# Restricted dispersal in a sea of gene flow

# Benestan, L., Fietz, K., Loiseau, N.,Guerin, P.-E., Trofimenko, E., P., Rühs, S., Schmidt, C., Rath, W., Biastoch, A., Pérez-Ruzafa, A., Baixauli, P., Forcada, A., Arcas, E., Lenfant, P., Mallol, S., Goñi, R.8, Velez, L., Höppner, M., Kininmonth,S., Mouillot, D., Puebla, O., Manel, S.

**Electronic Supplementary Material**

# Supplementary text

**Text S1**:

Filtering (81). After running the *process\_radtags* pipeline in STACKs, we kept only individuals with > 1,000,000 reads, the removal of reads with more than one mismatch in the barcode sequence, and the removal of low-quality reads (with an average raw phred-score < 20 within a 0.2 sliding window). Reads were further trimmed to a final length of 139bp. Taking advantage of paired-end information, we used clone\_filter to remove pairs of paired-end reads that matched exactly, as these are assumed to be PCR clones.

Standardized effect sizes (45). To rule out the possibility that changes in the IBD slope were due to a sample size effect, we standardized our dataset using the smallest number of pairwise comparisons at the smallest spatial scale of observation (0-5 km) for each spatial interval. This smallest number of pairwise comparisons ranged from 685 (*P. elephas*) to 2,346 (*D. sargus*). Error bars around slope estimates were generated by resampling 1,000 times the lowest number of pairwise comparisons, which implies that there are no error bars for the smaller spatial scale of observation. Standardized effect sizes values serve as a measure of departure from a pure null expectation. Departures from 0 indicate non-randomness: values greater than 1.96 or smaller than 1.96 mean that the is slope significantly greater or lower than expected at the *α* = 0.05 level.

**Text S2:** Input file of one genetic simulation. The range of parameters that were explored is presented in Table S4.

%%%%% SIMULATION PARAMETERS %%%%%%%%%%%%

Data\_File\_Name=ibdsim\_76

Genepop\_File\_Extension=.txt

Run\_Number=1

Random\_Seeds=87144630

Pause=Final

%%%%% MARKERS PARAMETERS %%%%%%%%%%%%%%%

Locus\_Number = 1500

Mutation\_Model = SNP

Polymorphic\_Loci\_Only = T

Ploidy=Diploid

%%%%%% OUTPUT FILE FORMAT OPTIONS %%%%%%%

Genepop=T

%%%%%% VARIOUS COMPUTATION OPTIONS %%%%%%%

%The code below can be specified in a single line %

DiagnosticTables=Iterative\_Identity\_Probability,Hexp,Fis

DiagnosticTables=Prob\_Id\_Matrix,Effective\_Dispersal,Iterative\_statistics

%%%%%%%% DEMOGRAPHIC OPTIONS %%%%%%%%%%%%%

%% LATTICE

Lattice\_Boundaries=absorbing

Total\_Range\_Dispersal=F

Lattice\_SizeX=3000

Lattice\_SizeY=300

Ind\_Per\_Pop=1

Void\_Nodes=1

Specific\_Density\_Design=false

Zone=F

%% SAMPLE

%% classical squared sample design:

