Ecotype formation in the European anchovy fuelled by structural variants of different origins and genetic interactions with a southern lineage

Laura Meyer, Pierre Barry, Christine Arbiol, Rita Castilho, Carl van der Lingen, Malika Chlaïda, Niall J Mckeown, Bruno Ernande, Alan Le Moan, François Bonhomme, et al.

Ecotype formation in the European anchovy fuelled by structural variants of different origins and genetic interactions with a southern lineage

Authors

Laura Meyer[1#] , Pierre Barry[1,2], Christine Arbiol[1], Rita Castilho[3], Carl Van der Lingen[4], Malika Chlaïda[5], Niall J McKeown[6], Bruno Ernande[7], Alan Le Moan[8], François Bonhomme[1*], Pierre-Alexandre Gagnaire[1*], Bruno Guinand[1*]

[1] ISEM, Univ. Montpellier, CNRS, EPHE, IRD, Montpellier, France
[2] CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos Universidade do Porto, Vairão, Portugal
[3] CCMAR, Centro de Ciências do Mar, Universidade do Algarve, Campus de Gambelas, Faro, Portugal
[4] Department of Biodiversity and Conservation Biology, University of the Western Cape Town, Cape Town, South Africa
[5] Research and Development Unit on Marine Biology, National Institute of Fisheries Research, Casablanca, Morocco
[6] Aberystwyth University, Department of Life Sciences, Ceredigion, Wales, UK
[7] MARBEC, Univ. Montpellier, Ifremer, CNRS, IRD, Montpellier, France
[8] CNRS-Sorbonne Université, Station Biologique de Roscoff, Roscoff, France

[#] Corresponding author: laura.meyer@umontpellier.fr
* Co-last authors

Abstract

The speciation of ecotypes can unfold in diverse ways and likely depends on multiple processes. The variants involved in ecotype divergence can include new mutations as well as older allelic variation that evolved in different contexts. Among the different types of variants that can contribute to reproductive isolation between ecotypes, structural variants (SVs) represent likely candidates due to their ability to protect divergent haplotypes from recombination and gene flow. The European anchovy (*Engraulis encrasicolus*) is known to be subdivided into marine and coastal ecotypes, and their divergence shows patterns that are consistent with SVs. Here, we present the first genome-scale study investigating genetic structure in the *E. encrasicolus* species complex. We generated a reference genome and produced whole-genome resequencing data for anchovies from the North-East Atlantic and Mediterranean Sea, as well as from South Africa. We complemented this approach

with the analysis of RAD-seq data in order to study ecotypic structure across the entire distribution range. We found that genetic diversity is not only characterised by the presence of two genetic clusters, namely the marine and coastal ecotypes, but also by a third ancestry which corresponds to a southern Atlantic lineage. This lineage occurs off South Africa but also in southern Morocco and the Canary Islands, and shows a gradient of admixture with northern populations nearing the Atlantic-Mediterranean transition zone. Genomic landscapes of differentiation showed evidence for large regions of high linkage disequilibrium, likely representing SVs that differentiate the three anchovy lineages. We found evidence that three of the SVs contributing to the gene flow barrier between ecotypes originated in the southern lineage, suggesting that the coastal and southern lineages have a partly shared evolutionary history. In addition to these barriers, three other SVs contributing to ecotype differentiation appear to have evolved *in situ*. Anchovies thus present an interesting case for the study of ecotype speciation, since the barriers involved in reproductive isolation have different origins and have partly diverged in geographic isolation.

Introduction

Ecotype formation has classically been studied as an intermediate step in ecological speciation. However, recent avenues of research have highlighted that speciation is a multi-step and multi-faceted process that depends on an interplay between many different factors (Bolnick et al., 2023; Johannesson et al., 2024). The speciation of ecotypes is no exception, and the mechanisms involved have proven to be more complex than previously thought. Many aspects remain poorly understood and questions about the chronology and context of ecotype formation are among the most puzzling. Studies that use the age of habitat formation or the time of divergence inferred from neutral loci tend to see the formation of ecotypes as a rapid process. By contrast, genomic regions associated with ecotype divergence often show much older allelic variation, with divergence times sometimes far exceeding dates derived from paleoenvironmental or demographic reconstructions.

Reconciling these different views of ecotype speciation requires understanding the origin of the genetic variation involved in the build-up of reproductive isolation (RI) mechanisms. Several lines of evidence have shown how standing genetic variation that plays a role in recent divergence can correspond to ancient variants that possibly evolved in different demographic, genomic and environmental contexts. Specifically, this “two-time frame” model has been used to describe how alleles that diverged in geographic isolation can subsequently be spatially redistributed and contribute to RI between lineages that are in close contact (Belleghem et al., 2018). One possible origin for such divergent alleles is through introgression with a closely related species or a divergent lineage. For example, if allopatric species or lineages come into secondary contact, RI might emerge due to intrinsic incompatibilities between the lineages or adaptation to different environments (Kulmuni et al., 2020). However, the barriers involved in RI between lineages or ecotypes may have diverse origins, and may have partly evolved in allopatry and partly in sympatry (Butlin et al., 2008). In addition to ancient variants that play a role in RI, other barriers could correspond to alleles that are involved in local adaptation or the reinforcement of RI between nascent lineages (Butlin & Smadja, 2018).

Not all genetic architectures seem to be equally likely to contribute to ecotype formation. Structural variants (SVs) that include chromosomal inversions, fusions/fissions, duplications or translocations could represent a type of variation that is disproportionately involved in ecotype divergence

(Campbell et al., 2021; Gould et al., 2017; Hager et al., 2022; Jay et al., 2022; Lundberg et al., 2023). This is because recombination is largely suppressed in SVs like chromosomal inversions, facilitating the accumulation of divergence and protecting haplotypes from the homogenising effect of gene flow (Kirkpatrick & Barton, 2006; Navarro & Barton, 2003; Rieseberg, 2001). Due to these properties, SVs may represent "pre-packaged" divergent haplotypes that can contribute to RI following their introgression (Della Torre et al., 1997; Edelman et al., 2019; Jay et al., 2018). Similarly, SVs have been hypothesised to better resist re-homogenisation after secondary contact compared to the collinear genome (Hoffmann & Rieseberg, 2008; Noor et al., 2001; Rafajlović et al., 2021). The role of SVs in RI has often been ascribed to local adaptation and thus linked to the formation of ecotypes, since an SV may capture multiple loci subject to environmental selection (Kirkpatrick & Barton, 2006). However, SVs may also contain co-adapted alleles or Dobzhansky-Muller incompatibilities that promote the isolation of nascent lineages carrying opposite haplotypes (Navarro & Barton, 2003). Recent evidence indicates that SVs may thus provide a sufficient barrier to gene flow to ensure at least moderate RI between lineages, especially if there are multiple SVs separating the lineages and if there is coupling between the SVs (Le Moan et al., 2024).

Here, we studied speciation and the role of SVs in RI in the European anchovy, *Engraulis encrasicolus sensu lato*. This species complex has been shown to be subdivided into a marine ecotype (offshore and pelagic, *E. engraulis s. stricto*) and a coastal ecotype (nearshore, lagoonal and estuarine, *E. maeoticus*) that are able to co-exist in quasi-sympatry despite frequent hybridisation (Le Moan et al., 2016). Partial reproductive isolation between ecotypes is reflected in their differences in genetic makeup as well as their phenotypic characteristics (for a review see Bonhomme et al., 2021). The divergence between ecotypes involves a condensed genomic architecture, suggesting a potential role for SVs in the speciation of anchovy ecotypes (Le Moan et al. 2016). Furthermore, the most likely demographic model underlying ecotype divergence involves secondary contact after a long period of geographic isolation during the last glacial maximum. The current distribution range of *E. encrasicolus* covers a very large area of the North-East Atlantic, Mediterranean and Black seas. Classically, the southern range limit of *E. encrasicolus* along the western African coast is considered as the south of the Gulf of Guinea, while *E. capensis* has been described further south in the Benguela system off South Africa. However, the clear occurrence of *E. capensis*-like genotypes around the Atlantic-Mediterranean transition zone (Silva et al., 2017; Zarraonaindia et al., 2012) shows that there is relatively little genetic difference between these anti-tropical entities. Hence, *E. capensis* could be considered as the third component of the *E. encrasicolus* species complex, with still unknown consequences on the divergence between the marine and coastal ecotypes.

We present the first study using whole-genome sequencing (WGS) to provide a full description of the genetic structure in the *E. encrasicolus* species complex. Through reference-based mapping and anchoring to a chromosome-scale reference genome, we reveal the genomic architecture associated with divergence between the different lineages of anchovy. We combined this WGS approach with the use of RAD-sequencing data to describe the eco-geographic structure at a large geographical scale. In the northern part of the range, our sampling scheme included anchovies from the marine and coastal ecotypes. We also included anchovies collected in the Canary Islands, off the coast of Morocco and off South Africa. We aimed to determine whether and how these lineages have genetically interacted during their evolutionary history and whether SVs have played a role in their divergence.

