The Fate of a Polygenic Phenotype Within the Genomic Landscapes of Introgression in the European Seabass Hybrid Zone

Maeva Leitwein [ID],[1,*] Ghislain Durif [ID],[2] Emilie Delpuech [ID],[1] Pierre-Alexandre Gagnaire [ID],[3] Bruno Ernande [ID],[1] Marc Vandeputte [ID],[1] Alain Vergnet [ID],[1] Maud Duranton [ID],[1] Frederic Clota [ID],[1] François Allal [ID][1]

[1]UMR Marbec, Université Montpellier, CNRS, Ifremer, IRD, INRAE, 34000 Montpellier, France
[2]IMAG—Institut Montpelliérain Alexander Grothendieck, 34000 Montpellier, France
[3]ISEM, Univ Montpellier, CNRS, EPHE, IRD, 34000 Montpellier, France

*Corresponding author: E-mail: maeva.leitwein.pro@gmail.com.
Associate editor: Mary O'Connell

Abstract

Unraveling the evolutionary mechanisms and consequences of hybridization is a major concern in biology. Many studies have documented the interplay between recombination and selection in modulating the genomic landscape of introgression, but few have considered how associations with phenotype may affect this landscape. Here, we use the European seabass (*Dicentrarchus labrax*), a key species in marine aquaculture that undergoes natural hybridization, to determine how selection on phenotype modulates the introgression landscape between Atlantic and Mediterranean lineages. We use a high-density single nucleotide polymorphism array to assess individual local ancestry along the genome and improve the mapping of muscle fat content, a polygenic trait that is divergent between lineages. Taking into account variation in recombination rates, we reveal a purging of Atlantic ancestry in the admixed Mediterranean populations. While Atlantic individuals had higher muscle fat content, we observed that genomic regions associated with this trait in Mediterranean populations displayed reduced introgression of Atlantic ancestry. These results emphasize how selection against maladapted alleles shapes the genomic landscape of introgression.

Key words: hybridization, introgression, phenotype, European seabass, selection, recombination.

Introduction

Hybridization is widespread and common in nature (Baack and Rieseberg 2007; Schumer et al. 2015; Payseur and Rieseberg 2016; Moran et al. 2021; Adavoudi and Pilot 2022), understanding the genomic consequences of hybridization between divergent populations is of primary interest in evolutionary biology (Baack and Rieseberg 2007; Abbott et al. 2016; Payseur and Rieseberg 2016; Moran et al. 2021). Hybridization induces gene flow, resulting in a mosaic of introgressed migrant tracts within the receiving population (Anderson and Stebbins 1954; Leitwein et al. 2019b). This mosaic of hybridization may result in phenotypic variation through new combinations of phenotypic traits (Lewontin and Birch 1966; Tobler and Carson 2010). While hybridization could be globally neutral, the genome-wide pattern of introgression can also be shaped by both natural selection and recombination rates (Martin and Jiggins 2017; Schumer et al. 2018; Moran et al. 2021). Recombination shortens migrant tracts at each generation and positive or negative selection may act on phenotypic traits carried on those tracts

(Martin and Jiggins 2017; Schumer et al. 2018; Leitwein et al. 2019b). Many studies have investigated the pattern of introgression and the associated evolutionary mechanisms (Abbott et al. 2016; Martin and Jiggins 2017; Schumer et al. 2018; Moran et al. 2021); however, only a few have linked this to the direct effects of introgression on phenotypes (McArthur et al. 2021; Reilly et al. 2022).

To understand how introgression may affect a quantitative trait, it is important to understand how selection and recombination interact to shape the mosaic of introgressed tracts. While recombination shortens migrant tracts at each generation (Schumer et al. 2018), it also varies along the genome with introgressed tracts shortening faster in highly recombining regions (Martin and Jiggins 2017; Leitwein et al. 2019b). Consequently, in cases where hybridization is not neutral, selection may act at a larger scale in low recombining regions or at a finer scale in high recombining regions. At a larger scale, the multiple effects of individual alleles are combined in long introgressed tracts (Anderson and Stebbins 1954; Leitwein et al. 2018, 2019b). Thus, in the

Discoveries

Open Access

case of selection against the maladapted alleles within introduced tracts (i.e. genetic barrier to introgression or outbreeding depression), the purging is expected to be faster and more efficient in low recombining regions (Duranton et al. 2018; Schumer et al. 2018; Leitwein et al. 2019b; Martin et al. 2019; Veller et al. 2019). Such patterns have been observed in the swordtail fish, where deleterious combinations of epistatically interacting introgressed alleles were counter-selected more quickly in the low recombining regions than in the high recombining regions (Schumer et al. 2018). Inversely, in case of positive effects of the introgressed tracts, by masking the effect of potential recessive deleterious alleles (i.e. associative overdominance), selection will favor the introgressed tracts in low recombining regions (Harris and Nielsen 2016; Leitwein et al. 2019a, 2019b, 2021). Selection will act at a finer scale in high recombining regions, where consequently both maladaptive (Sankararaman et al. 2014; Leitwein et al. 2019b; Moran et al. 2021) and adaptive effect of introgression (Hedrick 2013; Sankararaman et al. 2014; Racimo et al. 2015) may arise.

The European seabass (*Dicentrarchus labrax*) is an interesting model to evaluate the evolutionary consequences of hybridization and its impacts on phenotype. This is an important species extensively produced and developed in aquaculture (Vandeputte et al. 2019). Its phenotypes are well described and studied not only in aquaculture lineages but also in wild populations (Vandeputte et al. 2019). There are two lineages, the Atlantic and Mediterranean, hybridizing in the Alboran Sea (Lemaire et al. 2005) with gene flow mainly occurring from the Atlantic to the Mediterranean population (Duranton et al. 2018). A strong phenotypic differentiation between Atlantic and Mediterranean populations was shown for some traits. For instance, Doan et al. (2017) reported large differences in their resistance to betanodavirus infection. Variation in growth and somatic indexes was also identified between some Atlantic and Mediterranean populations (Vandeputte et al. 2014). Interestingly, Vandeputte et al. (2014) pointed out, in a previous common garden experiment that the northern Atlantic European seabass population was characterized by a higher muscle fat content than that of the Western and North-Eastern Mediterranean populations. This difference in muscle fat content could be reflecting local adaptation to temperature (Purchase and Brown 2001; Olsen et al. 2021), which is on average higher in the Mediterranean than in the Atlantic. Additionally, Duranton et al. (2018) observed selection against the introgression of Atlantic ancestry within the Mediterranean basin is illustrated in two ways. First, a reduced introgression of Atlantic ancestry in specific genomic regions, pointing to a possible role of those regions in reproductive isolation. Second, introgression varies spatially, and this reflects both temporal and spatial variation in selection against introgressed tracts.

The aim of this study was to understand how selection and recombination during introgression act on muscle fat content, a quantitative trait which is divergent between the two lineages. To do so, we used genomic data from the 57 K single nucleotide polymorphisms (SNPs) DlabChip array to retrieve the individual ancestry of fish from three European seabass populations belonging to the two lineages (Atlantic and Mediterranean) raised in a common garden experiment. Then, as gene flow is mainly occurring from the Atlantic to the Mediterranean populations, we assessed the mosaic of Atlantic ancestry over both Mediterranean populations. More specifically, we first assessed how selection coupled with recombination shapes the genome-wide mosaic of introgression tracts of Atlantic ancestry within Mediterranean populations. Then, we took advantage of the common garden experiment to highlight genetic differences in muscle fat content between Atlantic and Mediterranean populations. Using the admixed Western Mediterranean population, which shows the highest level of variation in Atlantic ancestry we examined how muscle fat content, a divergent phenotype, was impacted either by selection or recombination, or by their interplay.