Sample\_SizeX=300

Sample\_SizeY=1

Min\_Sample\_CoordinateX=1250

Min\_Sample\_CoordinateY=150

Ind\_Per\_Pop\_Sampled=1

%% DISPERSAL

Dispersal\_Distribution=Pareto

Immigration\_Control=Simple1DProduct

Total\_Emigration\_Rate=0.1

Dist\_max=10

Pareto\_Shape=0.5

Geometric\_Shape=NA

%%%%%% EndOfSettings %%%%%%%%

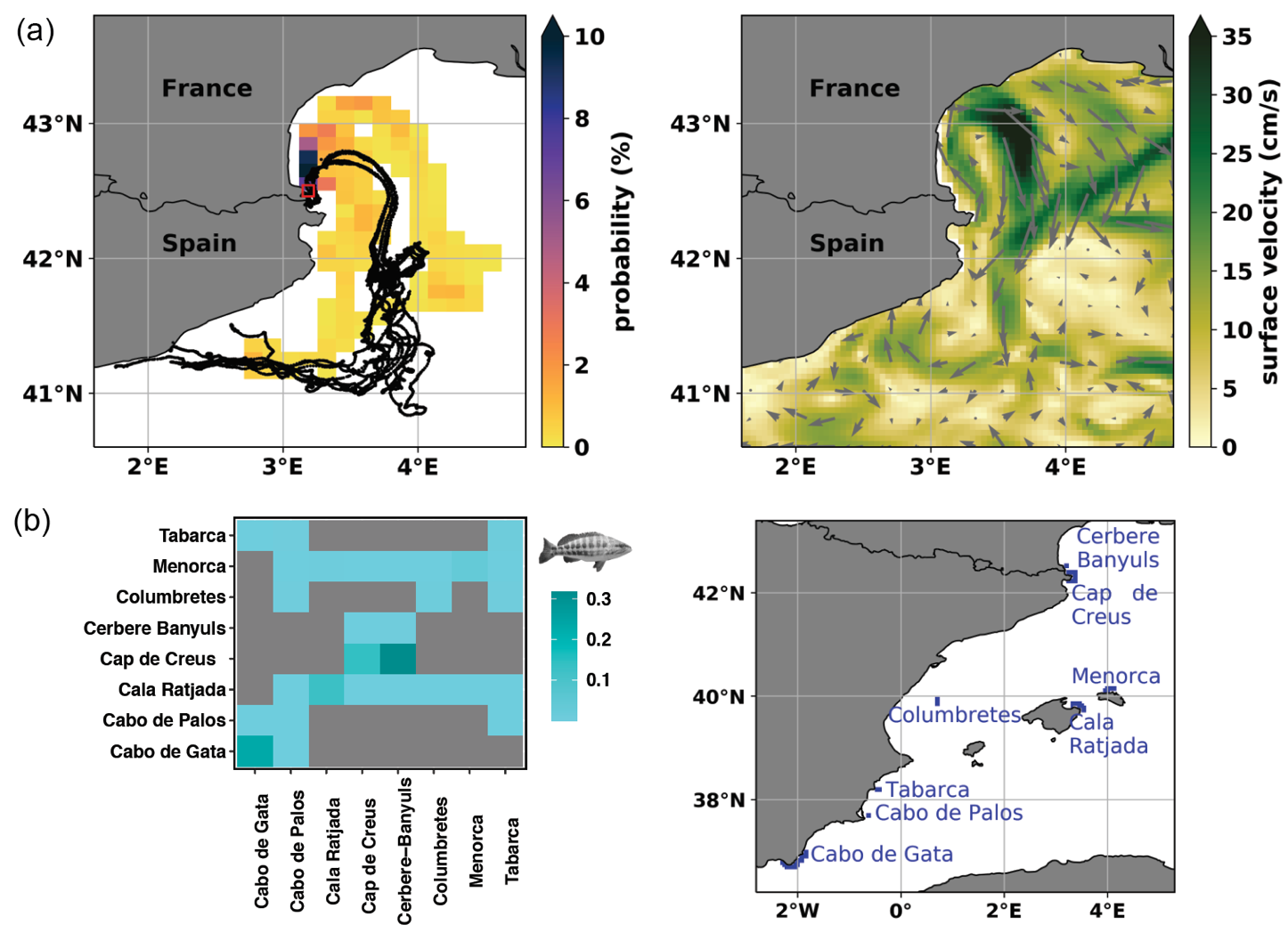
**Text S3: Details on the biophysical model**

To assess the performance of the modelling framework in the area of interest we tested how well it can simulate the observed tracks of 10 ARGOS-GPS (MAR-GE/T) surface drifters released within the marine protected area Cerbere-Banyuls on the 10th of July 2017 (Fig. S2a). Due to the turbulent nature of the ocean and uncertainties arising from the model’s numerical implementation and resolution one cannot expect that it exactly reproduces individual observed drifter tracks but has to use a probabilistic approach. We therefore simulated a set of O(104) virtual drifter trajectories by randomly releasing particles on the 10th of July over the four closest surface ocean grid cells which were then advected with the horizontal surface velocities with an integration time step of 5 minutes. The virtual drifter positions were stored hourly. The dispersal pattern after 20 days derived from the simulated drifters showed a reasonable agreement with the observed tracks when using not only the simulated surface currents but instead surface currents combined with Stokes drift for the velocities in the dispersal simulations. Remaining discrepancies may be partially related to the sail effect, which has not been considered within these simulations (and is not relevant for larval dispersal), as well as to (minor) model inaccuracies and to differences between ocean reanalysis and real currents.