Materials and methods

Sampling and DNA extraction

Samples were collected from multiple sites covering a large part of the species distribution area (Locations table, **Supplementary Table S2**) and were issued from various sampling expeditions and local fisheries (**Supplementary Table S1**, Type==”Tissue”). These samples were collected in different types of habitats, which were classified either as coastal or marine habitats. Also included in our sampling scheme were eight samples of anchovy collected off the South African coast (Gqeberha). Whole genomic DNA was extracted from muscle tissue or fin clips using commercial tissue kits (Qiagen and Macherey-Nagel). Extraction quality was checked on agarose gel for the presence of high molecular weight DNA, and double-stranded nucleic acid concentration was measured using Qubit 2.0 and standardised in concentration before library construction.

Reference genome assembly

We performed high-coverage linked-read (10X genomics) sequencing of a marine Atlantic *E. encrasicolus* individual from the Faro location (Algarve) to generate a new reference genome assembly (hereafter called Eencr_V1), following the same methodology as in Meyer et al. (2024). The *de novo* assembly obtained by analysing preprocessed linked-reads (raw coverage ~40X) with supernova v2.1.1 (Weisenfeld et al., 2017) reached a total length of ~926 Mb (925,873,119 bp; contig N50=13.08 kb; scaffold N50=20.36 kb) (**Supplementary Fig. S1**). All downstream analyses required for variant calling from RAD-seq and WGS data were performed on the subset of scaffolds longer than 10 kb to account for assembly fragmentation. Genomic landscapes of differentiation and local PCA (see below) were reconstructed by anchoring these scaffolds to the recently released chromosome-level assembly of an *E. encrasicolus* individual from the Black Sea (GenBank assembly accession: GCA_034702125.1) (**Supplementary Fig. S2**). Whole-genome alignment between our Eencr_V1 reference genome and the new assembly was performed with Minimap2 (Li, 2018) and visualised using D-GENIES (Cabanettes & Klopp, 2018).

Whole-genome resequencing data

Thirty-nine samples (**Supplementary Table S1**, WGS==”yes”) were selected for whole-genome sequencing (WGS), including samples from coastal and marine habitats in the Atlantic (*GAS*) and the Mediterranean Sea (*GDL*, *SPN*). We also included samples from the Atlantic-Mediterranean transition zone (*PRS*) and from South Africa to investigate the genetic makeup in these localities. Individual whole-genome sequencing libraries were prepared following the Illumina TruSeq DNA PCR-Free Protocol and sequenced to an average depth of ~10-30X on an Illumina NovaSeq 6000. Raw demultiplexed reads were processed using fastp (v0.19.05) (Chen et al., 2018) and aligned to our Eencr_V1 reference genome using BWA-MEM (BWA v0.7.17; Li, 2013). Picard (v2.26.8) (“Picard toolkit”, 2019) was used for sorting read alignments, marking duplicates and adding read groups.

Variants were called using the GATK best practices workflow (McKenna et al., 2010; Van der Auwera et al., 2013), without performing variant and base quality score recalibration steps. Firstly,

individual GVCF files were created from bam files with HaplotypeCaller (GATK v.4.3.0.0). This information was then stored in a GVCF database using GenomicsDBImport, and VCF files (one file per scaffold) were generated with GenotypeGVCFs. These files were concatenated into a single VCF file which was filtered using VCFtools v0.1.16 (Danecek et al., 2011) to retain only high quality SNPs. This included recoding genotypes as missing for low-quality calls (*--minGQ 20*) and hard-filtering sites based on their normalised variant quality, average genotype quality, mapping quality, strand bias, and average depth (greater than 90X, corresponding to the 97.5th quantile of the individual read depth distribution). The VCF was finally filtered for indels, multiallelic SNPs and sites containing more than 15% of missing genotypes ("--max-missing 0.85"). The final VCF file (hereafter referred to as the WGS dataset) contained ~5.9 M sites located on 9093 different scaffolds longer than 10 kb.

RAD sequencing data

RAD-seq libraries were prepared for 243 samples (**Supplementary Table S1**, RAD=="yes" and Type=="Tissue") by batches of 64 multiplexed individuals, following a similar protocol to Baird et al. (2008) using the SbfI restriction enzyme. Twenty-five of these samples were also used to produce WGS data, providing a link to understand the genetic structure in both datasets. RAD library sequencing was performed on an Illumina HiSeq2500 sequencer in 100 bp single-read mode. To complement our sampling, we also included raw sequencing data for 128 anchovy samples from Le Moan et al. (2016) which were generated using the same restriction enzyme. The data were demultiplexed using process_radtags (Stacks v2.60) and reads were aligned to our Eencr_V1 reference genome using BWA-MEM (BWA v0.7.17; Li, 2013). The reference-based Stacks pipeline was run in an integrated workflow developed by the MBB bioinformatic platform (https://web.mbb.cnrs.fr/subwaw/workflowmanager.php) (Penaud et al., 2020). Gstacks was run using default parameters ("--model marukilow --var-alpha 0.05 --gt-alpha 0.05 --max-clipped 0.2") with the minimum PHRED-scaled mapping quality set to 20 ("--min-mapq 20"). Thereafter, genotypes were exported in VCF format using the populations module ("--min-populations 2 --min-samples-per-pop 0.7 --min-maf 0.05") and filtered to remove SNPs with more than 15% missing data across individuals. The VCF was also filtered to only retain sites that were present in the WGS dataset, since our objective was to describe the same genetic variation but at a larger geographic scale. Lastly, all samples from the WGS dataset were integrated into the RAD VCF. The final VCF file (hereafter referred to as the RAD dataset) contained genotype data for 385 samples at 3880 variable sites.

Population structure

To describe the genetic structure in both WGS and RAD datasets, we conducted genome-wide and chromosome-wide Principal Component Analysis (PCA) using the R package SNPRelate (v1.28.0) (Zheng et al., 2012) and calculated individual heterozygosity per chromosome using VCFtools (v0.1.16) (Danecek et al., 2011). We used ADMIXTURE v1.3.0 (Alexander et al., 2009) to estimate individual ancestry proportions in all samples using the RAD dataset, assuming K=3 parental ancestries and using default parameters. Genetic differentiation (F_{ST}), nucleotide diversity (π) and absolute genetic divergence (d_{XY}) were calculated for the WGS data in non-overlapping 5 kb windows (with "--minSites 15") using the popgenWindows.py script (Martin, 2018; https://github.com/simonhmartin/genomics_general).

Identification and genotyping of structural variants

Our analyses of anchovy population structure revealed the presence of large structural variants (SVs) of several megabases segregating across the distribution range. We therefore used chromosome-wide PCA to identify clusters of individuals representing alternate genotypes at each SV, and assigned individuals' genotypes based on their cluster membership using both the WGS and RAD datasets. The PCA axis representing structural variation was in most cases assumed to be PC1, as it explains the highest amount of genotype variation. Samples which did not show clear cluster membership based on their PCA coordinates were not genotyped. For samples that had WGS data, we corroborated genotype assignment with the relative positions of each sample in local PCA, which was conducted in non-overlapping windows of 5 kb using lostruct (v0.0.0.9; Li & Ralph, 2019). In total, we genotyped individuals at 13 large SVs that occur on different chromosomes.

Divergence history of structural variants

To study the evolutionary relationships between clusters of individuals carrying different haplotypes at each of the 13 SVs, we constructed neighbour-joining trees for each chromosome carrying a large SV using the "phylo" command from VCF-Kit (Cook & Andersen, 2017). Because our chromosome-wide PCAs indicated low recombination between alternate SV haplotypes on the 13 chromosomes (**Supplementary Fig. S3 & S5**), these SV trees can be used to resolve the evolutionary relationships among haplotypes without the confounding effect of recombination. For these analyses, we used a subset of high coverage individuals that are homozygous for the SV considered to avoid phasing issues and to facilitate visualisation of haplotype relationships. We rooted trees on the branch separating alternate homozygote groups. For comparison, trees were also constructed for chromosomes not carrying SVs and rooted using a South African sample ("ATL_MAR_ZDA_61_1162").

Results

Genetic structure in the *E. encrasicolus* species complex - not two but three distinct ancestries

We used a combination of both WGS and RAD-seq data as complementary approaches to study the genetic structure of anchovies in the eastern Atlantic, Mediterranean and Black Seas. We used our RAD dataset with mean per-sample coverage of 52.7X to describe the overall genetic structure in 385 individuals from the entire range distribution. This analysis gives, for the first time, a clear picture of the genetic structure in the *E. encrasicolus* species complex. These results, obtained with a reduced representation SNP dataset, were very similar to those obtained using our WGS dataset (per-sample coverage 10-30X) which contains a smaller subset of samples (detailed results provided in the supplied HTML report, see **Supplementary Appendix**). Firstly, we observed genetic differentiation between samples collected in marine and coastal habitats in the northern part of the range, corresponding to the previously described marine and coastal ecotypes (Le Moan et al. 2016; Bonhomme et al. 2020). This can be observed along the second axis of variation (PCA 2 in **Fig. 1A**),

where most coastal samples are positioned at the top of the plot, while marine samples fall in the bottom right corner. As for PCA 1, this axis shows a different signal that reflects geographic structure rather than ecological structure. On the horizontal axis, South African samples and other individuals collected off the African Atlantic coast (Morocco and the Canary Islands) are spread out towards the left-hand side of the plot, while the majority of other samples group to the right. Hence, PCA at a genome-wide scale shows the existence of three distinct genetic ancestries, which were further confirmed using admixture analysis. Inferred individual ancestry was represented in a ternary plot (**Fig. 1B**) that shows the relative proportions of coastal (top), marine (right) and southern (left) ancestry for each individual. Ongoing gene flow between the three ancestries was revealed by considerable levels of admixture, in particular between the marine and coastal ancestries. A large number of samples also fell in the central area of the plot, reflecting relatively balanced proportions of the three ancestry components in these individuals.