Materials and Methods

Broodstock Origin
Males and females broodstock used in this study belong to the two European seabass (*Dicentrarchus labrax*) lineages. First, the Atlantic line (later called ATL) was formed of 531 ATL (F1) offspring from the mating of 9 dams and 25 sires from Brest, France. Secondly, the Mediterranean line was subset in two subpopulations: (i) the Western Mediterranean line (later called WEM) was formed of 526 WEM offspring from the mating of 23 dams and 39 sires from the Gulf of Lions, France; (ii) the Eastern Mediterranean line (later called EST) was formed of 695 EST offspring from the mating of 13 dams and 38 sires from Eilat, Israel and Beymelek, Turkey. The WEM population is the closest to the ATL lineage. All artificial mating of wild parents and fish rearing were carried out under classical recirculating aquaculture system (RAS) conditions with a stable temperature of 21 °C after a larval rearing phase at 16 °C until 90 d post-hatching (dph), and ad libitum feeding using a diet formulated for seabass (Neogrower, Le Gouessant, France), at the IFREMER aquaculture facility (agreement number D-34-192-6, Palavas, France) in agreement with the French decree no. 2013–118 February 1, 2013 NOR:AGRG1231951D, regulating the use of animals for experiments.

A subsample of 1,182 F1 individuals were euthanized at 858 dph, and traits of each fish were recorded at that time. Measures of fat content were performed with a fat meter (Distell FatMeter®) for 372 ATL, 380 EST, and 430 WEM fish. All individuals were weighed and sexed whenever gonad observation was possible.

Phenotype Differences Between Populations
Muscle fat content was determined for each fish, and adjusted by a linear regression model with the fat content included as the response variable and both body weight and sex as explanatory variables. The residuals of this model were then used as the phenotype of interest, to eliminate the well-known phenotypic effect of body size on muscle

fat. Fat residuals were compared between the three populations by performing a one-way ANOVA test and a Tukey HSD (honestly significance difference) adjustment for multiple comparisons of means (Winer et al. 1971).

Genotyping, Imputation, and Phasing

All 1,899 parents and offspring fish were fin-clipped and genotyped using the Axiom European seabass 57k SNPs DlabChip array developed by Griot et al. (2021). SNP calling and quality control were performed using ThermoFisher software Axiom Analysis Suite software, with threshold values of 95% for SNP call rate and 90% for sample call rate. Vcftools v 0.1.16 (Danecek et al. 2011) was used to filter out SNPs with a Minor Allele Frequency (MAF) lower than 5%, leaving a total of 47,680 SNPs. Parentage assignment analysis was performed using the R package APIS v1 (Griot et al. 2020), using 800 markers displaying the highest MAF and a positive assignment error rate set to 1%. COLONY v 2.0.6.6 (Jones and Wang 2010) was used to infer the genotype of three missing male parents and one female parent. APIS was run a second time with the four missing parents to generate a pedigree for all offspring (supplementary table S1, Supplementary Material online). The imputation and the phasing of missing genotypes were performed with Fimpute v3 with *parentage_test, find_match_cnflt* and *save_genotype* parameters (Sargolzaei et al. 2014).

Ancestry Inference

Local ancestry inference of the 1,752 progenies from the three populations (ATL, WEM, and EST) was performed with Loter v. 1.0.1 (Dias-Alves et al. 2018). Loter does not require parameter tuning due to a model aggregating strategy (Dias-Alves et al. 2018). However, to infer the ancestry and identify the Atlantic migrant tracts within the Mediterranean line, Loter requires reference individuals for both lineages (ATL and MED). Therefore, we ran Admixture 1.3.0 (Buerkle and Lexer 2008) and kept the 9 ATL females and 17 ATL males with more than 99% of ATL ancestry (estimated by Admixture) as the Atlantic ancestry reference provided to Loter. We then kept 5 EST females and 37 EST males with less than 1% of Atlantic ancestry as the Mediterranean ancestry reference for Loter. All 1,752 progenies were used as the admixed population with ancestry to be estimated. Loter was run separately for each of the 24 chromosomes. We then retrieved the number and length of ATL ancestry tracts in each individual within the EST and WEM populations. Linear modeling was used to investigate if the number and length of Atlantic ancestry tracts differed between the two Mediterranean populations. Atlantic ancestry tract length was log-transformed and the population of origin (WEM or EST) was introduced as an explanatory term. Significance was tested by comparing models with or without the explanatory variable. All analyses were performed with R v4.2.1 (R Core Team 2022). We then estimated the per SNP local ATL ancestry rate (i.e. the proportion of ATL ancestry in the population for each SNP marker along the genome), for both EST and WEM populations.

Linkage Map and Recombination Rate

A genetic linkage map for the Mediterranean lineage was constructed with LepMap3 (Rastas 2017) using the offspring and parent genotypes of the 222 families of the EST population, as this is the population with the highest proportion of Mediterranean background (as opposed to the WEM population which is an admixture of the ATL and EST populations). The recommended procedure of LepMap3 was used with a segregation distortion for Filtering2 step set to 0.0001, a missing genotype (missing limit) set at 0.2, and a LOD score of 50. The LOD score was then set at 30 for the JoinSingles2All step. Here and throughout, we will refer to "linkage group" (LG) when referring to the genetic position in cM and "chromosome" when referring to the physical positions in bp.

To estimate the local recombination rate, we compared our newly constructed genetic map (for both sexes separately) to the physical positions from the European seabass reference genome (GCA_000689215.1) using MAREYMAP (Rezvoy et al. 2007). The LOESS method was used to determine the local recombination rate in cM per Mb. We estimated the mean population local recombination rate by estimating sex-averaged local recombination rates. For missing values, a weighted mean recombination rate was estimated using the two closest markers (based on their relative physical positions) with a local recombination rate value.

The mean recombination rate per 2-mb sliding windows has then been estimated for comparison with the population-scaled recombination rate rho estimated by Duranton et al. (2018).

Local Ancestry Rate as a Function of Recombination Rate

To investigate the main evolutionary mechanisms shaping the introgressed ATL ancestry within the Mediterranean genetic background, we assessed the relationship between local recombination rate and local ancestry rate along the genome, or more precisely the proportion of ATL ancestry across all individuals and SNP markers for the EST and WEM populations. To do so, we performed generalized linear mixed models via penalized quasi-likelihood (glmmPQL from R package MASS). The proportion of local ATL ancestry was set as the response variable and the scaled mean Mediterranean local recombination rate as an explanatory variable. As the local ATL ancestry rate is defined as a proportion, we used quasi-binomial mixed logistic regressions with chromosomes as random effects. GlmmPQL allows us to estimate the slope of the regression within random effects (i.e. chromosomes) but not to estimate the uncertainty of the effects (i.e. confidence interval per chromosome). For a graphical representation of the confidence intervals, quasi-binomial logistic regressions (GLMs) were performed separately for each chromosome. All graphical representations were performed with R v4.2.1 (R Core Team 2022).

Genome-Wide Association Studies

Genome-wide association studies (GWAS) were performed using GEMMA v.0.98.5 based on a dataset consisting of the

ancestry information for 1,181 individuals and their fat residuals. For each of the 42,796 SNPs, the genotype was replaced by the ancestry obtained from Loter with 0, 1, and 2 for Atlantic homozygous, Atlantic/Mediterranean heterozygous, and Mediterranean homozygous, respectively. The statistical model used to test one marker at a time was described by the following equation: $y = x\beta + Zu + \varepsilon$ (Zhou and Stephens 2012). In this equation, y represents a vector containing the n fitted phenotypes of all individuals and x is the vector of marker genotypes coded as ancestry as defined above, β is the effect size of ancestry at the marker, Z is an appropriate incidence matrix and u is the random effect on phenotypes taking into account the population structure related to ancestry, with u an nx1 vector of random effects following a multivariate normal distribution MVNn $(0, \lambda\tau^{-1}$ **A**) ε is a $n \times 1$ vector of residual errors following MVNn $(0, \lambda\tau^{-1}$ **I**$_n$), τ^{-1} is the variance of the residual errors, λ is the ratio between the two variance components, **A** is a known $m \times m$ relatedness matrix (equivalent to the genomic relationship matrix but built with ancestry coefficients), **I**$_n$ is a $n \times n$ identity matrix. To account for multiple testing issues, the significance threshold was obtained after a Bonferroni correction: $-\log 10 * (0.05/\text{mean number of independent tests})$. For each chromosome, we retrieved the quantitative trait locus (QTL) with the most significant P-value and displayed for each individual the fat residuals and the ancestry on a violin plot performed in R v4.2.1 (R Core Team 2022). For comparison, a standard GWAS using actual SNP genotypes was also performed.