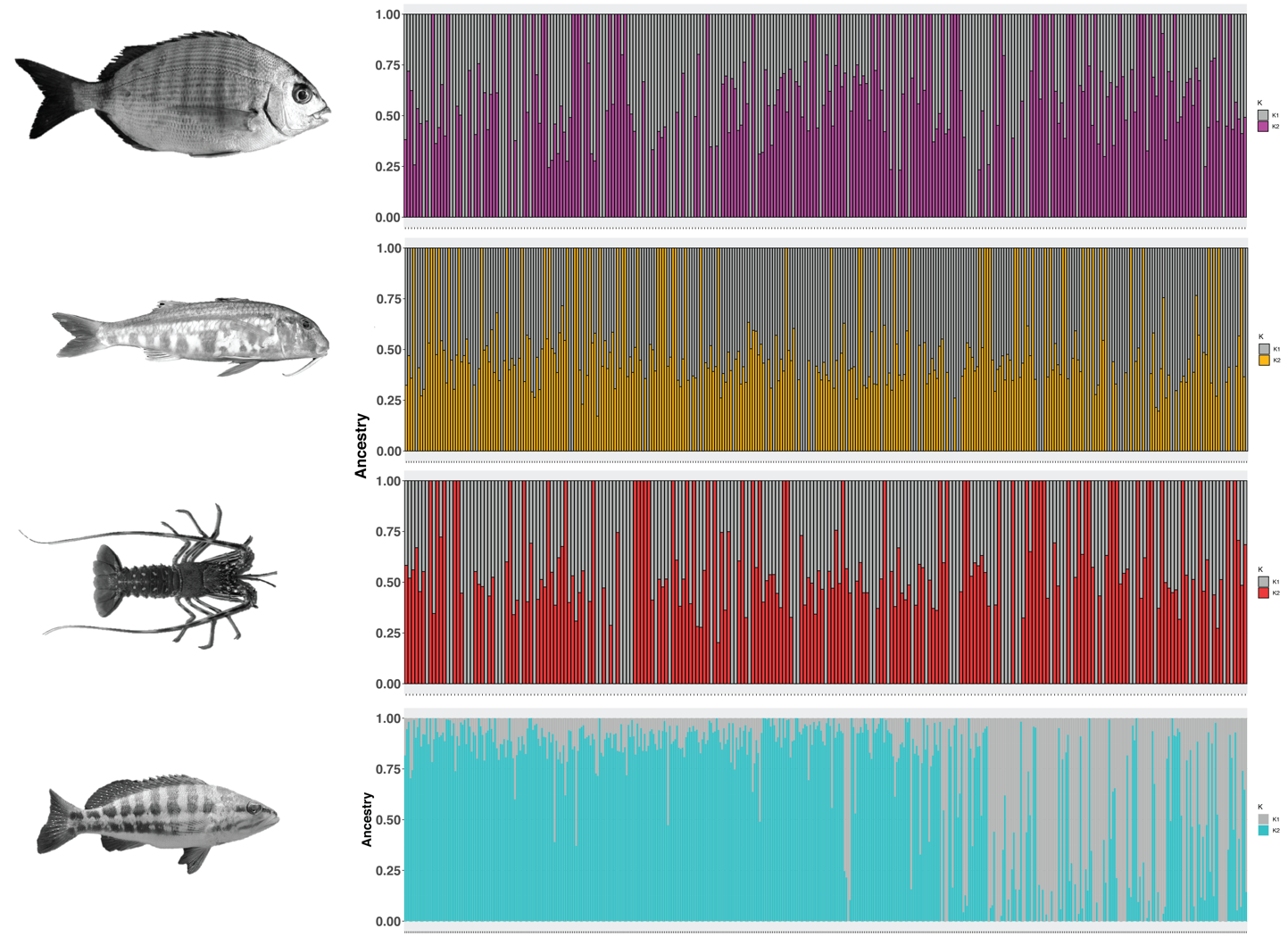
# Supplementary figures



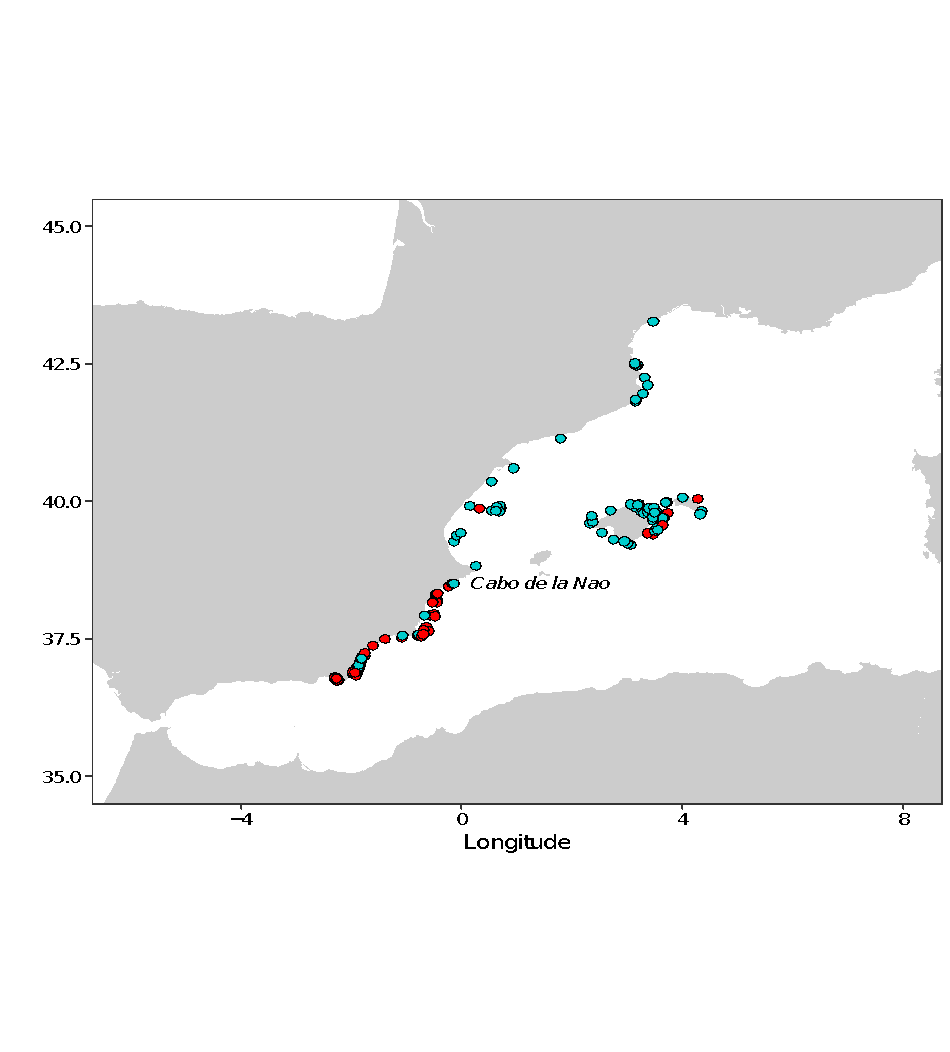
**Fig. S1.** Density distribution of the geographic distance between pairs of samples for each species (5-km bins). The distributions present hills and valleys, which suggests that individual densities are not homogenous across the study area, but we were nonetheless able to cover a continuous range distances from 1 to 950 km.



