48**: null.

593    Gompert Z, Parchman TL, Buerkle CA (2012). Genomics of isolation in hybrids. *Philos Trans R Soc Lond*
594            *B Biol Sci* **367**: 439–450.

595    Grossen C, Seneviratne SS, Croll D, Irwin DE (2016). Strong reproductive isolation and narrow
596            genomic tracts of differentiation among three woodpecker species in secondary contact. *Mol*
597            *Ecol* **25**: 4247–4266.

598    Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009). Inferring the joint
599            demographic history of multiple populations from multidimensional SNP frequency data.
600            *PLoS Genet* **5**: e1000695.

601    Harrison RG (1990). Hybrid zones: windows on evolutionary process. *Oxf Surv Evol Biol* **7**: 69–128.

602    Harrison RG, Larson EL (2016). Heterogeneous genome divergence, differential introgression, and the
603            origin and structure of hybrid zones. *Mol Ecol* **25**: 2454–66.

604    Harrison RG, Rand DM (1989). Mosaic hybrid zone and the nature of species boundaries.

605    Hewitt GM (1988). Hybrid zones-natural laboratories for evolutionary studies. *Trends Ecol Evol* **3**:
606            158–167.

607    Jeffery NW, DiBacco C, Van Wyngaarden M, Hamilton LC, Stanley RRE, Bernier R, *et al.* (2017). RAD
608            sequencing reveals genomewide divergence between independent invasions of the European
609            green crab (Carcinus maenas) in the Northwest Atlantic. *Ecol Evol* **7**: 2513–2524.

610    Jombart T (2008). adegenet: a R package for the multivariate analysis of genetic markers.
611            *Bioinformatics* **24**: 1403–5.

612    Jombart T, Ahmed I (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.
613            *Bioinformatics* **27**: 3070–1.

614    Kruuk LE, Baird SJ, Gale KS, Barton NH (1999). A comparison of multilocus clines maintained by
615         environmental adaptation or by selection against hybrids. *Genetics* **153**: 1959–71.

616    Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–9.

617    Larson EL, Andrés JA, Bogdanowicz SM, Harrison RG (2013). Differential introgression in a mosaic
618         hybrid zone reveals candidate barrier genes. *Evolution* **67**: 3653–3661.

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

622    Martinsen GD, Whitham TG, Turek RJ, Keim P (2001). Hybrid populations selectively filter gene
623         introgression between species. *Evolution* **55**: 1325–1335.

624    Mayr E (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard
625         University Press.

626    Moyle LC, Nakazato T (2010). Hybrid incompatibility "snowballs" between Solanum species. *Science*
627         **329**: 1521–1523.

628    Orr HA (1995). The population genetics of speciation: the evolution of hybrid incompatibilities.
629         *Genetics* **139**: 1805–1813.

630    Ouanes K, Bahri-Sfar L, Ben Hassine OK, Bonhomme F (2011). Expanding hybrid zone between Solea
631         aegyptiaca and Solea senegalensis: genetic evidence over two decades. *Mol Ecol* **20**: 1717–
632         1728.

633    Patarnello T, Volckaert FA, Castilho R (2007). Pillars of Hercules: is the Atlantic–Mediterranean
634         transition a phylogeographical break? *Mol Ecol* **16**: 4426–4444.

635    Payseur BA (2010). Using differential introgression in hybrid zones to identify genomic regions
636         involved in speciation. *Mol Ecol Resour* **10**: 806–820.

637    Payseur BA, Rieseberg LH (2016). A genomic perspective on hybridization and speciation. *Mol Ecol*
638         **25**: 2337–60.

639    Porter AH, Wenger R, Geiger H, Scholl A, Shapiro AM (1997). The Pontia daplidice-edusa hybrid zone
640         in northwestern Italy. *Evolution* **51**: 1561–1573.

641    Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Magnussen E, Jónsson B, *et al.* (2014). Assessing
642         patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide
643         polymorphisms. *Heredity* **112**: 627.

644    Raj A, Stephens M, Pritchard JK (2014). fastSTRUCTURE: variational inference of population structure
645         in large SNP data sets. *Genetics* **197**: 573–89.