Based on their ternary coordinates, samples were classified as belonging to one of seven ancestry categories, each corresponding to an area on the plot (black demarcations **Fig. 1B**, triangle in **Fig. 1C**). A sample was thus considered to belong to a given genetic cluster (coastal/*C*: green; southern/*S*: red; marine/*M*: blue) if that ancestry reached more than 80% of its total genetic ancestry. Secondly, we distinguished three different classes of admixed individuals where ancestry proportions were mainly made up of two ancestries (the third not amounting to more than 10%). These classes were *CS* (admixed between *C* and *S*; brown), *SM* (admixed between *S* and *M*; purple) and *MC* (admixed between *M* and *C*; seagreen). The last admixed class, *MCS*, consisted of individuals with balanced proportions of all three ancestries (admixed between *M*, *C* and *S*; grey). We summarised the genetic variation present at each sampling location by calculating the fraction of individuals in each ancestry category (**Fig. 1C**) in order to describe the eco-geographical distribution of the three ancestries. We found that individuals belonging to the *C* cluster (green) were only found in coastal habitats in the northern part of the range (diamond symbols), while *M* individuals (blue) mainly occurred in marine environments here (square symbols), thus corresponding to the two anchovy ecotypes previously described. This pattern was especially marked in the Mediterranean Sea, where almost all coastal samples were part of the *C* cluster (e.g. coastal habitats in *SIC*, *TNO* and *GDL*). However, this signal of ecotypic differentiation becomes diluted nearer to the Atlantic-Mediterranean boundary, where a gradient of increasing southern ancestry is observed. This admixture gradient can be seen through the increasing proportion of *MCS* individuals (grey) in the Alboran Sea (*ALB*), off the southern coast of Portugal (*PRS*) and in northern Morocco (MA4 and MA3). Finally, we observed that samples from locations to the south of the Canary islands (*CNR*), including the sampling site in South Africa (*ZDA*, inset map), all belonged to the *S* cluster (red).

After describing the three ancestries as well as their ecogeographic distribution patterns, we aimed to study their genomic architecture of differentiation. Genomic differentiation landscapes reconstructed between the coastal, marine and southern clusters using WGS data, yielded highly heterogeneous patterns that strongly varied from chromosome to chromosome (**Fig. 2**). The background level of differentiation between the marine and coastal clusters was lower (**Fig. 2A**) compared to the background F_{ST} between the southern and coastal clusters (**Fig. 2B**) and between the southern and marine clusters (**Fig. 2C**). Genomic landscapes of differentiation were generally similar whether marine and coastal individuals originated from the Atlantic (*ATL*: first row of each comparison) or from the Mediterranean Sea (*MED*: second row), even though some differences were observed (e.g. on chromosomes CM068262 and CM068273). On average, genetic

differentiation was lower between the coastal and southern clusters than between the marine and southern clusters (mean F_{ST} 0.02 lower in **Fig. 2B** than in **Fig. 2C** using *ATL* samples; mean F_{ST} 0.01 lower in **Fig. 2B** than in **Fig. 2C** using *MED* samples).

To investigate the genetic architecture potentially explaining heterogeneous F_{ST} values across the genome, we performed PCA on individual chromosomes using both the WGS and RAD datasets (**Supplementary Fig. S3 and S5**). Chromosome-wide PCAs revealed consistent patterns of clustering between the WGS and RAD datasets, but these varied widely from chromosome to chromosome, as did the amount of genetic variation explained by the first two PC axes. Representative examples of chromosome-wide PCAs shown in **Fig. 3A-H** illustrate the main diversity of patterns that were observed. While some chromosomes show largely continuous ancestry gradients (**Fig. 3A & E**), others show several discrete clusters where samples are closely grouped (**Fig. 3B-D & F-H**). The presence of tight PCA clusters indicates that a large number of SNPs are in strong LD, resulting in the segregation of a limited number of non-recombining haplotypes. Combined with elevated levels of heterozygosity in the intermediate clusters (**Fig. 3K**), these patterns suggest the presence of SVs. This was further supported by continuous clustering patterns observed across numerous consecutive local PCA windows (**Fig. 3L**) and continuous F_{ST} plateaus, providing evidence for the segregation of large SVs.

We assessed to what extent these SVs are associated with ecotype or lineage divergence by analysing the genomic landscape of differentiation,. Regions showing extremely high differentiation values (F_{ST} above the 95th quantile) were almost all clustered in the continuous F_{ST} plateaus suggesting that most divergent regions are associated with these putative SVs regions. We identified 13 chromosomes (indicated by asterisks in **Fig. 2**) that showed evidence for SVs spanning at least 2.5% of the windows on the chromosome. Using individual coordinates from chromosome-wide PCAs, these 13 SVs were successfully genotyped (**Supplementary Fig. S4 and S6**) and individuals were classified as either *00* homokaryotes (pink), *01* heterokaryotes (salmon) or *11* homokaryotes (gold) (**Fig. 3I-L**). Individuals that could not be confidently assigned to any given group were not genotyped (grey). For consistency, we always assigned the *00* genotype to the group containing the most southern samples, in order to polarise *0* haplotype with respect to southern lineage ancestry.

Anchovy lineages are differentiated at multiple SVs

Based on the SV genotypes that we assigned to samples, we studied the frequency patterns of the two haplotypes at each SV (*0* and *1*) as well as the genotype frequencies per genetic cluster (**Fig. 4**). We observed that the coastal, marine and southern clusters (background colours in **Fig. 4**) carry different sets of SV genotypes. The southern cluster largely harbours homokaryotic *00* genotypes at SVs (bottom row of pie charts), while marine individuals (top two rows) are mostly homokaryotic for the opposite haplotype (*11*). In coastal individuals, on the other hand, some SVs are nearly fixed for the *00* genotype, while others are fixed for the *11* genotype (see details below). Overall, we found that the three clusters show substantial frequency differences at multiple SVs, which is coherent with the F_{ST} plateaus that were observed in the differentiation landscapes (**Fig. 2**). Admixed individuals that carry more than one type of ancestry (e.g. *MCS*) were often heterokaryotes at SVs (**Supplementary Fig. S7**), consistent with their admixed status.

Even though SV frequency differences between the coastal, marine and southern clusters are observed, these SVs are not always fixed for a given haplotype but display a degree of haplotype sharing between the three clusters. This is clearly visible on CM068262 and CM068273, for example, where the SVs are polymorphic in almost all populations (**Fig 4**). Patterns of haplotype sharing can also be observed in neighbour-joining trees of these SV regions (**Fig. 5B & F**), where samples group according to their SV genotype (pink and gold circles) and not according to their genetic cluster membership (label colour). By looking at the general patterns of haplotype distributions across different SVs, it can be observed that southern ancestry haplotypes (*0*) are common in marine and/or coastal populations in the north. Furthermore, these southern haplotypes are slightly more common in the Atlantic (first and third rows in **Fig. 4**) than in the Mediterranean (second and fourth rows).

One of our main findings is that there is a substantial excess of haplotype sharing between the southern and coastal clusters, which is not observed between the southern and marine clusters. This is particularly evident on CM068255, CM068258 and CM068265, where *00* haplotypes are highly predominant (sometimes fixed) in the coastal samples. These three chromosomes are among the six showing high F_{ST} between the marine and coastal clusters (asterisks in **Fig. 2A**), which can thus be explained by the presence of southern haplotypes in coastal samples. This was confirmed in phylogenies of the SV regions (**Fig. 5C & G**), since coastal and southern samples are grouped in the same branch, while marine samples (with the alternative haplotype) are in a separate branch. On the other hand, high F_{ST} values between coastal and marine samples on CM068256, CM068270 and CM068271 are not due to the presence of southern haplotypes. On these chromosomes, northern populations almost exclusively carry *1* haplotypes (rows 1-4 in **Fig. 4**) that are, in addition to their differentiation from the *0* haplotype, also divergent between *C* and *M*. The presence of haplotype structure within the *1* haplogroup is indeed suggested by the separation of coastal and marine samples on PCA 2 (**Fig. 3D & H**), and also supported by the subdivision between different *1* haplotypes in the phylogenies (brown and gold circles in **Fig. 5D & H**). Overall, these patterns suggest that there are three distinct haplotypes segregating on CM068256, CM068270 and CM068271, potentially resulting from the presence of multiple SVs on each chromosome.