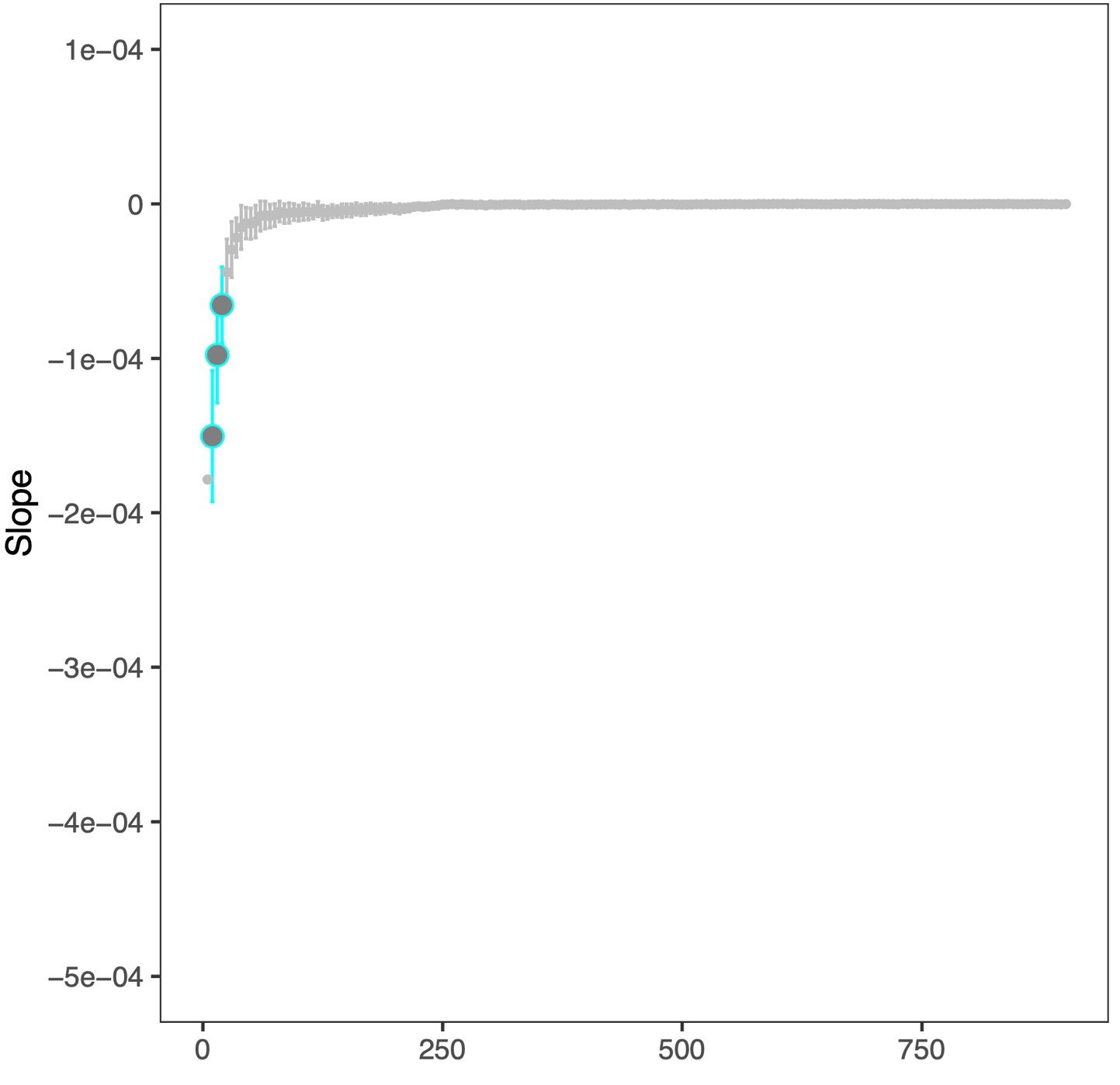
**Fig. S2. (a)** Comparison between simulated and observed surface drifter trajectories when Stokes drift was included in the biophysical model. Left: real surface drifter trajectories (black lines) and mean spreading pathways of simulated drifters (color shading, the darker the color the more drifter occurrences) for the first 20 drift days. Right: snapshot of surface velocities including Stokes drift on the drifter release day (10th of July 2017). **(b)** Left: connectivity matrix among eight marine protected areas (MPAs) for *Serranus cabrilla.* Right: Map of marine protected areas, which represent the larvae release regions.