646    Rougeux C, Bernatchez L, Gagnaire P-A (2017). Modeling the Multiple Facets of Speciation-with-
647         Gene-Flow toward Inferring the Divergence History of Lake Whitefish Species Pairs
648         (Coregonus clupeaformis). *Genome Biol Evol* **9**: 2057–2074.

649    Roux C, Fraisse C, Romiguier J, Anciaux Y, Galtier N, Bierne N (2016). Shedding light on the grey zone
650         of speciation along a continuum of genomic divergence. *PLoS Biol* **14**: e2000234.

651  Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013). Crossing the species barrier: genomic hotspots of
652      introgression between two highly divergent Ciona intestinalis species. *Mol Biol Evol* **30**:
653      1574–1587.

654  She JX, Autem M, Kotulas G, Pasteur N, Bonhomme F (1987). Multivariate-Analysis of Genetic
655      Exchanges between Solea-Aegyptiaca and Solea-Senegalensis (Teleosts, Soleidae). *Biol J Linn*
656      *Soc* **32**: 357–371.

657  Souissi A, Gagnaire PA, Bonhomme F, Bahri-Sfar L (2017). Introgressive hybridization and
658      morphological transgression in the contact zone between two Mediterranean Solea species.
659      *Ecol Evol* **7**: 1394–1402.

660  Sousa VC, Carneiro M, Ferrand N, Hey J (2013). Identifying loci under selection against gene flow in
661      isolation-with-migration models. *Genetics* **194**: 211–233.

662  Szymura JM, Barton NH (1986). Genetic analysis of a hybrid zone between the fire-bellied toads,
663      Bombina bombina and B. variegata, near Cracow in southern Poland. *Evolution*: 1141–1159.

664  Taylor SA, Curry RL, White TA, Ferretti V, Lovette I (2014). Spatiotemporally consistent genomic
665      signatures of reproductive isolation in a moving hybrid zone. *Evolution* **68**: 3066–81.

666  Taylor SA, Larson EL, Harrison RG (2015). Hybrid zones: windows on climate change. *Trends Ecol Evol*
667      **30**: 398–406.

668  Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, O'Brien JE, *et al.* (2008). Genome-
669      wide patterns of gene flow across a house mouse hybrid zone. *Genome Res* **18**: 67–76.

670  Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RST, *et al.* (2014). \European sea bass
671      genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat*
672      *Commun* **5**.

673  Turner JR (1967). Why does the genotype not congeal? *Evolution*: 645–656.

674  Visser M, de Leeuw M, Zuiderwijk A, Arntzen JW (2017). Stabilization of a salamander moving hybrid
675      zone. *Ecol Evol* **7**: 689–696.

676  Wang LY, Luzynski K, Pool JE, Janousek V, Dufkova P, Vyskocilova MM, *et al.* (2011). Measures of
677      linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house
678      mouse hybrid zone. *Mol Ecol* **20**: 2985–3000.

679  Wielstra B, Burke T, Butlin RK, Avcı A, Üzüm N, Bozkurt E, *et al.* (2017). A genomic footprint of hybrid
680      zone movement in crested newts. *Evol Lett*.

681

682

683

**Conflict of Interest**

All the authors declare no conflict of interest concerning the data presented here

**Data Archiving**

Raw sequences data and VCF files will be deposed upon acceptance on Genbank and Dryad

697 **Table 1:** Results of model fitting for seven alternative models of divergence between *S. senegalensis* (SEN) and *S. aegyptiaca* (AEG). In the order of appearance in the table: the model fitted, the Akaike information criterion (AIC), the maximum-likelihood estimate over 20 independent runs (after 3 successive optimization steps: simulated annealing hot, cold and BFGS) and theta parameter for the ancestral population before split. Following are the inferred values for the model parameters (scaled by theta): the effective size of *S. aegyptiaca* ($N_1$) and *S. senegalensis* ($N_2$) populations, 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 (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 proportion of correctly orientated SNPs ($O$) when inferring ancestral allelic states using *S. solea* as an outgroup. *SI*: Strict isolation, *IM*: Isolation with migration, *AM*: Ancient migration, *IM2m*: Isolation with two rates of migration, *AM2m*: Ancient migration with two rates of migration, SC: Secondary contact, SC2m: Secondary contact with two rates of migration.