By studying the branch lengths separating different groups of haplotypes in phylogenies, we gathered information about divergence in different genomic regions. We aimed to evaluate their concordance with alternative divergence scenarios and observed different types of patterns that are illustrated by the four columns in Fig. 5 (for all phylogenies see **Supplementary Fig. S8** and for D_{xy} landscapes see **Supplementary Fig. S9**). We found that chromosomes that do not carry large SVs are characterised by short branch lengths and low levels of divergence (**Fig. 5A & E**,). Such collinear regions did not show pronounced genetic structure between the three clusters and supported highest genetic similarity between the coastal and marine clusters. These patterns contrasted with phylogenies reconstructed from SV regions, where long branches were found to separate samples carrying different genotypes (**Fig. 5B-D & F-H**). Similar to what was described before, we observed that certain SVs were divergent between ecotypes (**Fig. 5C-D & G-H**) while others rather showed patterns of haplotype sharing (**Fig. 5B & F**). We note that we were not able to compare divergence levels between different SVs, as SV block delimitation (based on F_{ST} and local PCA patterns) was probably not precise enough to exclude all recombinant regions from the analysis. It is however plausible that the divergence between coastal *1* haplotypes and marine *1*

haplotypes on CM068256, CM068270 and CM068271 is younger than the divergence between *0* and *1* haplotypes, as is reflected by topologies and branch lengths in phylogenies (**Fig. 5D & H**).

Discussion

We present the first genome-scale study investigating genetic structure in the *E. encrasicolus* species complex of anchovies from the eastern Atlantic, Mediterranean and Black Seas. Our results reveal that two previously described ecotype lineages (marine and coastal) genetically interact with a third southern lineage. This lineage has not previously been described in the northeast Atlantic but is related to the southern African anchovy *E. capensis*. The coastal, marine and southern lineages are almost exclusively differentiated by their genotype combinations at multiple large SVs, the remainder of the genome being only weakly differentiated. We further found that the SVs involved in ecotype divergence have two distinct origins. While three of the SVs have likely established in association with habitat type in the northern hemisphere, three others found in the coastal lineage most likely originate from the southern lineage. We here consider different evolutionary scenarios that could explain these patterns.

Previous genetic studies have shown that the European anchovy is subdivided into marine and coastal ecotypes that are present from the Bay of Biscay, through the Mediterranean to the Black Sea. Here, we show that there is a third component of genetic structure in this species, corresponding to an Atlantic lineage occurring off the African coastline. This southern lineage shows genetic homogeneity at a very large spatial scale, with genetic similarity between individuals sampled in Morocco, the Canary Islands and even as far as South Africa. However, from northern Morocco and southern Portugal into the Alboran Sea, we observe genetic admixture forming a gradient of decreasing southern ancestry. Zarraonaindia et al. (2012) also reported the presence of strong genetic structure in this region, but interpreted this signal as being due to the presence of differentiated populations inhabiting narrow-shelf waters associated with upwelling. We instead propose that this region corresponds to a three-way contact zone between the southern lineage and the two northern (marine and coastal) lineages. Here, we observe post-F1 introgressive hybridisation resulting in widespread admixture and gene flow between the southern, coastal and marine clusters, as is reflected by gradual ancestry gradients in the PCA plot (Fig. 1). Almost all individuals that were identified as belonging to the three-way admixed class (MCS) were sampled in the contact zone between the three lineages (Fig. 1C). This pattern of three-way admixture also extends further north in the Atlantic, where a few individuals carrying a MCS background were detected. This is further consistent with the presence of admixed genotypes in the Bay of Biscay as reported by Zarranoianda et al. (2012). The existence of gene flow between coastal and marine ecotypes has already been illustrated in previous work (Le Moan et al., 2016), but our results reveal that admixture with the southern lineage also contributes to global diversity patterns. Previous studies reporting various signals of spatial structure in this species may have unknowingly captured different aspects of these complex admixed ancestries, leading to many different and conflicting interpretations in the literature.

We found evidence for multiple megabase-scale SVs that segregate in the marine, coastal and southern anchovy lineages. Although the presence of SVs was only supported through indirect evidence based on LD, divergence and heterozygosity patterns, these regions showed many of the signals typically associated with chromosomal inversions (Mérot et al., 2020). In certain regions, we

further found evidence for the presence of three distinct haplotypes, which could potentially be explained by multiple rearrangements occurring on the same chromosome. In the literature, it has often been reported that SVs play a role in differentiating evolutionary lineages within species complexes, suggesting that they can play an important role in the formation of cryptic lineages or ecotypes (reviewed in Zhang et al., 2021). This is probably because co-adapted and locally advantageous alleles are protected from recombination in rearranged portions of the genome, allowing RI barriers to be protected from gene flow and favouring their accumulation (Navarro & Barton, 2003; Noor et al., 2001). The results that we obtained in anchovies seem to point in this direction, since markers differentiating lineages and ecotypes were largely concentrated in SVs. We found that the SVs harbour divergent haplotypes, suggesting that recombination between lineages has been suppressed in these regions for a significant amount of time. In contrast, collinear regions of the genome showed low differentiation levels which likely reflects the homogenising effect of gene flow and recombination, as is supported by the detection of numerous admixed genotypes.

Ecotype differentiation between marine and coastal anchovies was shown to involve multiple SVs that cover roughly 25% of the genome. This estimate is in line with a previous study that estimated barriers to gene flow between marine and coastal ecotypes to span 20 to 25% of the genome (Le Moan et al., 2016). By reconstructing the genomic landscape of ecotype divergence, we determined that the two ecotypes were differentiated at six large SVs. We do note however that the exact size and fine-scale structure of the SVs will need to be confirmed by long-range sequencing in order to directly identify breakpoints and to resolve the possibly complex variation at each SV. In addition, there is also the possibility that smaller genomic islands of differentiation, located in the collinear genome, could also play an important role in ecotype divergence and that not all of them may have been detected in our study.

We found that the coastal ecotype carries the same haplotype as the southern lineage at a minimum of three SVs. If these SVs were already segregating in the population that was ancestral to the marine, coastal and southern lineages, we could have expected similar patterns due to incomplete lineage sorting (ILS). However, we do not observe any SVs that show the opposite pattern where the same haplotype fixed in the marine and southern lineages, an outcome that could have been expected under ILS. This suggests that a different explanation could be necessary and raises the question of whether the coastal and southern lineages have a partly shared evolutionary history. Overall, the patterns of population structure in the collinear genome as well as haplotype distributions at SVs (whether shared between lineages or private) suggest that there are two possible scenarios (**Fig. 5M & N**):

1) In the first scenario, the deepest split represents the onset of divergence between the southern lineage and the branch giving rise to the two northern lineages (marine and coastal ecotypes) (**Fig. 5M**). After this initial split, genomic rearrangements could have taken place in either branch, resulting in the segregation of distinct haplotypes that accumulated divergence over time. Subsequent gene flow between the southern and coastal lineages could then have caused the introgression of SVs, explaining the patterns of haplotype sharing between these lineages. In this case, the introgression of SVs probably took place during an earlier period of contact compared to the contemporary period of gene flow, since current admixture in the contact zone appears insufficient to explain how coastal samples from across the distribution range almost exclusively carry *0* haplotypes at these SVs. Evidence for the

introgression of SVs is supported in other systems (Jay et al., 2018), and may serve as a source of divergent haplotypes that in some cases lead to speciation. In anchovies, for example, this influx of divergent haplotypes might have been the initial trigger for ecotype speciation, especially if the SVs contained genetic incompatibilities that evolved in allopatry. SVs introgressed from the southern lineage could alternatively have conferred a selective advantage in coastal habitats, although this is not strongly supported since the southern lineage is not confined to such areas.

2) The second scenario that could underlie differentiation patterns among the three anchovies lineages, considers that the coastal lineage shares its most recent common ancestor with the southern lineage and not with the marine lineage (**Fig. 5N**). This could explain how coastal anchovies carry southern haplotypes at SVs (through recent common ancestry), without invoking the introgression of multiple SVs and their successful passing of various selective filters. Instead, this scenario suggests that when the proto-coastal lineage entered into secondary contact with the marine lineage, haplotypes at certain SVs remained intact while divergence in the background genome (collinear regions and other SVs such as on CM068268) was eroded through gene flow.