To further control for population structure, we performed a GWAS where instead of the fat content, we introduced the individual average proportion of Atlantic ancestry as a trait in the model. This allowed us to detect regions associated with ancestry only, not taking the phenotype into account.

Genomic Scores
In order to evaluate the relationship between fat and ancestry, we computed a genomic score for each of the 430 WEM offspring. The genomic score was estimated by using all the significant markers in the genomic regions of interest detected by the GWAS analysis. For each individual, we calculated the number of occurrences having a Mediterranean ancestry. For example, a pure Atlantic individual has a genomic score of zero (no Mediterranean ancestry over all positions in the genomic regions of interest), conversely, a pure East Mediterranean individual will display a maximal genomic score (number of markers × 2). We then performed in R v4.2.1 (R Core Team 2022). A Spearman correlation between the genomic scores and the residuals of fat estimated for the GWAS analysis.

Genes and Functional Enrichments
We retrieved the genes present and their function for each genomic region of interest detected by GWAS analysis from the Ensembl databases with biomaRt packages in R. We used the gene annotations from Tine et al. (2014). For enrichment analysis, we used the Protein Analysis Through Evolutionary Relationships (PANTHER 19.0) tool to determine Gene Ontology (GO) with a statistical over-representation test based on the biological process.

Results

Phenotypes Characteristics
ATL fish displayed significantly more muscle fat content than WEM (Tukey's test P-adj < 0.01, Fig. 1), and EST fish (Tukey's test P-adj < 0.01, Fig. 1) and WEM fish were also fatter than EST fish (Tukey's test P-adj < 0.05, Fig. 1).

SNP Calling, Ancestry Tracts, and Rate
Among the 57k SNPs of the Axiom DlabChip array, a total of 47,681 SNPs with MAF higher than 5% were retained. SNPs assigned to the ungroup chromosome "UN" for the reference genome were removed, leaving a total of 42,796 SNPs distributed along the 24 European seabass chromosomes for further analysis. Overall, the mean length of Atlantic ancestry tracts was significantly shorter within the EST population (mean$_{length}$ = 1.31 ± 1.05 Mb) than within the WEM population (mean$_{length}$ = 1.35 ± 1.27 Mb, Fig. 2a, F = 16.48 on 113,371 degrees of freedom, P-value = 5e−5). In addition, the number of Atlantic ancestry tracts by individual was significantly greater in the WEM population (mean$_{number}$ = 199.5 ± 15.33, Fig. 2b) than in the EST population (mean$_{number}$ = 47.43 ± 30, Fig. 1b, F = 8,084 on 730 degrees of freedom, P-value < 2e−16).

The genome-wide level of Atlantic ancestry rate was estimated for both the EST and WEM populations using the 42,796 SNPs distributed along the 24 European seabass chromosomes. The Atlantic ancestry rate in the WEM population was highly variable among and along the 24 chromosomes with a mean of 0.259 (95% CI, 0.078–0.432). For each chromosome, local genomic regions displaying excess and deficit (exceeding the CI) of Atlantic introgression could be observed (supplementary fig. S1a, Supplementary Material online). The Atlantic ancestry rate in the EST population was also highly variable within and between chromosomes with excess and deficit of introgression but displayed a much lower mean ancestry than the WEM population, with a mean of 0.019 (95% CI, 0 to −0.048) (supplementary fig. S1b, Supplementary Material online).

Linkage Map and Recombination Rate Estimation
A total of 42,741 markers were homogeneously assigned to 24 linkage groups (LGs) for the EST European seabass genetic map corresponding to the 24 known chromosomes of the species. The male genetic map was shorter, with a total length of 1154.86 cM and a density of 37 markers per cM, compared to the female genetic map, with a total length of 1657.01 cM and a density of 25.8 markers per cM (supplementary fig. S2a and b, Supplementary Material online and supplementary table S2, Supplementary Material
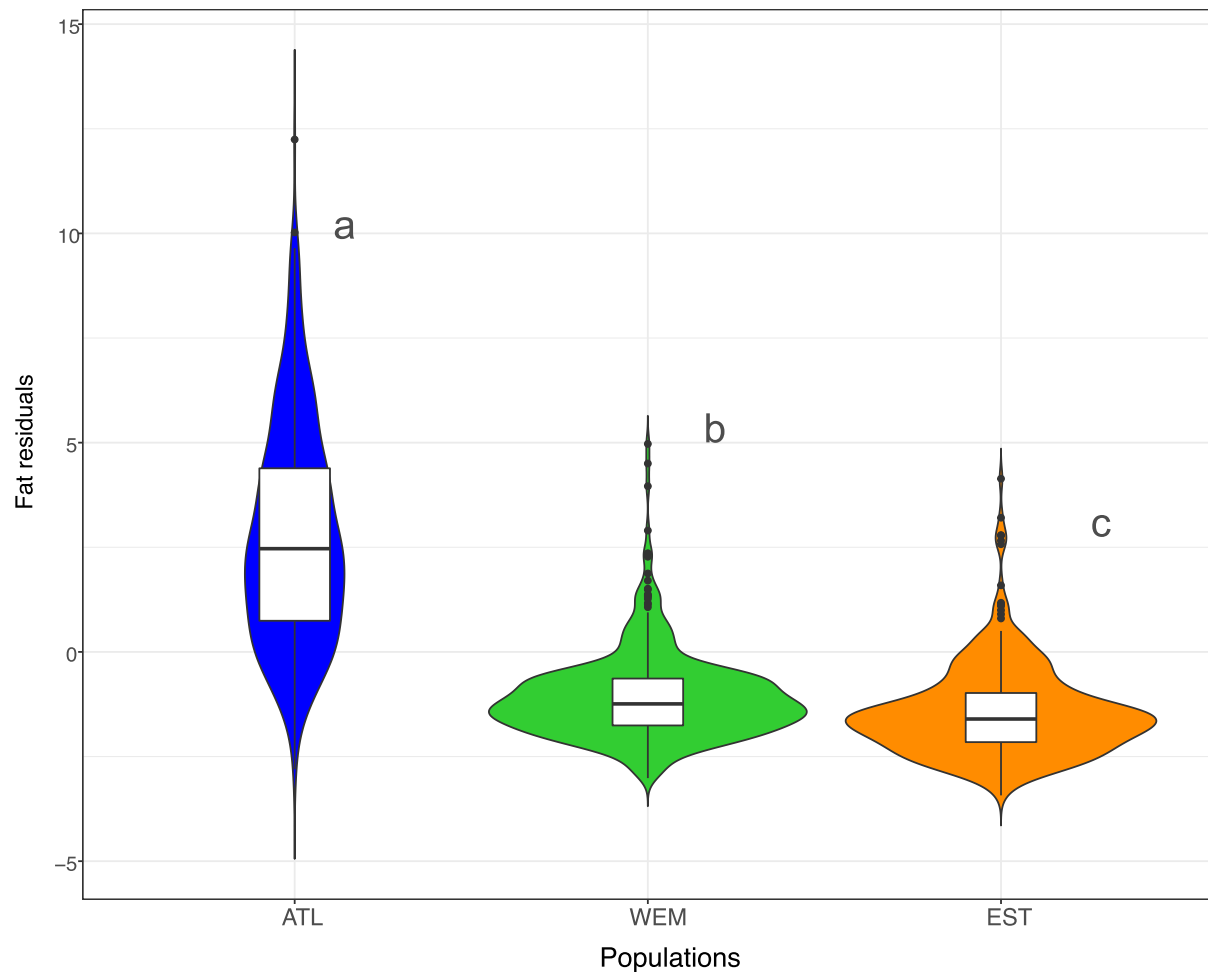
Fig. 1. Violin plot of the muscular fat content residuals for the three populations Atlantic (ATL) and both west and east Mediterranean populations (WEM and EST). Letters represent the significance of Tukey's test (P-adj < 0.01), average with no significant differences are shown with the same letter.

online). As for many species, the male linkage map tends to be smaller than the female linkage map due to a lower recombination rate in males (Sutherland et al. 2016; Doan 2017; Leitwein et al. 2017; Sardell and Kirkpatrick 2020; Guerrero-Cózar et al. 2021). For further population analyses, we chose to use the consensus between male and female genetic maps (supplementary fig. S2c, Supplementary Material online). The resulting mean total genetic map length was 1,405 cM, with individual LGs ranging from 51.53 cM (LG 24) to 65.79 cM (LG 1) (supplementary fig. S2c, Supplementary Material online and supplementary table S2, Supplementary Material online).