| Model | AIC | Log-likelihood | *theta* | *nu* A | $N_1$ | $N_2$ | $m_{12}$ | $m_{21}$ | $m'_{12}$ | $m'_{21}$ | $T_S$ | $T_{AM}$ | $T_{SC}$ | $P$ | $O$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *SI* | 4611.53 | -2301.76 | 2569.18 | 1 | 0.226 | 0.287 | - | - | - | - | 0.571 | - | - | - | 0.988 |
| *IM* | 3049.93 | -1518.96 | 1018.56 | 1 | 0.572 | 0.768 | 0.035 | 0.053 | - | - | 3.809 | - | - | - | 0.981 |
| *AM* | 3048.24 | -1517.12 | 514.41 | 1 | 1.14 | 1.53 | 0.017 | 0.026 | - | - | 8.55 | 0.0002 | - | - | 0.977 |
| *IM2m* | 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 |
| *AM2m* | 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 |
| *SC* | 2564.84 | -1275.42 | 2160.72 | 1 | 0.291 | 0.354 | 0.215 | 0.521 | - | - | 1.011 | - | 0.023 | - | 0.981 |
| **SC2m** | **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 |

716    **Figure legends**

717    **Figure 1:** Map of sampling locations: **1)** Dakar, **2)** Cadiz, **3)** Annaba, **4)** Mellah Lagoon, **5)**
718    Tabarka, **6)** Bizerte Lagoon, **7)** Gulf of Tunis, **8)** Kerkennah Islands, **9)** El Bibane Lagoon,
719    **10)** Bardawill Lagoon.

720    **Figure 2:** Population genetic structure of *Solea senegalensis* and *Solea aegyptiaca* analysed
721    using 10,758 SNPs. Principal Component Analysis (PCA) of the 153 individuals with panel
722    **a)** representing individual coordinates along PC1 and PC2 axes and panel **b)** showing PC1
723    and PC3 axes. **c)** *fastStructure* analysis performed with $K$=2 genetic clusters. Numbers
724    indicate the 10 different sampling locations.

725    **Figure 3:** Demographic divergence history of *Solea senegalensis* and *Solea aegyptiaca*. **a)**
726    The observed joint allele frequency spectrum (JAFS) for *S. aegyptiaca* (AEG, y axis) and *S.*
727    *senegalensis* (SEN, x axis) showing the number of SNPs (coloured scale on the right side of
728    each spectrum) per bin of derived allele counts using 20 individuals per species. **b)** The
729    maximum-likelihood JAFS obtained under the best-fit model of secondary contact with
730    heterogeneous introgression rates (SC2$m$ model). **c)** The maximum-likelihood JAFS obtained
731    from *msms* simulations under the free introgression model (Υ). **d)** The maximum-likelihood
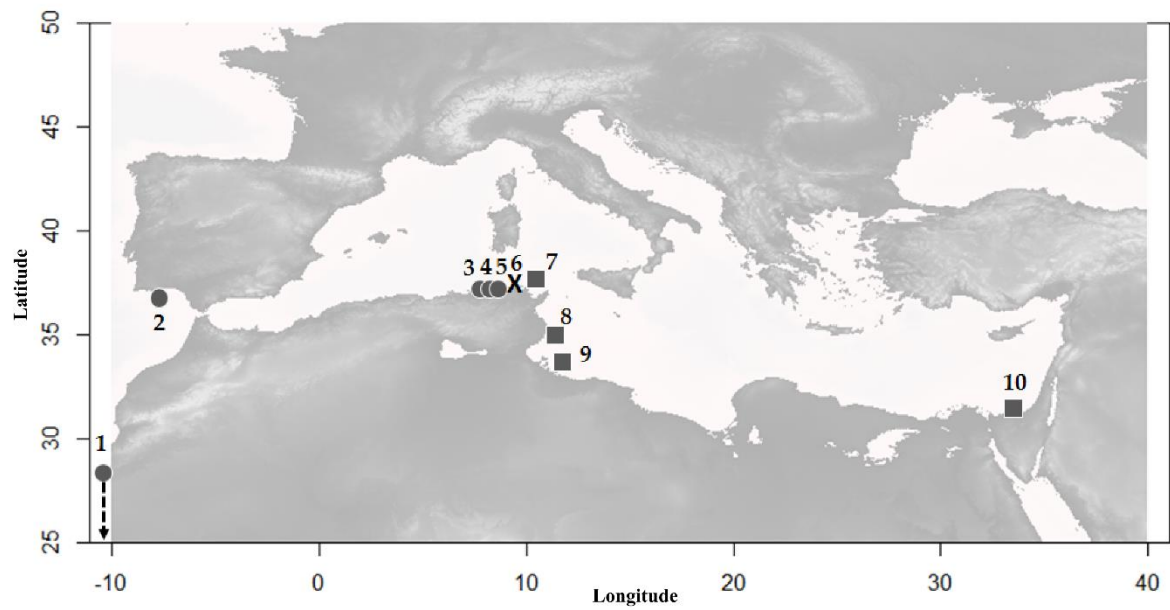732    JAFS obtained from *msms* simulations under the reduced introgression model (Φ).