Demographic analysis has indeed shown that ecotype divergence in the European anchovy is most likely explained by secondary contact following a long period of allopatric isolation (Le Moan et al., 2016). Furthermore, two deeply divergent mitochondrial haplotypes have been described and similarly proposed to result from secondary contact (Magoulas et al., 1996). Although the aforementioned studies did not consider the existence of a third anchovy lineage, the period of allopatric divergence that was inferred may in fact correspond to the split between the branches leading to the marine and the southern lineages (**Fig. 5M & N**). The divergence between these ancient lineages may subsequently have been partially eroded during periods of gene flow, particularly through the emergence of the coastal lineage which shows a mosaic of genetic ancestry from both evolutionary branches. In addition to the “hybrid” status of the coastal lineage, the recent re-mixing of these divergent ancestries is reflected by patterns of shared polymorphism at SVs and mitochondrial haplotypes across the three lineages. For instance, mitochondrial clade A is predominant (but not fixed) off the Moroccan coast, while clades A and B are both common in the marine and coastal lineages (Chahdi Ouazzani, 2016; Oueslati et al., 2014; Silva et al., 2017), supporting that these haplotypes have been shared through gene flow and showing similar patterns to some of the SVs that we described (e.g. on CM068260). Overall, by describing these distinct evolutionary branches and associating them with observed diversity, our results are thus in line with earlier findings suggesting that the patterns observed in anchovies are a result of one or more secondary contacts.

Although this study presents the most complete study to date of genome-scale variation in anchovies of the *E. encrasicolus* species complex, results do not allow to determine which of the two alternative evolutionary scenarios is the most likely. This would probably require more in-depth analyses and complementary information regarding the past geographic distributions of the different lineages, especially during the last glacial period. However, we may hypothesise that the two evolutionary branches diverged while being isolated from each other in the northern and southern hemispheres, as it broadly seems to be the case today. Periods of secondary contact then could have taken place during cooler periods when transequatorial dispersals were possible and when the

lineages were confined to lower latitudes due to ice sheets near the poles. During such an episode, anchovies from the southern hemisphere could have moved north and come into contact with northern populations to participate in giving rise to the coastal lineage. In more recent times, such dispersal events could have established the population of southern anchovies off the Moroccan coast, although other studies rather suggest that the South African population was established through southward migration (Bouchenak-Khelladi et al., 2008; Grant & Bowen, 2006). It would be particularly interesting to further investigate whether the anchovy lineages were already associated with particular habitats or how/when this association took place to result in a coastal and marine lineage. Our study brings an interesting perspective for understanding the formation of ecotypes, since our results suggest that secondary contact between geographically isolated lineages favoured the emergence of the ecotypes. This joins a body of other studies that have proposed that ecotype formation may involve phases of allopatric divergence and other historical contingencies, and that it is rarely based on strict *in situ* adaptation alone (Belleghem et al., 2018; Bernatchez et al., 1996; Bierne et al., 2011; Hendry, 2009; Le Moan et al., 2021; Rundell & Price, 2009). Instead, RI barriers between ecotypes may be a result of genetic incompatibilities that evolved in allopatry, or may be due to alleles that are under extrinsic selection but which evolved in different contexts altogether.

One salient result of our study is that SVs play an important role in differentiating marine and coastal anchovy ecotypes, and that these SVs which provide a barrier effect between ecotypes have different origins. While three of the SVs (CM068255, CM068258 and CM068265) find their likely origin in the southern lineage, three others (CM068256, CM068270 and CM068271) appear to have evolved *in situ* in the northern part of the range. As for the SVs with a southern origin, distinguishing between the two proposed evolutionary scenarios (although they are not necessarily mutually exclusive) could be interesting for understanding the exact mechanisms underlying the role of SVs in ecotype speciation. For example, the first scenario (**Fig. 5M**) seems to suggest that the introgression of SVs was associated with a form of selective advantage, while the second scenario (**Fig. 5N**) highlights that SVs can play a role in preserving past divergence.

Regardless of their source of origin, the six SVs differentiating anchovy ecotypes cover a large portion of the genome and contain many hundreds of genes that could function as different RI barriers. It seems likely that coupling between these various barriers and between the SVs could have provided a sufficiently strong barrier to counter gene flow between the ecotypes. For coupling to have taken place, we expect that there was a progressive build-up of associations between different SVs as the ecotypes continued to diverge. We suggest that the three SVs of southern origin could represent the oldest RI barriers, and that these subsequently became associated with younger barriers specific to the northern lineages, further accentuating RI between the ecotypes. The three younger SVs that evolved *in situ* in the north may contain alleles that play a role in local adaptation or the reinforcement of RI (e.g. immigrant inviability and other forms of prezygotic isolation). The European anchovy could thus present an interesting case of ecotype speciation that results from the coupling of barriers of different origins, which partly evolved in allopatry (SVs from the southern lineage) and partly in sympatry/parapatry (depending on the extent to which gene flow was already reduced). Future directions of study should aim to confirm the origin and the fine-scale structure of the SVs, as well as determine the specific selective forces maintaining this variation. It would be pertinent to evaluate the degree of coupling between SVs (e.g. by conducting marine-coastal transects) and to determine whether we can expect coupling to get stronger in the future (i.e. towards the completion of speciation). Future studies should also assess the degree and the nature

of RI between ecotypes as well as RI with the southern lineage, in order to consider implications for management.

Acknowledgements

The data were partly produced and analysed with the support of the GenSeq genotyping and sequencing platform. This project benefited from the Montpellier Bioinformatics Biodiversity platform (MBB) supported by the LabEx CeMEB, an ANR "Investissements d'avenir" program (ANR-10-LABX-04-01). We thank the Montpellier GenomiX platform for constructing the reference genome's 10× Chromium library. We would like to thank all of our colleagues who aided with sampling.

Figures

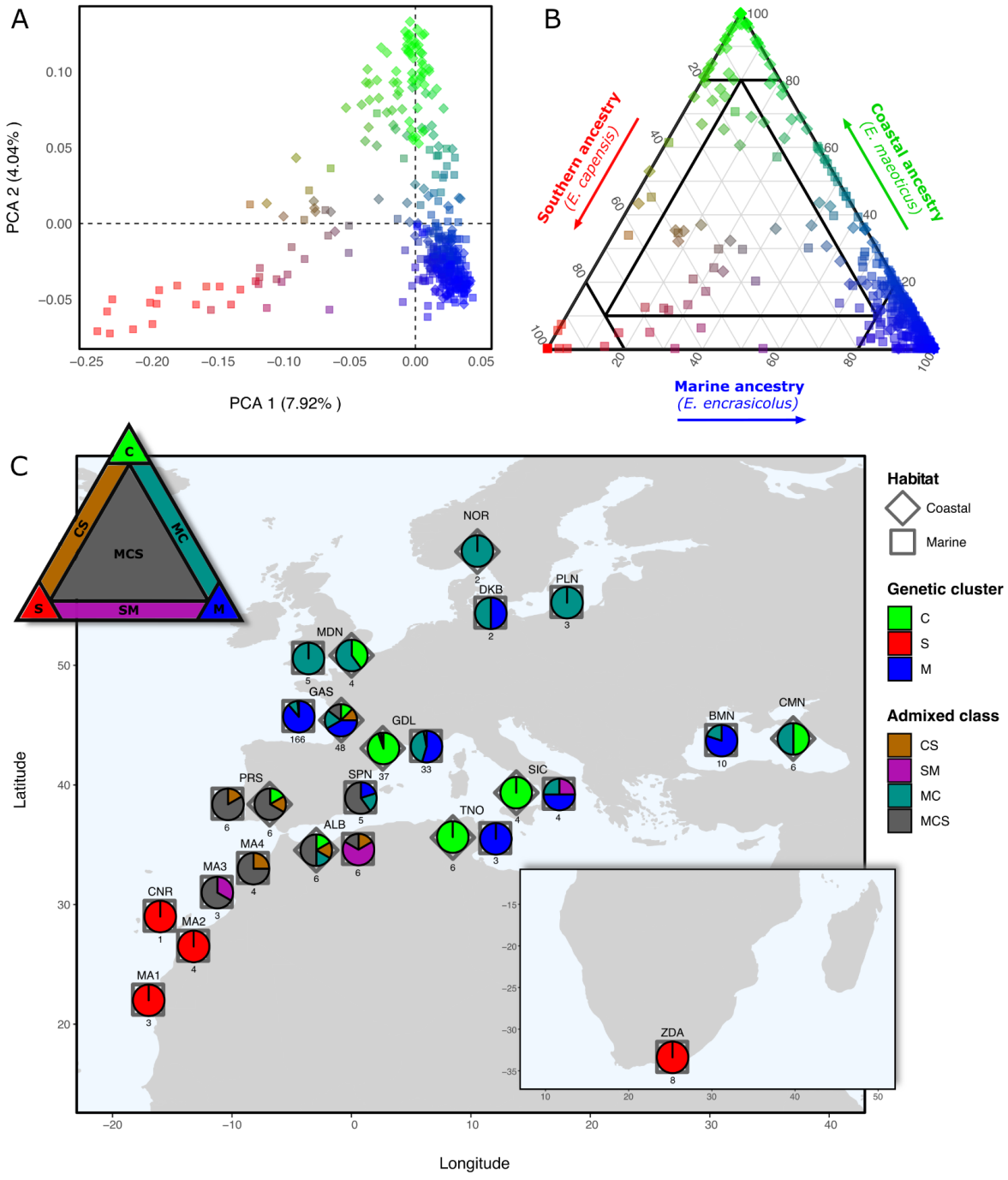

Fig. 1. A) Principal Component Analysis (PCA) performed on the entire dataset of 385 anchovy samples. Sites used in the analysis were high-quality variants present both in the WGS data as well as in the RAD data, corresponding to a total of 3881 SNPs. Shapes indicate habitat type and colours reflect ancestry proportions as determined by admixture analysis (see B). B) Ternary plot showing the admixture level between three genetic ancestries: coastal (green), southern (red), and marine (blue) ancestry. Coordinates, as well as RGB colours, reflect the relative ancestry proportions of samples along each of the three axes. Samples were classified as belonging to a genetic cluster (black

lines demarcating seven areas) based on their position in the plot. Clusters *C*, *S* and *M* represent "non-admixed" parental lineage ancestries, while *CS*, *SM*, *MC* and *MCS* represent various levels of admixture. B) Map with sampling locations where symbols represent habitat type and pie charts show the proportions of different genetic clusters present. Numbers beneath pie charts indicate sample sizes. Locations are described in **Supplementary Table S$**.