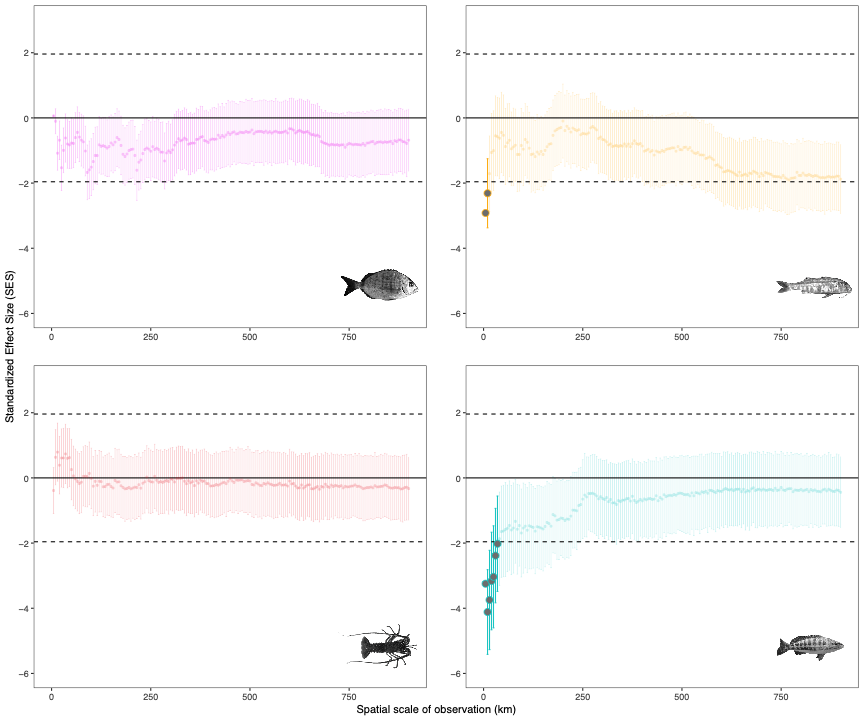
**Fig S3.** Admixture analysis of the four species *D. sargus*, *M. surmuletus, P. elephas* and *S. cabrilla* (from top to bottom). Each vertical bar represents one individual, with the colors indicating the likelihood of this individual belonging to a particular ancestral population considering *K* = 2. *K* refers to the number of ancestral populations that were assumed to be present in the dataset. Samples sorted according to latitude from left to right.



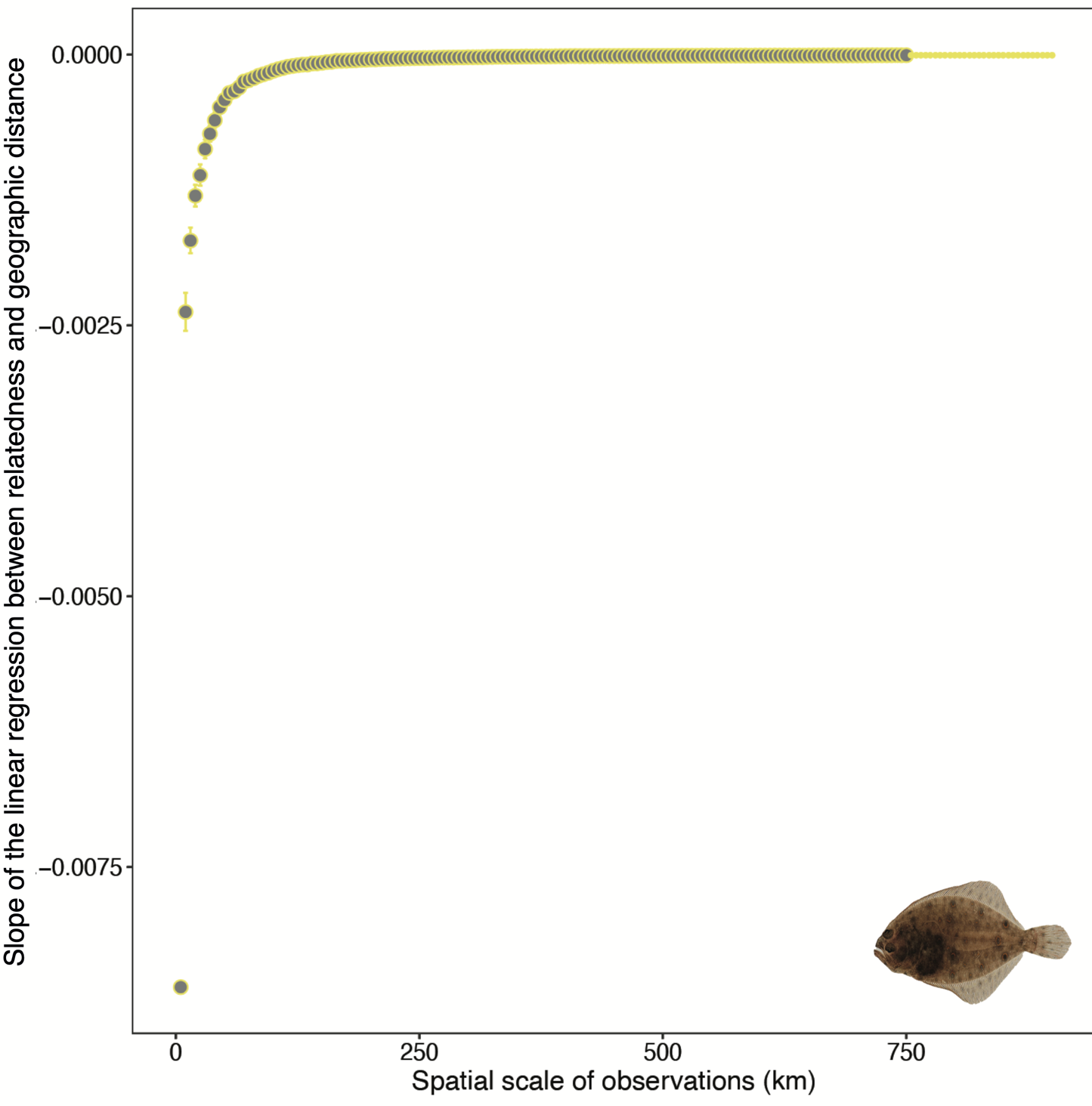
**Fig. S4.** Geographic distribution of the two genetic clusters identified in *Serranus cabrilla*. A northern and a southern group were identified, separated by the *Cabo de la Nao.*