733    **Figure 4:** Distribution of geographic cline slope ($S$) for the 10,758 loci as a function of their
734    cline center ($C$). The box represents a zoom in the central part of the contact zone.

735    **Figure 5:** Correlation between genomic cline parameter ($\alpha$) and geographic position of the
736    cline center ($C$) on both sides of the hybrid zone: **a)** in the *Solea senegalensis* geographic
737    range and **b)** in the *Solea aegyptiaca* geographic range.

738    **Figure 6:** Correlation between the geographic cline slope parameter ($S$) and positive values of
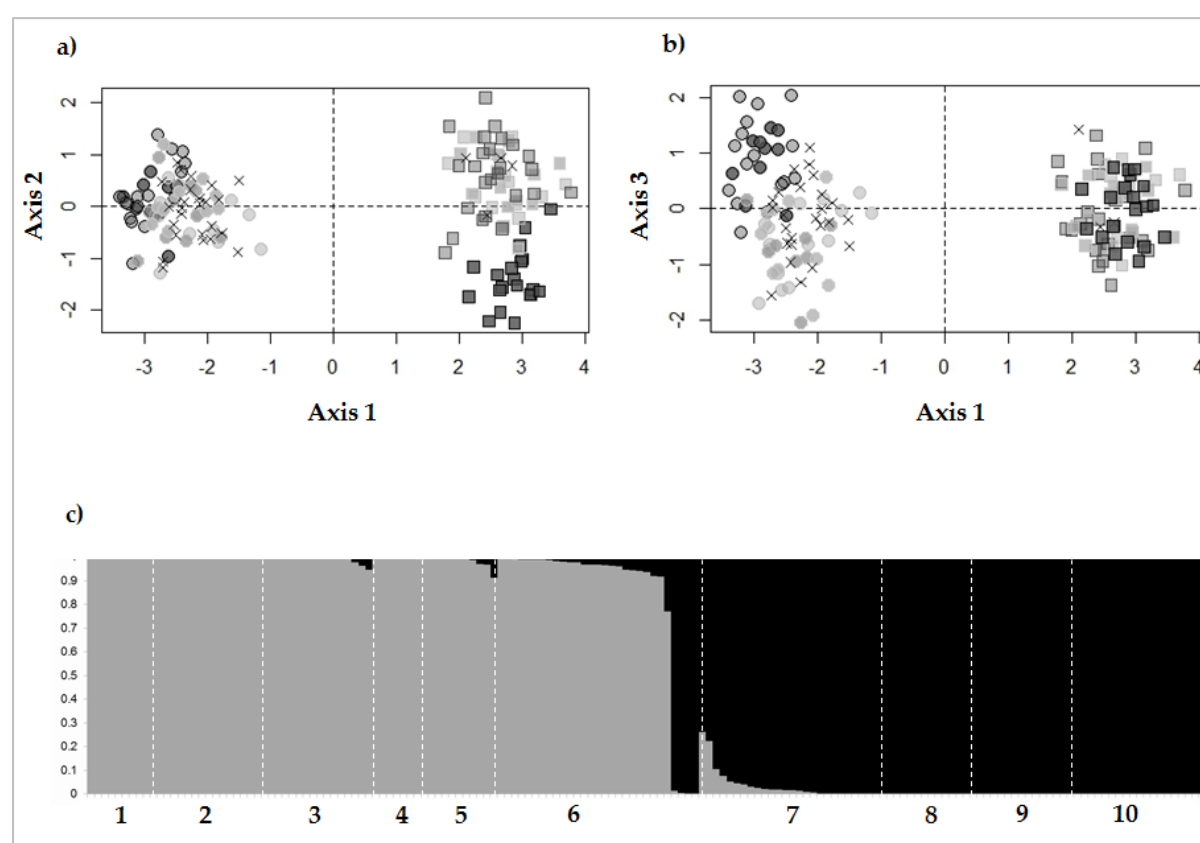739    the genomic cline parameter $\beta$.