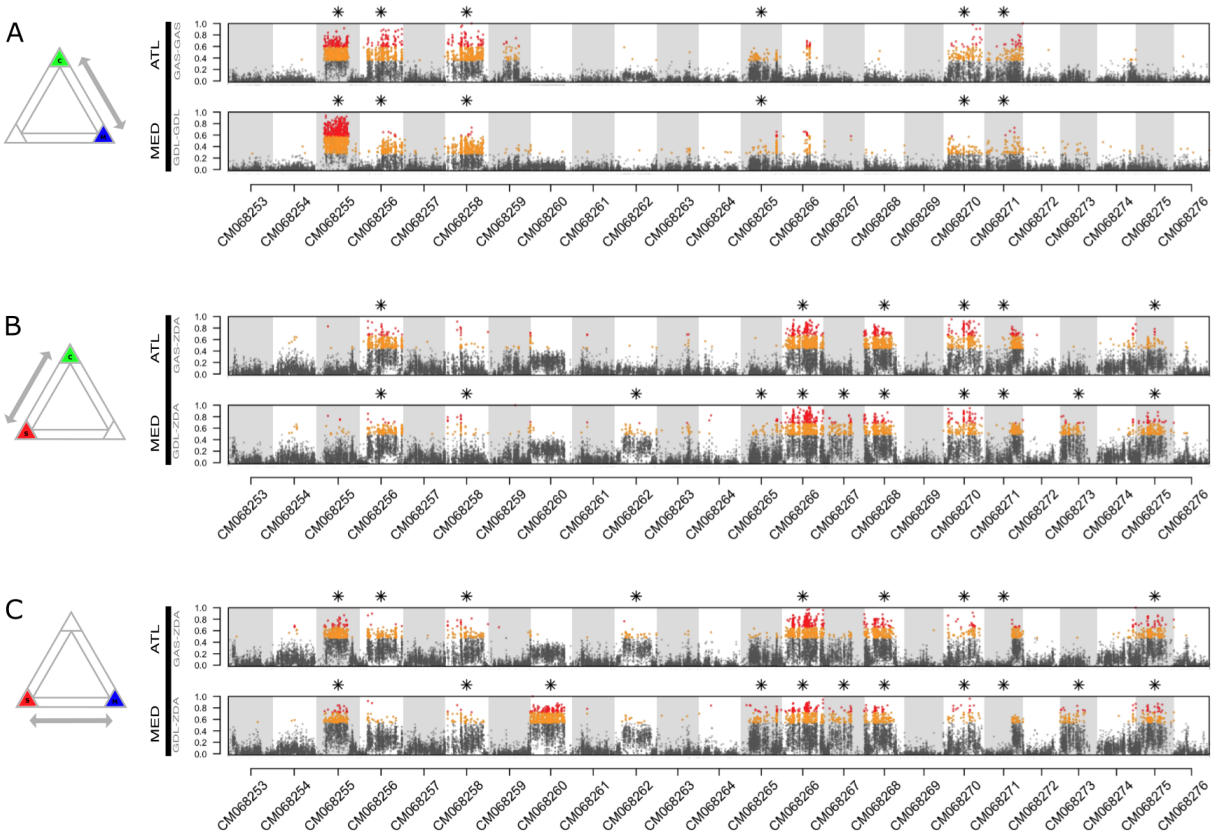

Fig. 2. Genomic landscapes of differentiation (F_{ST}) calculated in 5 kb sliding windows between groups of samples (3 individuals per group) from different genetic clusters (see Fig. 1). Differentiation landscapes are shown for three different comparisons (A: Coastal vs. Marine; B: Coastal vs. southern; C: southern vs. Marine). Each panel consists of two rows, representing cases where Coastal/Marine samples either originated from the Atlantic (ATL) or from the Mediterranean Sea (MED). Orange points are windows where F_{ST} was higher than the 95th quantile, while red points are above the 99th quantile. Stars indicate chromosomes where more than 2.5% of windows showed F_{ST} higher than the 95th quantile. Grey and white rectangles delimit the 24 chromosomes of *E. encrasicolus*.

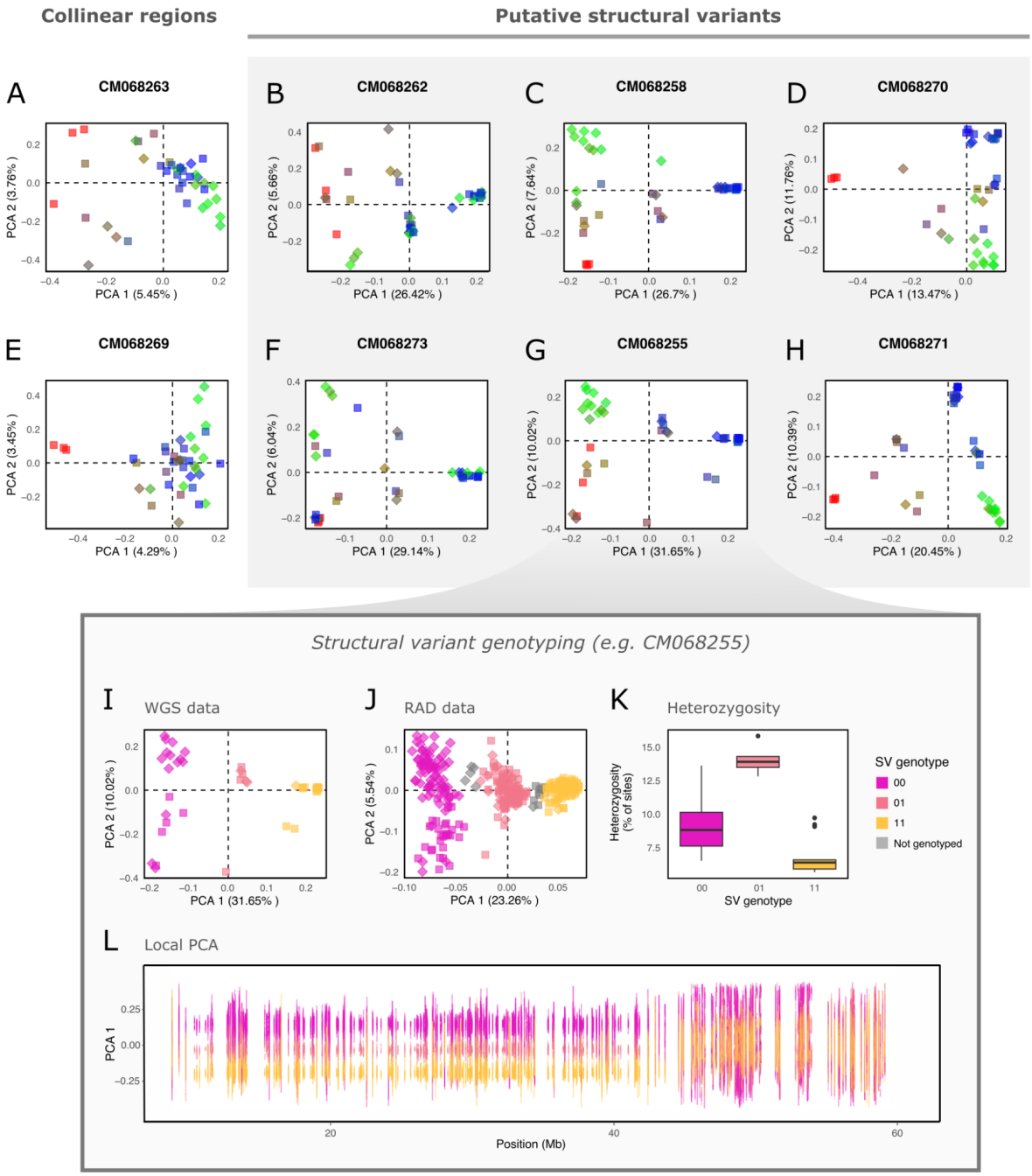

Fig. 3. Examples of relatedness patterns on different chromosomes. PCA conducted on whole-genome data (n=39) at a chromosome-wide level (A-H) reveals that certain chromosomes show signs of reduced recombination through the presence of tight PCA clusters (B-D, F-H). These chromosomes show evidence for the presence of structural variants (SVs), which we aimed to genotype in all individuals (I-L). SV genotypes were assigned for whole-genome and RAD data (n=385) based on PCA coordinates (I & J). For whole-genome data, these assignments were further corroborated by individual heterozygosity levels (K) and local PCA patterns (L). In (L), PCA 1 coordinates plotted for non-overlapping 5 kb windows along the chromosome show sustained clustering patterns in the SV region (ending around ~40 Mb). Gaps represent regions on CM068255 for which we did not have data (scaffolds < 10 kb or not present in our reference genome). Colours either reflect genome-wide ancestry proportions (A-H) or the SV genotype that was assigned to each individual (I-L).