Then, by using the male and female genetic maps compared to the genomic position, we estimated the mean local recombination rate in cM/Mb for 40,856 markers along the 24 chromosomes. The local mean recombination rate was highly variable between and within LGs with a mean value of 2.69 ± 2.29 cM/Mb (Fig. 3 and supplementary table S3, Supplementary Material online).

The mean 2Mb sliding window recombination rate was significantly positively correlated with the population-scaled recombination rate estimated by Duranton et al. (2018) (rho spearman = 0.703; supplementary fig. s3, Supplementary Material online).

Recombination and Atlantic Ancestry Rate

At the genome-wide level, we observed a positive correlation between the local recombination rate and the local Atlantic ancestry rate in both the WEM and EST populations (Fig. 4). The estimated slope coefficients for the generalized linear mixed models were 0.46 for the WEM population and 0.42 for the EST population. The penalized quasi-likelihood test was significant for both models ($P < 0.001$). Consequently, at the genome-wide level, for both populations, we observed more Atlantic ancestry in high recombination rate regions and less Atlantic ancestry in low recombination rate regions (Fig. 4a).

At the chromosomal level, the relationship between the local Atlantic ancestry rate and the local recombination rate was also predominantly positive. Figure 4b displays the slope coefficient within chromosomes of the GlmmPQL models for both WEM and EST populations. Three chromosomes (LG2, LG22-25, and LG24) displayed
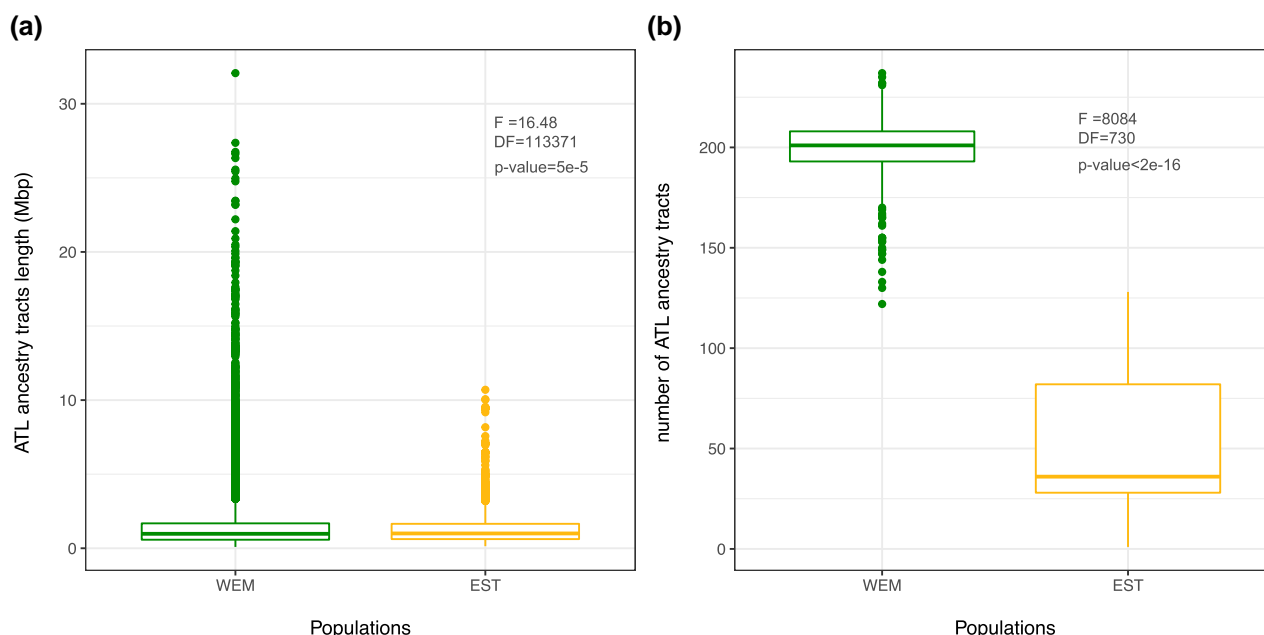
Fig. 2. Atlantic (ATL) ancestry tracts for both WEM and EST populations. a) Boxplot of the ATL ancestry tracts length in Mbp within each population. b) Boxplot of the number of ATL ancestry within each population. *F* indicates the significance of the models and DF the degree of freedom followed by the *P*-value.
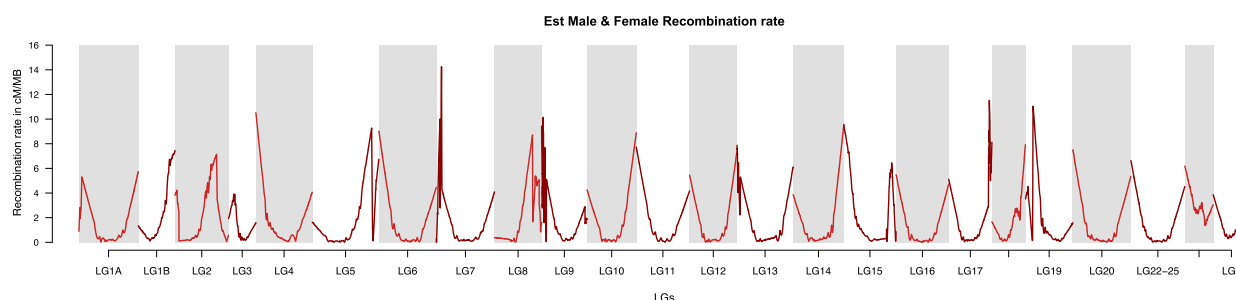

Fig. 3. Mean recombination rate estimates (in cM/Mb) between males and females of the EST line along the 24 chromosomes of the European seabass.

a negative relationship between local Atlantic ancestry and local recombination rate for the EST population (Fig. 4b and supplementary fig. S4, Supplementary Material online). All other chromosomes displayed a positive relationship between the two rates (Fig. 4b). In LG2, we could observe that genomic regions driving the negative correlation are located around 5 MB with a large local Atlantic ancestry rate associated with a low local recombination rate, and around 16Mb with a low local Atlantic ancestry rate associated with a highly recombining region (supplementary fig. S5, Supplementary Material online).