740    **Figure 7:** Geographic distribution of **a)** the geographic cline center ($C$) and **b)** the genomic
741    cline parameter $\alpha$ for loci showing a probability of introgression greater or equal to 95%.
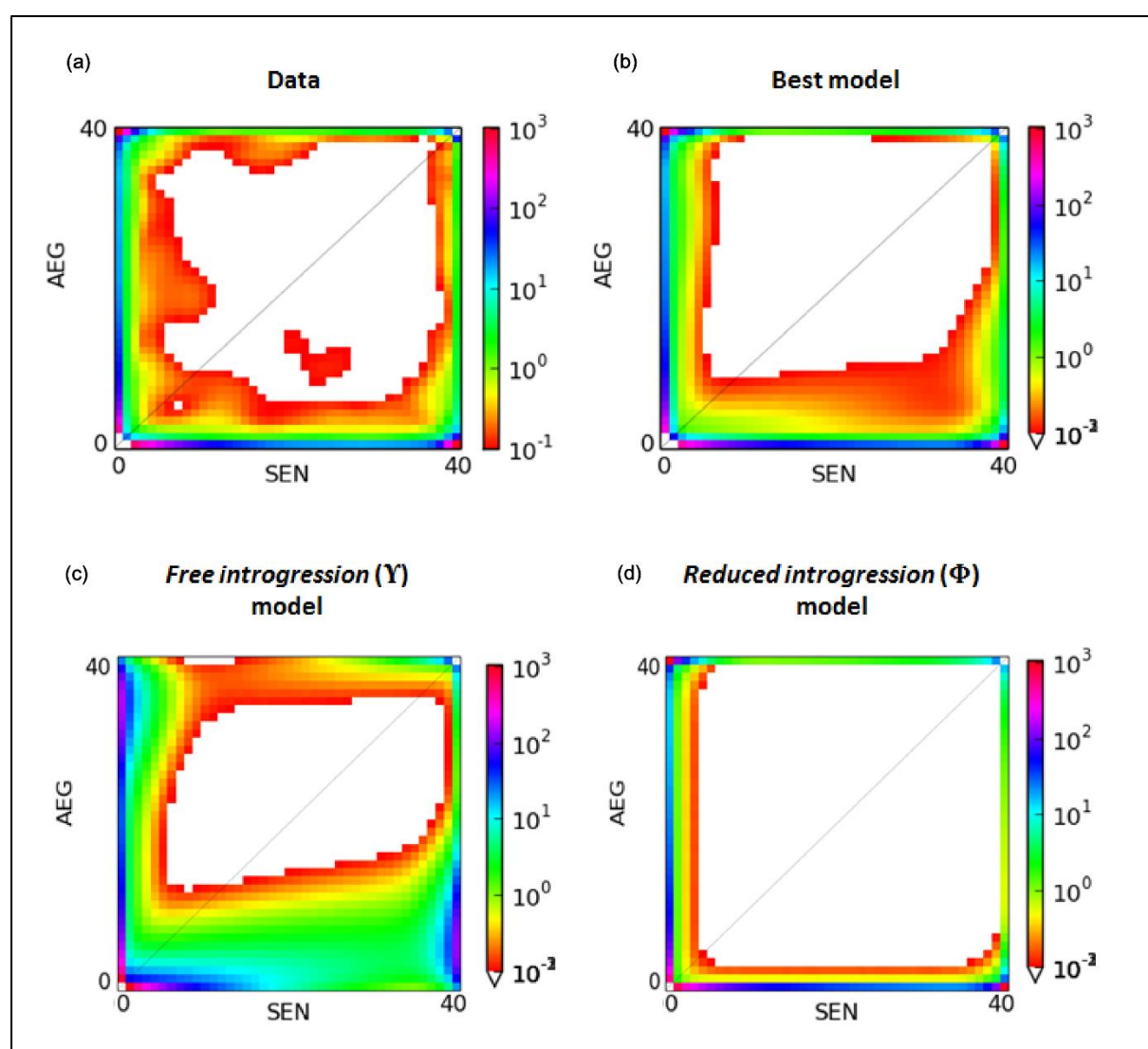
742