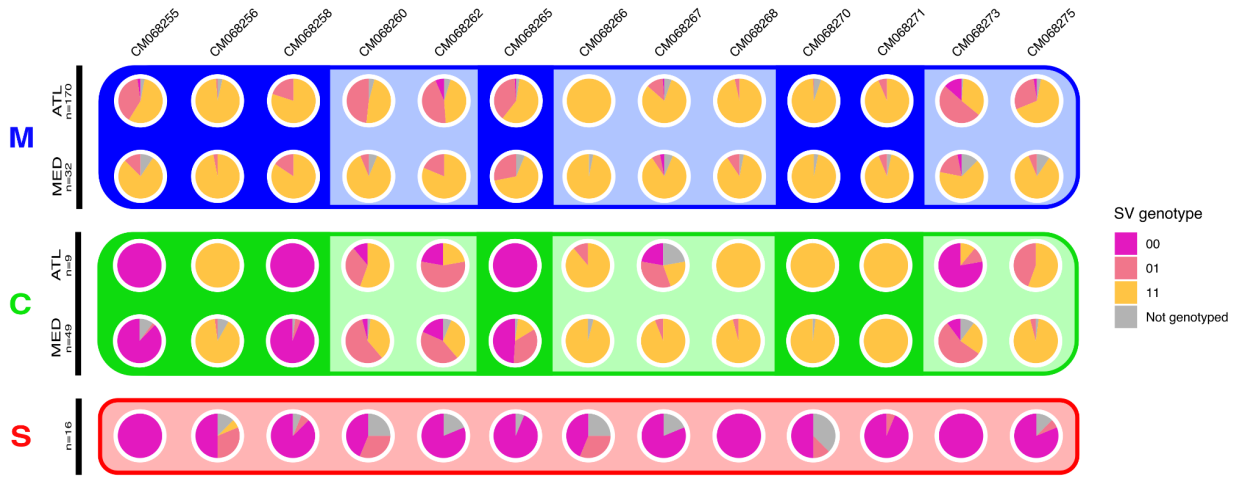

Fig 4. Genotype frequencies for 12 SVs on different chromosomes (columns). Pie charts show frequencies for the *00* (pink), *01* (salmon) and *11* (gold) genotypes in the Marine cluster (blue background), Coastal cluster (green background) and southern cluster (red background), with upper and lower rows corresponding to Atlantic and Mediterranean samples respectively (for *M* and *C*). Darker background colour in *M* and *C* indicates chromosomes that show elevated F_{ST} when comparing Marine and Coastal individuals (Fig. 2A). For three of these chromosomes (CM068255, CM068258 and CM068265), ecotype differentiation involves southern haplotypes (*00*) that are present at high frequency in the Coastal individuals, while this is not the case for CM068270 and CM068271.

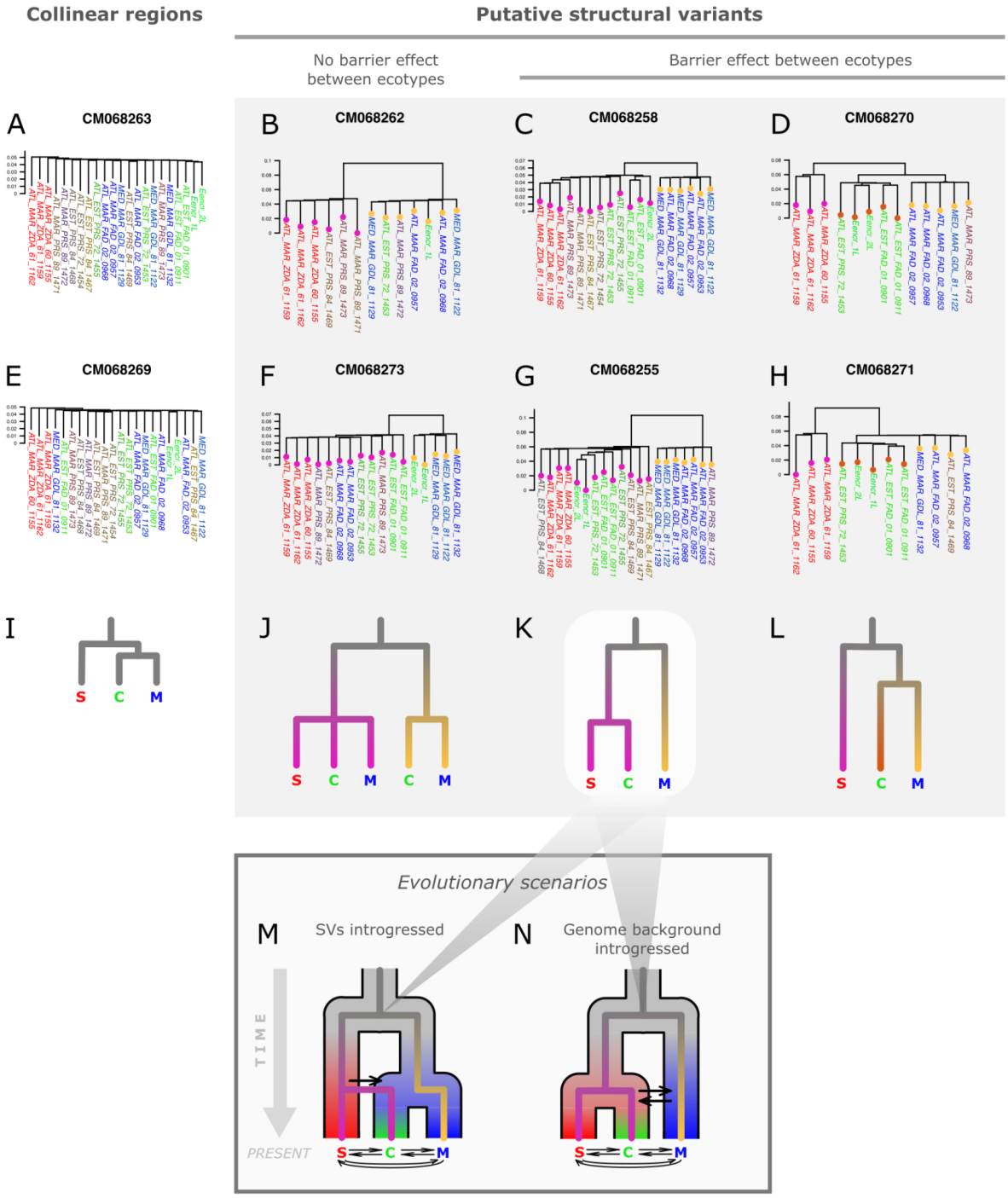

Fig. 5. A-H) Neighbour-joining trees showing interindividual relationships on the same subset of chromosomes as in Fig. 3. Short branch lengths for normally recombining chromosomes (A & E) contrast with long branches between divergent haplotypes at SVs (B-D, F-H). These SVs show varying patterns of being shared or being private to the coastal, marine or southern clusters. This is what we illustrate with schematic trees in (I-L), where pink or gold colouration reflects divergence between haplotypes. In (L), brown colouration indicates the divergence of a third haplotype differentiating coastal and marine samples (e.g. CM068270 and CM068271). In (K), we propose two evolutionary scenarios (M & N) which could explain the observed patterns of haplotype distributions across SVs. For (A & E), trees were constructed using SNPs on the entire chromosome, whereas in (B-D, F-H), the region was limited

to that covered by the SV. For trees of SV regions, intermediate samples that present heterokaryotes are not displayed. Leaf labels are coloured according to individual ancestry proportions and tip symbols (circles) correspond to SV genotype. Trees were plotted on the same vertical scale.