****

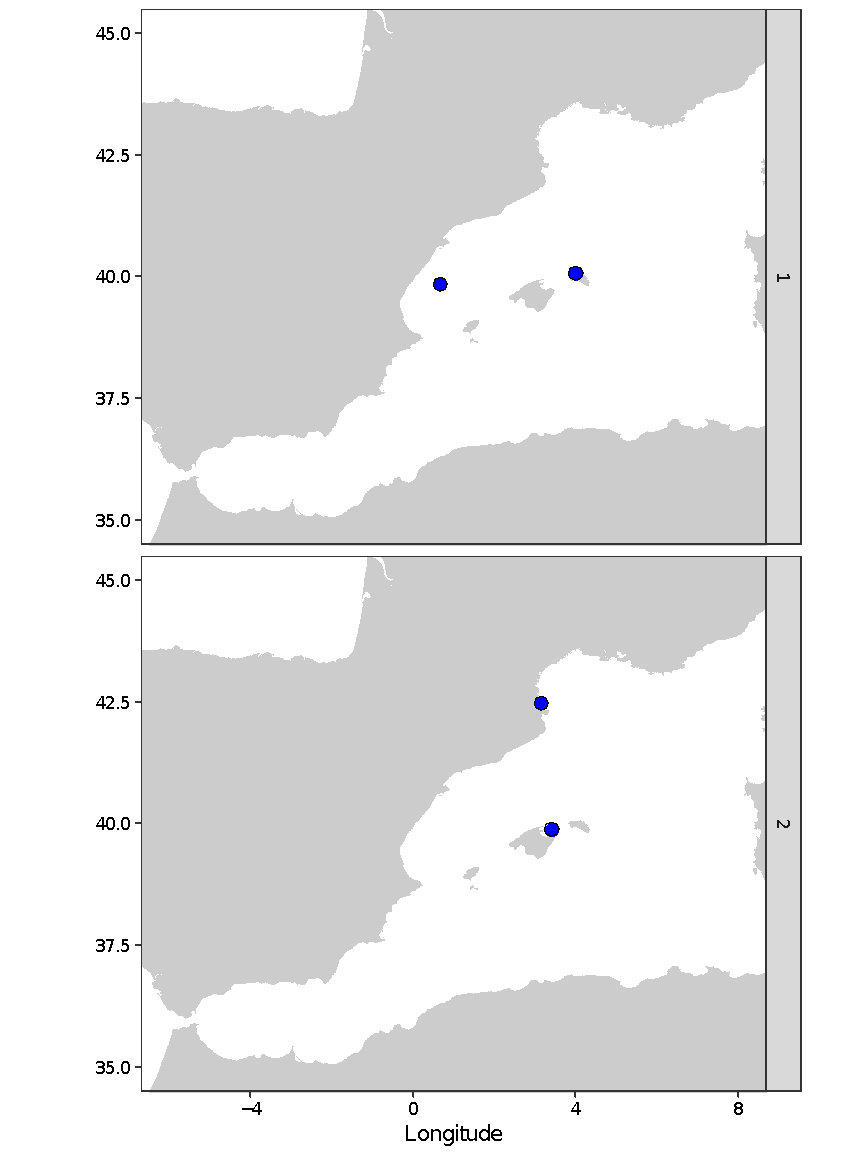
**Fig. S5.** Slope of the linear regression between genetic relatedness and geographic distance among pairs of *Serranus cabrilla* individuals at various spatial scales of observation after removal of five pairs of individuals that are close kin and geographically close. The slopes that are statistically significant following the Standardized Effect Size (SES) procedure (see Methods) are highlighted in bold. The broad pattern remains unchanged, indicating that the decrease at small spatial scales is not driven by such pairs.



**Fig. S6.** Standardized Effect Size (SES) analysis for *D. sargus, M. surmuletus, P. elephas* and *S. cabrilla* (from upper left to lower right). The horizontal dashed lines show the limits for statistical significance.