743    **Figure 1:**



744

745 **Figure 2 :**

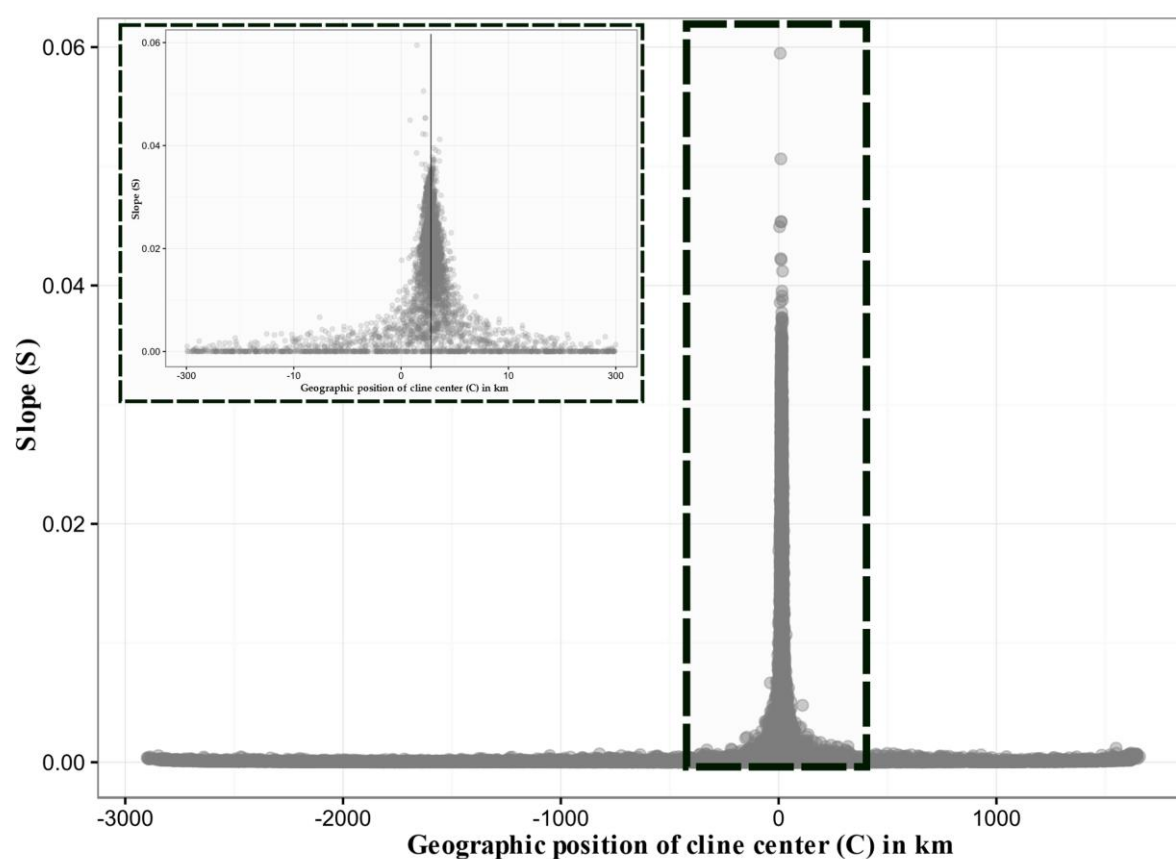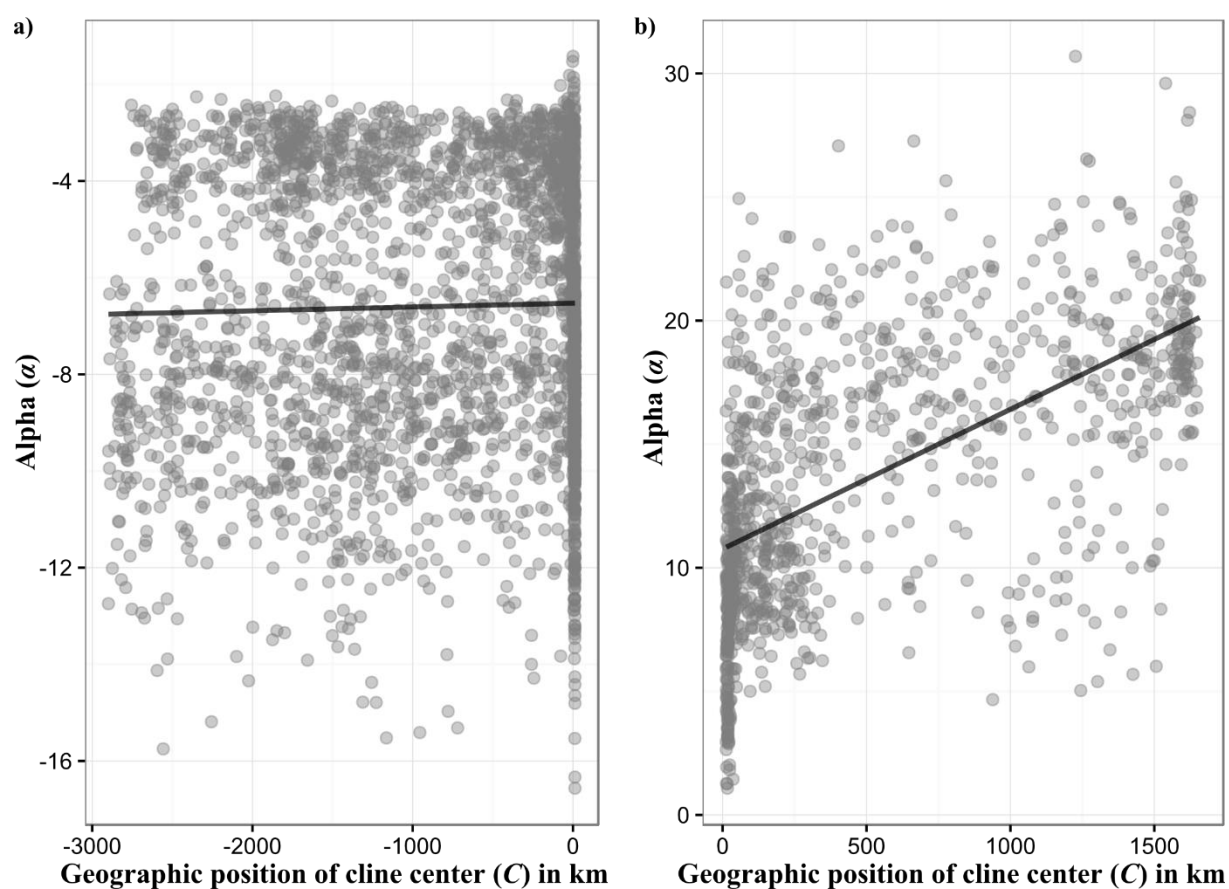

746

747 **Figure 3 :**



748

749    **Figure 4:**



750

751 **Figure 5:**



752

753  **Figure 6:**



754

755 **Figure 7:**



756