Relationship Between Ancestry and Phenotype

While the standard GWAS (supplementary fig. S6, Supplementary Material online) using actual SNP genotypes did not show any significant association of SNPs with muscle fat, a total of 637 SNP positions, distributed on 14 chromosomes, displayed a significant association between ancestry and fat (Fig. 5a). In parallel, 72 of those positions distributed on four chromosomes may also reflect the population structure as the GWAS for local ancestry and the individual's average ATL ancestry rate were both significant at these positions (blue vertical line Fig. 5a and supplementary fig. S7, Supplementary Material online). For the 14 most significant QTLs in each chromosome, homozygous individuals in terms of ATL ancestry presented a higher fat residual than heterozygous and homozygous Mediterranean individuals (Fig. 5b and supplementary fig. S8, Supplementary Material online). For four chromosomes (LG1B, LG4, LG9, LG11, and LG14) and their associated QTL, heterozygous individuals displayed more fat than individuals with homozygous Mediterranean ancestry (Fig. 5b and supplementary fig. S7, Supplementary Material online).
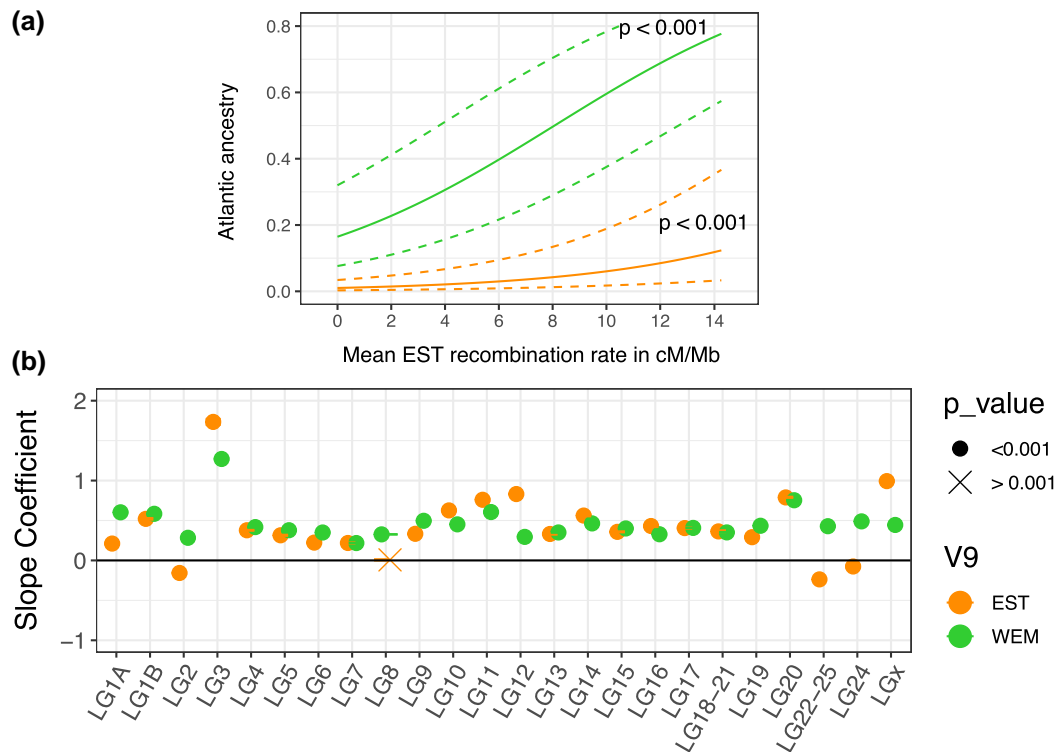
(a)

(b)

Fig. 4. a) Whole-genome relationship between local Atlantic ancestry rate and local recombination rate (cM/Mb) for both WEM (green) and EST (orange) populations. The colored plain curve is the mean population response of local Atlantic ancestry rate to local recombination rate predicted by the quasi-binomial mixed logistic regressions with penalized quasi-likelihood (GlmmPQL). The slope coefficient was 0.46 and 0.42 for the WEM (green) and EST (orange) with both significant P-value < 0.001. Dashed-colored curves represent the confidence interval (CI 95%) of the predicted response. b) Slope of the response of local Atlantic ancestry rate to local recombination rate estimated for each of the 24 European seabass chromosomes introduced as random effects in the GlmmPQL for both EST (orange) and WEM (green) populations. Symbols' shape represents the significance of each separated logistic regression model (GLMs) per chromosome with circles and crosses representing P-values < 0.001 and P > 0.001, respectively.

Additionally, we highlight in pink vertical lines in Fig. 5a; the genomic regions for which the Atlantic ancestry rates in the WEM population were less than 5%, in order to highlight potential regions of isolation between Atlantic and Mediterranean lineages. The 637 SNP positions in the significant QTLs were located in genomic regions displaying low Atlantic ancestry introgression within WEM and EST populations as well as low recombination rate (Fig. 5c and supplementary fig. s7, Supplementary Material online).

Using the 637 SNPs located in the QTL regions, we estimated for each of the 430 WEM individuals a genomic score of the ancestry, and fish with no Atlantic ancestry at these 637 SNPs had a genomic score of 1,274 (Fig. 6). We observed a significant positive correlation ($rho_{spearman} = -0.145$, $P < 0.0001$; Fig. 6) between the genomic score and the fat residuals. This suggests that individuals with Atlantic ancestry at these QTLs are those with higher fat content.

Biological Process Analysis
We found 399 genes within the genomic regions of interest detected by the GWAS analysis (supplementary table S4, Supplementary Material online). The analysis revealed strong enrichment of genes involved in transmembrane

transport activities (GO: 0055085, 3.56-fold enrichment, FDR = 2.65E−04, supplementary table S4, Supplementary Material online). Two genes abcc6a and SPATA20 located in LG1B (SPATA20, from 981,354 bp to 1,023,555 bp supplementary table S4, Supplementary Material online) and LG7 (abcc6a, from 8,804,849 bp to 8,829,425 bp, supplementary table S4, Supplementary Material online) are of particular interest (see discussion).

Discussion
The main objective of this study is to understand how a quantitative polygenic trait is affected in the case of selection against hybridization between the two main lineages of the European seabass (*Dicentrarchus labrax*). The Atlantic lineage is characterized by a higher muscle fat content than the Mediterranean lineage. In order to assess the relationship between ancestry and phenotype, we needed to first evaluate the evolutionary processes modulating the mosaic of introgressed ancestry tracts in the hybrid zone. To this end, we assessed the genome-wide pattern of Atlantic ancestry within the West and East Mediterranean populations of European seabass. We observed more and longer Atlantic tracts in the West
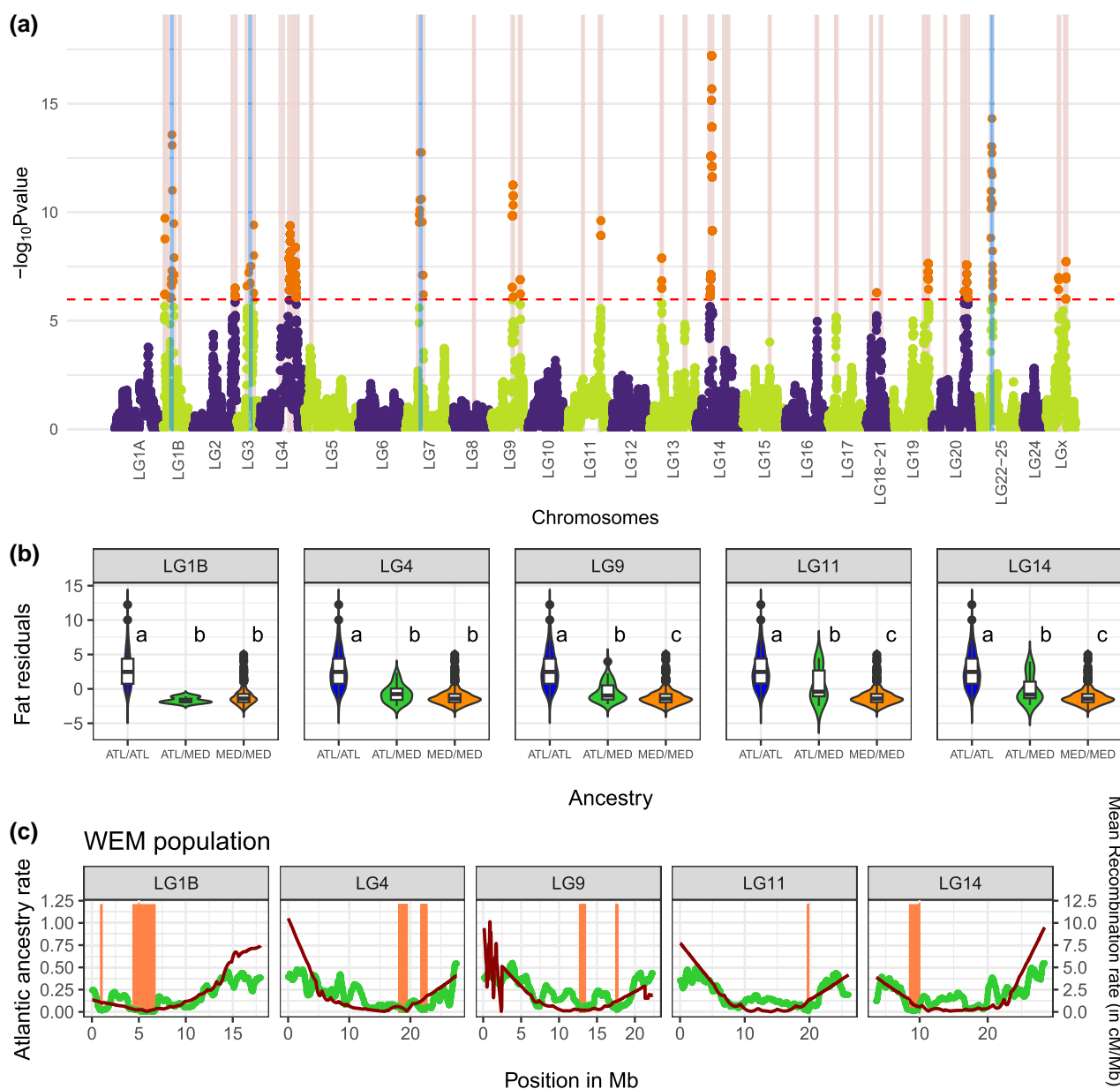
Fig. 5. a) Manhattan plot displaying the significance of each "ancestry" variant association with muscle fat content. Each dot represents an SNP carrying the ancestry information distributed along the 24 European seabass chromosomes. The *y* axis represents the strength of the association as −log 10 of the *P* values. The dashed horizontal line represents the significance threshold computed with a Bonferroni correction. Blue vertical line represents the position of QTLs associated with population structure. Pink vertical line represents genomic regions with less than 5% of ATL ancestry. b) Violin and box plots of the fat residuals as a function of the ancestry (homozygous ATL, heterozygous ATL/MED, and homozygous MED) for the most significant QTL per LG. The letters indicate the significance of ANOVA and Tukey's test. c) Plots of local Atlantic ancestry rate and local recombination rate (in cM/Mb) of five chromosomes for the WEM populations. Green dots represent the local Atlantic ancestry rate within the WEM population at each SNP position. The red line represents the mean local recombination rate in cM/Mb. The *x* axis is the position in Mb within each chromosome. Vertical dark orange lines represent the QTLs position retrieved from the GWAS analysis.