****

**Fig S7.** Slope of the linear regression between genetic relatedness and geographic distance among pairs of individuals at various spatial scales of observation for theSummer flounder (*Paralichthys dentatus*) dataset published by Hoey & Pinsky [1] [2]. Slope averages and error bars from a resampling procedure (see Methods). The slopes that are statistically significant following the Standardized Effect Size (SES) procedure (see Methods) are indicated in greater size.



**Fig. S8.** The two pairs of *Serranus cabrilla* samples that are closely related genetically but geographically distant.

# Supplementary Tables

**Table S1:** Biological parameters of the four target species, from the FishMed database [2] and the references indicated below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Order** | **Perciformes** | **Perciformes** | **Perciformes** | **Decapoda** |
| Family | Sparidae | Mullidae | Serranidae | Palinuridae |
| Species | *Diplodus sargus* | *Mullus surmuletus* | *Serranus cabrilla* | *Palinurus elephas* |
| IUCN Red List Category | DD | LC | LC | VU |
| Reproductive mode | Spawning | Spawning | Spawning | Spawning |
| Minimum depth (m) | 1 | 3 | 5 | 5 |
| Maximum depth (m) | 40 | 80 | 500 | 150 |
| Reproduction | Oviparous | Oviparous | Oviparous | Oviparous |
| Mating system | Protandrous hermaphrodite | Gonochoric | Simultaneous hermaphrodite | Gonochoric |
| Adult behavior | Gregarious | Gregarious | Solitary | Gregarious |
| Spawning season | March - June | March - July | February – July | December-April |
| Spawning migration | Yes [3] | Unknown | Unknown | Yes [4,5] |
| Larval vertical migration | no | no | no | yes |
| Pelagic larval duration (PLD) | 26-30 days[6] | 25-35 days [6] | 21-28 days [6] | 4-6 months [7] |
| Dispersal potential in terms of PLD | medium | medium | low | high |
| Habitat | Rock/Posidonia | Sand | Sand/Rock/Posidonia | Rock |

**Table S2**: Filtering steps for the genetic data. The table indicates the number of loci remaining after each filtering step. The data were filtered with respect to missing data (<30%), inbreeding coefficient (-0.5 ≤ *FIS* > 0.5), linkage disequilibrium (LD, >5000 base pairs physical distance and *R2* <0.8), minor allele frequency (MAF, > 0.01), Hardy-Weinberg equilibrium (HWE) and selection (pcadapt). To do so, we used the R-package pcadapt [5]to identify loci putatively under selection, and removed them to obtain a neutral dataset. We set *K*=1 and used a Bonferroni correction with a P-value of 0.05 to identify and remove outlier loci from the putatively neutral dataset identified for each species. Final number of loci highlighted in bold.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Filtering step** | **Vcftools command** | ***Palinurus elephas*** | ***Mullus surmuletus*** | ***Diplodus sargus*** | ***Serranus cabrilla*** |
| *STACKs (output from populations)* |  | 126,598 | 37,764 | 47,588 | 101,347 |
| Missing < 30% | max-missing 0.3 | 53,296 | 37,764 | 47,588 | 93,680 |
| ≤ -0.5 *FIS* > 0.5 | *custom script* | 46,368 | 35,772 | 44,875 | NA |
| LD5000 | - | NA | 16,828 | 21,585 | NA |
| LD *R*2 < 0.8 | geno-r2 | 45,684 | 16,744 | 21,509 | NA |
| MAF > 0.01 | maf 0.01 | NA | 16,744 | 21,509 | 64,057 |
| HWE (P-value = 0.01) | hwe 0.01 | 27,408 | 14,768 | 19,059 | 13,282 |
| Pcadapt Outliers |  | **25,230** | **14,318** | **18,512** | **13,101** |

**Table S3:** Correlations between our estimator (here named Ajk) and 5 other methods to estimate the pairwise relatedness between individuals of *Serranus cabrilla* obtained with the CO-ANCESTRY software [8]. The relatedness estimates named wang, lynchli, lynchrd, ritland and quellergt are respectively described in Wang [9], Li *et al.* [10], Lynch & Ritland [11], Ritland [12] and Queller & Godnight [13].