References

Belleghem, S. M. V., Vangestel, C., Wolf, K. D., Corte, Z. D., Möst, M., Rastas, P., Meester, L. D., & Hendrickx, F. (2018). Evolution at two time frames : Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLOS Genetics*, *14*(11), e1007796. https://doi.org/10.1371/journal.pgen.1007796

Benjamin Penaud, Rémy Dernat, Iago Bonnici, & Khalid Belkhir. (2020). *MBB-Framework : A set of tools to build and run reproducible and portable workflows* [Logiciel]. https://web.mbb.cnrs.fr/subwaw/workflowmanager.php

Bernatchez, L., Vuorinen, J. A., Bodaly, R. A., & Dodson, J. J. (1996). Genetic Evidence for Reproductive Isolation and Multiple Origins of Sympatric Trophic Ecotypes of Whitefish (Coregonus). *Evolution*, *50*(2), 624-635. https://doi.org/10.2307/2410836

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis : Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*(10), 2044-2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Bolnick, D. I., Hund, A. K., Nosil, P., Peng, F., Ravinet, M., Stankowski, S., Subramanian, S., Wolf, J. B. W., & Yukilevich, R. (2023). A multivariate view of the speciation continuum. *Evolution*, *77*(1), 318-328. https://doi.org/10.1093/evolut/qpac004

Bonhomme, F., Meyer, L., Arbiol, C., BĂnaru, D., Bahri-Sfar, L., Fadhlaoui-Zid, K., Strelkov, P., Arculeo, M., Soulier, L., Jean-pierre, Q., & Gagnaire, P. (2021). Systematics of European coastal anchovies (genus Engraulis Cuvier). *Journal of Fish Biology*, *100*. https://doi.org/10.1111/jfb.14964

Bouchenak-Khelladi, Y., Durand, J.-D., Magoulas, A., & Borsa, P. (2008). Geographic structure of European anchovy : A nuclear-DNA study. *Journal of Sea Research*, *59*. https://doi.org/10.1016/j.seares.2008.03.001

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric : The most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 2997-3007. https://doi.org/10.1098/rstb.2008.0076

Butlin, R. K., & Smadja, C. M. (2018). Coupling, Reinforcement, and Speciation. *The American Naturalist*, *191*(2), 155-172. https://doi.org/10.1086/695136

Campbell, M. A., Anderson, E. C., Garza, J. C., & Pearse, D. E. (2021). Polygenic Basis and the Role of Genome Duplication in Adaptation to Similar Selective Environments. *Journal of Heredity*, *112*(7), 614-625. https://doi.org/10.1093/jhered/esab049

Chahdi Ouazzani, K. (2016). Genetic population structure of the European anchovy (Engraulis encrasicolus) based on mitochondrial DNA sequences along the Moroccan coast. *AACL Bioflux*, *9*.

Cook, D. E., & Andersen, E. C. (2017). VCF-kit : Assorted utilities for the variant call format. *Bioinformatics*, *33*(10), 1581-1582. https://doi.org/10.1093/bioinformatics/btx011

Della Torre, A., Merzagora, L., Powell, J. R., & Coluzzi, M. (1997). Selective introgression of paracentric inversions between two sibling species of the Anopheles gambiae complex. *Genetics*, *146*(1), 239-244. https://doi.org/10.1093/genetics/146.1.239

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., García-Accinelli, G., Van Belleghem, S. M., Patterson, N., Neafsey, D. E., Challis, R., Kumar, S., Moreira, G. R. P., Salazar, C., Chouteau, M., Counterman, B. A., Papa, R., Blaxter, M., Reed, R. D., … Mallet, J. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science*, *366*(6465), 594-599. https://doi.org/10.1126/science.aaw2090

Gould, B. A., Chen, Y., & Lowry, D. B. (2017). Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive differentiation and reproductive isolation. *Molecular Ecology*, *26*(1), 163-177. https://doi.org/10.1111/mec.13881

Grant, W. S., & Bowen, B. W. (2006). Living in a tilted world : Climate change and geography limit speciation in Old World anchovies (Engraulis; Engraulidae). *Biological Journal of the Linnean Society*, *88*(4), 673-689. https://doi.org/10.1111/j.1095-8312.2006.00651.x

Hager, E. R., Harringmeyer, O. S., Wooldridge, T. B., Theingi, S., Gable, J. T., McFadden, S., Neugeboren, B., Turner, K. M., Jensen, J. D., & Hoekstra, H. E. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science*, *377*(6604), 399-405. https://doi.org/10.1126/science.abg0718

Hendry, A. P. (2009). Ecological speciation ! Or the lack thereof?This Perspective is based on the author's J.C. Stevenson Memorial Lecture delivered at the Canadian Conference for Fisheries Research in Halifax, Nova Scotia, January 2008. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(8), 1383-1398. https://doi.org/10.1139/F09-074

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the Impact of Inversions in Evolution : From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual review of ecology, evolution, and systematics*, *39*, 21-42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Jay, P., Leroy, M., Le Poul, Y., Whibley, A., Arias, M., Chouteau, M., & Joron, M. (2022). Association mapping of colour variation in a butterfly provides evidence that a supergene locks together a cluster of adaptive loci. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1856), 20210193. https://doi.org/10.1098/rstb.2021.0193

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. Á., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Current Biology*, *28*(11), 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072

Johannesson, K., Faria, R., Moan, A. L., Rafajlović, M., Westram, A. M., Butlin, R. K., & Stankowski, S. (2024). Diverse pathways to speciation revealed by marine snails. *Trends in Genetics*, *40*(4), 337-351. https://doi.org/10.1016/j.tig.2024.01.002

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*(1), 419-434. https://doi.org/10.1534/genetics.105.047985

Kulmuni, J., Butlin, R. K., Lucek, K., Savolainen, V., & Westram, A. M. (2020). Towards the completion of speciation : The evolution of reproductive isolation beyond the first barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1806), 20190528. https://doi.org/10.1098/rstb.2019.0528

Le Moan, A., Bekkevold, D., & Hemmer-Hansen, J. (2021). Evolution at two time frames : Ancient structural variants involved in post-glacial divergence of the European plaice (Pleuronectes platessa). *Heredity*, *126*(4), Article 4. https://doi.org/10.1038/s41437-020-00389-3

Le Moan, A., Stankowski, S., Rafajlović, M., Ortega-Martinez, O., Faria, R., Butlin, R. K., & Johannesson, K. (2024). Coupling of twelve putative chromosomal inversions maintains a strong barrier to gene flow between snail ecotypes. *Evolution Letters*, qrae014. https://doi.org/10.1093/evlett/qrae014

Li, H., & Ralph, P. (2019). Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics*, *211*(1), 289-304. https://doi.org/10.1534/genetics.118.301747

Lundberg, M., Mackintosh, A., Petri, A., & Bensch, S. (2023). Inversions maintain differences between migratory phenotypes of a songbird. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-36167-y

Magoulas, A., Tsimenides, N., & Zouros, E. (1996). Mitochondrial DNA phylogeny and the reconstruction of the population history of a species : The case of the European anchovy (Engraulis encrasicolus). *Molecular Biology and Evolution*, *13*(1), 178-190. https://doi.org/10.1093/oxfordjournals.molbev.a025554

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends in Ecology & Evolution*, *35*(7), 561-572. https://doi.org/10.1016/j.tree.2020.03.002

Moan, A. L., Gagnaire, P.-A., & Bonhomme, F. (2016). Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, *25*(13), 3187-3202. https://doi.org/10.1111/mec.13627

Navarro, A., & Barton, N. H. (2003). Accumulating postzygotic isolation genes in parapatry : A new twist on chromosomal speciation. *Evolution; International Journal of Organic Evolution*, *57*(3), 447-459. https://doi.org/10.1111/j.0014-3820.2003.tb01537.x

Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, *98*(21), 12084-12088. https://doi.org/10.1073/pnas.221274498

Oueslati, S., Fadhlaoui-Zid, K., Kada, O., Augé, M. T., Quignard, J. P., & Bonhomme, F. (2014). Existence of two widespread semi-isolated genetic entities within Mediterranean anchovies. *Marine Biology*, *161*(5), 1063-1071. https://doi.org/10.1007/s00227-014-2399-5

Rafajlović, M., Rambla, J., Feder, J. L., Navarro, A., & Faria, R. (2021). Inversions and genomic differentiation after secondary contact : When drift contributes to maintenance, not loss, of differentiation. *Evolution*, *75*(6), 1288-1303. https://doi.org/10.1111/evo.14223

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, *16*(7), 351-358. https://doi.org/10.1016/s0169-5347(01)02187-5

Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, *24*(7), 394-399. https://doi.org/10.1016/j.tree.2009.02.007

Silva, G., Cunha, R. L., Ramos, A., & Castilho, R. (2017). Wandering behaviour prevents inter and intra oceanic speciation in a coastal pelagic fish. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-02945-0

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research*, *27*(5), 757-767. https://doi.org/10.1101/gr.214874.116

Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M. A., Manzano, C., Grant, W. S., Irigoien, X., & Estonba, A. (2012). Multiple SNP Markers Reveal Fine-Scale Population and Deep Phylogeographic Structure in European Anchovy (Engraulis encrasicolus L.). *PLOS ONE*, *7*(7), e42201. https://doi.org/10.1371/journal.pone.0042201

Zhang, L., Reifová, R., Halenková, Z., & Gompert, Z. (2021). How Important Are Structural Variants for Speciation? *Genes*, *12*(7), 1084. https://doi.org/10.3390/genes12071084