Mediterranean population, which is the closest population to the hybrid zone, compared to the East Mediterranean population, which is the furthest. Coupling information about local ancestry rate and local recombination rate, we highlight that, at both genome-wide and chromosomal scales, selection against introgressed Atlantic ancestry occurs, resulting in a low introgression rate in low recombining regions. This suggests selection against the Atlantic ancestry through a genetic barrier to introgression or outbreeding depression. We found 14 QTLs where local ancestry affects muscle fat content. Individuals with less Atlantic ancestry in these regions are also those with low muscle fat content. Furthermore, these regions are located in low introgressed and low recombining regions, suggesting there is selection against loci associated with muscle fat content, in Mediterranean populations.
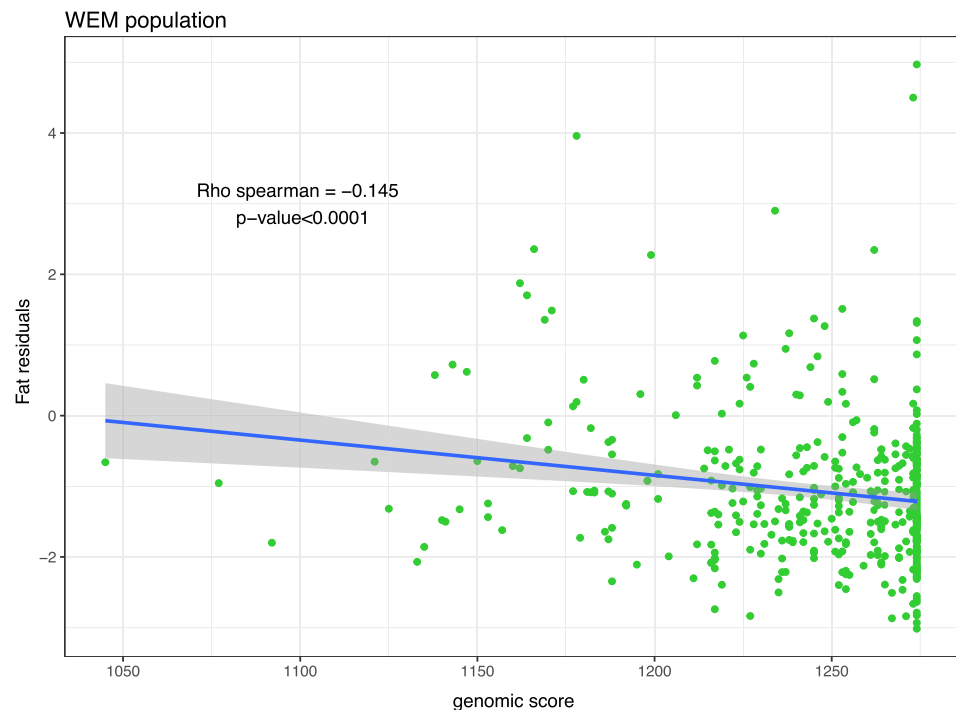
Fig. 6. Negative correlation between the ancestry genomic scores for the 637 QTLs and the fat residuals for each WEM individual represented by green dots (rho$_{spearman}$ = −0.145, P-value < 0.0001).

Mediterranean Recombination Rate Variation

The consensus recombination Mediterranean landscape was highly variable within and between linkage groups. We can observe many "U" patterns within linkage groups with lower recombination rates in the center and higher recombination at the extremities. Such patterns could be explained by higher recombination rates near telomeres and lower recombination rates near centromeres (Nachman and Payseur 2012; Guerrero-Cózar et al. 2021). However, the exact location of the centromere remains unclear (Oral et al. 2017) and was beyond the scope of this study. Finally, our estimated landscape of local recombination rate was consistent with the population-scaled recombination rate rho, which is dependent on the effective population size (rho = 4Ner, Ne is the effective population size, and r the recombination rate for the window) estimated by Duranton et al. (2018).

Selective Effects Following Gene Flow

Hybridization occurring in the Alboran Sea is mainly oriented from the Atlantic to the Mediterranean with selection acting mainly against hybridization (Duranton et al. 2018). Consequently, the West Mediterranean (WEM) population, being the closest to the contact zone, displays more and longer Atlantic tracts. The East Mediterranean (EST) population, being the furthest from the hybrid contact zone, displays fewer and shorter Atlantic ancestry tracts due to dilution, recombination events, and selection for a higher number of generations. This is consistent with the patterns described by Duranton et al. (2018). Geographic distance from the hybrid contact zone is a proxy of the timing of hybridization, and the WEM

population thus consists of earlier generation hybrids than the EST population. Post-hybridization timing is important to consider, as selective effects will operate at a larger tract scale in the earlier hybrids and locally in later hybrid generations (Leitwein et al. 2019b). While selective effects vary with time they also vary with the recombination rate landscape (Schumer et al. 2018). Here, we observed both in WEM and EST populations a positive relationship between local introgression rate and local recombination rate. Such relationships and variable patterns of introgression among chromosomes highlight an initial rapid purging of introduced ancestry (Veller et al. 2019; Moran et al. 2021). This purging slows over time as the size of introgressed tracts is reduced by recombination (Veller et al. 2019). This is slightly reflected in our study, as the slope coefficient between introgression and recombination rate is weaker in the EST population than in the WEM population. Purging is the main evolutionary mechanism observed at both genome-wide and chromosomal scales revealing outbreeding depression (Barton and Hewitt 1985). Outbreeding depression can occur when mutations from both ancestral lineages negatively interact resulting in Dobzhansky–Muller hybrid incompatibilities (DMIs) (Unckless and Orr 2009) or due to rupture of co-adapted parental alleles (Schumer et al. 2015; Moran et al. 2021). Such negative epistasis is a common mechanism following hybridization (Moran et al. 2021; Adavoudi and Pilot 2022). For example, in swordtail fish negative epistatically interacting alleles drove the pattern of introgression with a purge of the minor parents' ancestry in low recombining regions (Schumer et al. 2018). In addition, genetic incompatibilities are a key component of reproductive isolation barriers and play a major role in speciation. Consequently, the genomic regions

free of Atlantic introgression, the so-called genomic islands of differentiation, as previously described by Duranton et al. (2018), could reflect a reproductive isolation barrier or even selection against maladaptive alleles.