****

**Table S4:** Range of parameters tested in the genetic simulations. We tested the effects of varying the number of loci (in steps of 100), the dispersal function and the total emigration rate for the geometric dispersal distribution (in steps of 0.1). For all simulations, ploidy was set to diploid and total range dispersal = false. Sample size = 300 on the x axis of the lattice and 1 on the y axis (min sample coordinate x = 1250, min sample coordinate y = 150). SS = stepping stone, G = geometric, P = pareto.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Simulation ID** | **Dispersal distribution** | **Total Emigration Rate (m)** | **Pareto Shape (n)** | **Geometric Shape (g)** |
| 106 (Fig. 3a) | SS | 0,1 | NA | NA |
| 75 (Fig. 3b) | G | 0,1 | NA | 1 |
| 76 (Fig. 3c) | P | 0,1 | 0,5 | NA |
| 77 (Fig. 3d) | P | 0,1 | 1 | NA |
| 78 | P | 0,1 | 1,5 |  |
| 79 | 2 |
| 80 | 2,5 |
| 81 | 3 |
| 82 | 3,5 |
| 83 | 4 |
| 84 | 4,5 |
| 85 | 5 |
| 86 | 5,5 |
| 87 | 6 |
| 88 | 6,5 |
| 89 | 7 |
| 115 | 8 |
| 116 | 9 |
| 117 | 10 |
| 118 | 11 |
| 119 | 12 |
| 120 | 13 |
| 121 | 14 |
| 122 | 16 |
| 123 | 18 |
| 66 | G | 0,1 | NA | 0,1 |
| 67 | 0,2 |
| 68 | 0,3 |
| 69 | 0,4 |
| 70 | 0,5 |
| 71 | 0,6 |
| 72 | 0,7 |
| 73 | 0,8 |
| 74 | 0,9 |
| 124 | 0,01 |
| 125 | 0,001 |
| 126 | 0,0001 |
| 127 | 0,00001 |
| 128 | 0,000001 |
| 129 | 0,0000001 |
| 130 | 0 |
| 107 | SS | 0,2 | NA | NA |
| 108 | 0,3 |
| 109 | 0,4 |
| 110 | 0,5 |
| 111 | 0,6 |
| 112 | 0,7 |
| 113 | 0,8 |
| 114 | 0,9 |

**Table S5**: Test of genetic isolation by distance across the entire study area for the four target species and the *P. dentatus* dataset [1] that we reanalyzed. Mantel’s r, one-tailed p-values (null-hypothesis r ≤ 0) and 95% confidence intervals (CI) estimated with the R package e*codist* with 10,000 permutations *(46)*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Species** | **Mantel r** | **P-value** | **CI 2.5%** | **CI 97.5%** |
| *D. sargus* | -0,016 | 0,999 | -0,020 | -0,011 |
| *M. surmuletus* | -0,042 | 1 | -0,046 | -0,037 |
| *S. cabrilla* | -0,376 | 1 | -0,389 | -0,361 |
| *P. elephas* | -0,010 | 0,968 | -0,016 | -0,004 |
| *P. dentatus* | 0,026 | >0.86 | NA | NA |

# References

1. Hoey JA, Pinsky ML. 2018 Genomic signatures of environmental selection despite near-panmixia in summer flounder. *Evol. Appl.* **11**, 1732–1747. (doi:10.1111/eva.12676)

2. Albouy C *et al.* 2015 FishMed: traits, phylogeny, current and projected species distribution of Mediterranean fishes, and environmental data. *Ecology* (doi:10.1890/14-2279.1)

3. Planes S, Lenfant P. 2002 Temporal change in the genetic structure between and within cohorts of a marine fish, Diplodus sargus, induced by a large variance in individual reproductive success. *Mol. Ecol.* **11**, 1515–1524. (doi:10.1046/j.1365-294X.2002.01521.x)

4. Goñi R, Latrouite D. 2005 Review of the biology, ecology and fisheries of Palinurus spp. species of European waters: Palinurus elephas (Fabricius, 1787) and Palinurus mauritanicus (Gruvel, 1911). *Cah. Biol. Mar.* **46**, 127-142.

5. Abecasis D, Bentes L, Lino PG, Santos MN, Erzini K. 2013 Residency, movements and habitat use of adult white seabream (Diplodus sargus) between natural and artificial reefs. *Estuar. Coast. Shelf Sci.* **118**, 80-85 (doi:10.1016/j.ecss.2012.12.014)

6. Galarza JA, Turner GF, Macpherson E, Rico C. 2009 Patterns of genetic differentiation between two co-occurring demersal species: The red mullet (Mullus barbatus) and the striped red mullet (Mullus surmuletus). *Can. J. Fish. Aquat. Sci.* **66**, 1478-1490 (doi:10.1139/F09-098)

7. Follesa MC, Cuccu D, Cannas R, Sabatini A, Deiana AM, Cau A. 2009 Movement patterns of the spiny lobster Palinurus elephas (Fabricius, 1787) from a central western Mediterranean protected area. *Sci. Mar.* **73**, 499-506 (doi:10.3989/scimar.2009.73n3499)

8. Wang J. 2011 Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol. Ecol. Resour.* **11**, 141-145 (doi:10.1111/j.1755-0998.2010.02885.x)

9. Wang J. 2002 An estimator for pairwise relatedness using molecular markers. *Genetics* **160**, 1203-1215 (doi:10.1093/genetics/160.3.1203)

10. Li CC, Weeks BE, Chakravarti A. 1993 Similarity of DNA fingerprints due to chance and relatedness. *Hum. Hered.* **43**, 45-52 (doi:10.1159/000154113)

11. Lynch M, Ritland K. 1999 Estimation of pairwise relatedness with molecular markers. *Genetics* **152**, 1753-1766 (doi:10.1093/genetics/152.4.1753)

12. Ritland K. 1996 Estimators for pairwise relatedness and individual inbreeding coefficients. *Genet. Res.* **67**, 175-185 (doi:10.1017/s0016672300033620)

13. Queller DC, Goodnight KF. 1989 Estimating Relatedness Using Genetic Markers. *Evolution (N. Y).* **8**, 70-73 (doi:10.2307/2409206)