Muscle Fat Content and Ancestry Association

The Atlantic European seabass population had a higher muscle fat content compared to the Mediterranean populations. These results confirm previous observations by Vandeputte et al. (2014). The difference in muscle fat content could possibly be explained by the fact that Atlantic individuals tend to experience longer periods of starvation in winter, and therefore more muscle fat may allow better winter survival. However, we lack information and physiological data to fully understand the potential adaptive benefit behind these differences in fat content. Admixed individuals are usually removed from GWAS analyses as complex population substructures can create false positives (Martin et al. 2018; Sul et al. 2018). Yet, a recent study (Atkinson et al. 2021) demonstrated that using both ancestry and genotypes improves the detection of association signals and allows us to include admixed individuals. However, in this approach, the authors tested the effect on traits of a genotype conditional on ancestry. In our study, we wanted to test the direct effect of ancestry on a phenotype instead, which is why we replaced the SNP genotype with ancestry at each locus in our GWAS analysis. As a result, by using the haplotype information, we were able to detect an association between ancestry and muscle fat content for 14 genomic regions. Muscle fat content is a quantitative polygenic trait. The use of haplotype information allows the detection of small associated effects (Leitwein et al. 2019b) that were not found by using directly local SNPs, despite these being well distributed along the genome. Admittedly, for four of the QTL regions associated with muscle fat content, confounding effects with population structure could occur as local ancestry and the mean individual percentage of Atlantic ancestry were associated with these regions. However, individuals with more Atlantic ancestry in these specific regions were also those with higher muscle fat content, strengthening the evidence of the role of these regions in muscle fat phenotype. Consistently, all of our QTLs were located in genomic regions with less than 5% Atlantic ancestry, and thus potentially regions involved in reproductive isolation. Interestingly, the QTL regions characterized by low Atlantic introgression also show a low recombination rate. This suggests a purge of Atlantic ancestry in these regions and possibly a counter-selection for muscle fat content in Mediterranean populations. Stocking more fat content requires energy at the expense of growth or reproduction, and may therefore be disadvantageous in the Mediterranean environment where little benefit in terms of overwintering survival is expected as winter starvation is less common.

From the biological process analysis, we highlight a strong enrichment of genes linked with transmembrane transport processes (GO: 0055085, 3.56 fold enrichment, FDR = 2.65E−04). While a direct link between such biological processes remains to be confirmed, two of those genes were previously reported as strongly involved in adipose tissue storage and peripheral lipid metabolism in humans, namely abcc6a (10.1186/1476-511X-13-118; https://doi.org/10.3390/ijms23169218) and SPATA20 (https://doi.org/10.1371/journal.pone.0217644), the latter being also involved in mice spermatogenesis. These two genes are located in QTL regions in the chromosomes LG1B (SPATA20) and LG7 (abcc6a); these regions have been pointed as genomic islands of differentiation underlying partial reproductive isolation between Atlantic and Mediterranean lineages (Duranton et al. 2018). While differentiation around these genes associated with fat deposition may explain phenotypic differentiation between the two lineages, elucidating their possible pleiotropic effect on partial reproductive isolation will require further examination. However, this study brings new insights into understanding further the link between genomic differentiation between Atlantic and Mediterranean European seabass and its phenotypic consequences.

Conclusion

Understanding the evolutionary consequences of hybridization between populations is of paramount interest in evolutionary biology. Genome-wide patterns of introgressed ancestry are shaped by both selection and recombination (Duranton et al. 2018; Schumer et al. 2018; Veller et al. 2019; Leitwein et al. 2019b; Moran et al. 2021). Here, we highlight how a divergent polygenic phenotype, muscle fat content in European seabass, is affected by this interplay between selection and recombination during hybridization. To do this, we used the haplotype information, which increases statistical power to detect genomic regions associated with the polygenic traits in admixed populations. Introgressed Atlantic ancestry is quickly purged in Mediterranean populations, suggesting a reproductive isolation barrier and outbreeding depression. In addition, regions associated with muscle fat content are also counter-selected in the Mediterranean, reflecting a potential maladaptation of the phenotype in the Mediterranean environment. However, it remains unclear whether selection against a maladapted phenotype shapes the genomic landscape of introgression or whether the landscape of introgression shapes the phenotype.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability

Raw data are available at SEANOE https://doi.org/10.17882/98771. Scripts are available at https://github.com/mleitwein/Ancestry_GWAS.

References

Abbott RJ, Barton NH, Good JM. Genomics of hybridization and its evolutionary consequences. Mol Ecol. 2016:**25**(11):2325–2332. https://doi.org/10.1111/mec.13685.

Adavoudi R, Pilot M. Consequences of hybridization in mammals: a systematic review. Genes (Basel). 2022:**13**(1):50. https://doi.org/10.3390/genes13010050.

Anderson E, Stebbins GL. Hybridization as an evolutionary stimulus. Evolution. 1954:**8**(4):378–388. https://doi.org/10.1111/j.1558-5646.1954.tb01504.x.

Atkinson EG, Maihofer AX, Kanai M, Martin AR, Karczewski KJ, Santoro ML, Ulirsch JC, Kamatani Y, Okada Y, Finucane HK, et al. Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. Nat Genet. 2021:**53**(2):195–204. https://doi.org/10.1038/s41588-020-00766-y.

Baack EJ, Rieseberg LH. A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev. 2007:**17**(6):513–518. https://doi.org/10.1016/j.gde.2007.09.001.

Barton N, Hewitt GM. Analysis of hybrid zones. Annu Rev Ecol Syst. 1985:**16**(1):113–148. https://doi.org/10.1146/annurev.es.16.110185.000553.

Buerkle CA, Lexer C. Admixture as the basis for genetic mapping. Trends Ecol Evol. 2008:**23**(12):686–694. https://doi.org/10.1016/j.tree.2008.07.008.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, 1000 Genomes Project Analysis Group, et al. The variant call format and VCFtools. Bioinformatics. 2011:**27**(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Dias-Alves T, Mairal J, Blum MGB. Loter: a software package to infer local ancestry for a wide range of species. Mol Biol Evol. 2018:**35**(9):2318–2326. https://doi.org/10.1093/molbev/msy126.

Doan QK. Genetic and genomic variation of resistance to viral nervous necrosis in wild populations of European seabass (*Dicentrarchus labrax*). Anim. Genet. Université Montpellier, 2017 [accessed 2017 Nov 28]. http://www.theses.fr/2017MONTT094.

Doan QK, Vandeputte M, Chatain B, Haffray P, Vergnet A, Breuil G, Allal F. Genetic variation of resistance to viral nervous necrosis and genetic correlations with production traits in wild populations of the European sea bass (*Dicentrarchus labrax*). Aquaculture. 2017:**478**:1–8. https://doi.org/10.1016/j.aquaculture.2017.05.011.

Duranton M, Allal F, Fraïsse C, Bierne N, Bonhomme F, Gagnaire P-A. The origin and remolding of genomic islands of differentiation in the European sea bass. Nat Commun. 2018:**9**(1):2518. https://doi.org/10.1038/s41467-018-04963-6.

Griot R, Allal F, Brard-Fudulea S, Morvezen R, Haffray P, Phocas F, Vandeputte M. APIS: an auto-adaptive parentage inference software that tolerates missing parents. Mol Ecol Resour. 2020:**20**(2):579–590. https://doi.org/10.1111/1755-0998.13103.

Griot R, Allal F, Phocas F, Brard-Fudulea S, Morvezen R, Bestin A, Haffray P, François Y, Morin T, Poncet C, et al. Genome-wide association studies for resistance to viral nervous necrosis in three populations of European sea bass (*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip. Aquaculture. 2021:**530**:735930. https://doi.org/10.1016/j.aquaculture.2020.735930.

Guerrero-Cózar I, Gomez-Garrido J, Berbel C, Martinez-Blanch JF, Alioto T, Claros MG, Gagnaire P-A, Manchado M. Chromosome anchoring in Senegalese sole (*Solea senegalensis*) reveals sex-associated markers and genome rearrangements in flatfish. Sci Rep. 2021:**11**(1):13460. https://doi.org/10.1038/s41598-021-92601-5.

Harris K, Nielsen R. The genetic cost of Neanderthal introgression. Genetics. 2016:**203**(2):881–891. https://doi.org/10.1534/genetics.116.186890.

Hedrick PW. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol. 2013:**22**(18):4606–4618. https://doi.org/10.1111/mec.12415.

Jones OR, Wang J. COLONY: a program for parentage and sibship inference from multilocus genotype data. Mol Ecol Resour. 2010:**10**(3):551–555. https://doi.org/10.1111/j.1755-0998.2009.02787.x.

Leitwein M, Cayuela H, Bernatchez L. Associative overdominance and negative epistasis shape genome-wide ancestry landscape in supplemented fish populations. Genes (Basel). 2021:**12**(4):524. https://doi.org/10.3390/genes12040524.

Leitwein M, Cayuela H, Ferchaud A-L, Normandeau É, Gagnaire P-A, Bernatchez L. The role of recombination on genome-wide patterns of local ancestry exemplified by supplemented brook charr populations. Mol Ecol. 2019a:**28**(21):4755–4769. https://doi.org/10.1111/mec.15256.

Leitwein M, Duranton M, Rougemont Q, Gagnaire P-A, Bernatchez L. Using haplotype information for conservation genomics. Trends Ecol Evol. 2019b:**35**(3):245–258. https://doi.org/10.1016/j.tree.2019.10.012.

Leitwein M, Gagnaire P-A, Desmarais E, Berrebi P, Guinand B. Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. Mol Ecol. 2018:**27**(17):3466–3483. https://doi.org/10.1111/mec.14816.

Leitwein M, Guinand B, Pouzadoux J, Desmarais E, Berrebi P, Gagnaire P-A. A dense brown trout (*Salmo trutta*) linkage map reveals recent chromosomal rearrangements in the Salmo genus and the impact of selection on linked neutral diversity. G3 Genes Genomes Genet. 2017:**7**(4):1365–1376. https://doi.org/10.1534/g3.116.038497.

Lemaire C, Versini J-J, Bonhomme F. Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. J Evol Biol. 2005:**18**(1):70–80. https://doi.org/10.1111/j.1420-9101.2004.00828.x.

Lewontin RC, Birch LC. Hybridization as a source of variation for adaptation to new environments. Evolution. 1966:**20**(3):315–336. https://doi.org/10.2307/2406633.

Martin SH, Davey J, Salazar C, Jiggins CD. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLoS Biol. 2019:**17**(2):e2006288. https://doi.org/10.1371/journal.pbio.2006288.

Martin S, Jiggins CD. Interpreting the genomic landscape of introgression. Curr Opin Genet Dev. 2017:**47**:69–74. https://doi.org/10.1016/j.gde.2017.08.007.

Martin ER, Tunc I, Liu Z, Slifer SH, Beecham AH, Beecham GW. Properties of global- and local-ancestry adjustments in genetic association tests in admixed populations. Genet Epidemiol. 2018:**42**(2):214–229. https://doi.org/10.1002/gepi.22103.

McArthur E, Rinker DC, Capra JA. Quantifying the contribution of Neanderthal introgression to the heritability of complex traits.

Nat Commun. 2021:**12**(1):4481. https://doi.org/10.1038/s41467-021-24582-y.

Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. The genomic consequences of hybridization. eLife. 2021:**10**: e69016. https://doi.org/10.7554/eLife.69016.

Nachman MW, Payseur BA. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos Trans R Soc Lond B Biol Sci. 2012:**367**(1587): 409–421. https://doi.org/10.1098/rstb.2011.0249.

Olsen L, Thum E, Rohner N. Lipid metabolism in adaptation to extreme nutritional challenges. Dev Cell. 2021:**56**(10):1417–1429. https://doi.org/10.1016/j.devcel.2021.02.024.

Oral M, Colléter J, Bekaert M, Taggart JB, Palaiokostas C, McAndrew BJ, Vandeputte M, Chatain B, Kuhl H, Reinhardt R, et al. Gene-centromere mapping in meiotic gynogenetic European seabass. BMC Genomics. 2017:**18**(1):449. https://doi.org/10.1186/s12864-017-3826-z.

Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. Mol Ecol. 2016:**25**(11):2337–2360. https://doi.org/10.1111/mec.13557.

Purchase CF, Brown JA. Stock-specific changes in growth rates, food conversion efficiencies, and energy allocation in response to temperature change in juvenile Atlantic cod. J Fish Biol. 2001:**58**: 36–52. https://doi.org/10.1111/j.1095-8649.2001.tb00497.x.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 2015:**16**(6):359–371. https://doi.org/10.1038/nrg3936.

Rastas P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics. 2017:**33**(23): 3726–3732. https://doi.org/10.1093/bioinformatics/btx494.

R Core team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2022.

Reilly PF, Tjahjadi A, Miller SL, Akey JM, Tucci S. The contribution of Neanderthal introgression to modern human traits. Curr Biol. 2022:**32**(18):R970–R983. https://doi.org/10.1016/j.cub.2022.08.027.

Rezvoy C, Charif D, Guéguen L, Marais GA. MareyMap: an R-based tool with graphical interface for estimating recombination rates. Bioinformatics. 2007:**23**(16):2188–2189. https://doi.org/10.1093/bioinformatics/btm315.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014:**507**(7492):354–357. https://doi.org/10.1038/nature12961.

Sardell JM, Kirkpatrick M. Sex differences in the recombination landscape. Am Nat. 2020:**195**(2):361–379. https://doi.org/10.1086/704943.

Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014:**15**(1):478. https://doi.org/10.1186/1471-2164-15-478.

Schumer M, Cui R, Rosenthal GG, Andolfatto P. Reproductive isolation of hybrid populations driven by genetic incompatibilities. PLoS Genet. 2015:**11**(3):e1005041. https://doi.org/10.1371/journal.pgen.1005041.

Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 2018:**360**(6389):656–660. https://doi.org/10.1126/science.aar3684.

Sul JH, Martin LS, Eskin E. Population structure in genetic studies: confounding factors and mixed models. PLoS Genet. 2018:**14**(12): e1007309. https://doi.org/10.1371/journal.pgen.1007309.

Sutherland BJG, Gosselin T, Normandeau E, Lamothe M, Isabel N, Audet C, Bernatchez L. Novel method for comparing RADseq linkage maps reveals chromosome evolution in salmonids. bioRxiv 039164. https://doi.org/10.1101/039164, 25 May 2016, preprint: not peer reviewed.

Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F, Belkhir K, Klages S, et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat Commun. 2014:**5**(1):5770. https://doi.org/10.1038/ncomms6770.

Tobler M, Carson EW. Environmental variation, hybridization, and phenotypic diversification in Cuatro Ciénegas pupfishes. J Evol Biol. 2010:**23**(7):1475–1489. https://doi.org/10.1111/j.1420-9101.2010.02014.x.

Unckless RL, Orr HA. Dobzhansky–Muller incompatibilities and adaptation to a shared environment. Heredity (Edinb). 2009:**102**(3):214–217. https://doi.org/10.1038/hdy.2008.129.

Vandeputte M, Gagnaire P-A, Allal F. The European sea bass: a key marine fish model in the wild and in aquaculture. Anim Genet. 2019:**50**(3):195–206. https://doi.org/10.1111/age.12779.

Vandeputte M, Garouste R, Dupont-Nivet M, Haffray P, Vergnet A, Chavanne H, Laureau S, Ron TB, Pagelson G, Mazorra C, et al. Multi-site evaluation of the rearing performances of 5 wild populations of European sea bass (*Dicentrarchus labrax*). Aquaculture. 2014:**424–425**:239–248. https://doi.org/10.1016/j.aquaculture.2014.01.005.

Veller C, Edelman NB, Muralidhar P, Nowak MA. Recombination, variance in genetic relatedness, and selection against introgressed DNA. bioRxiv 846147. https://doi.org/10.1101/846147, 18 November 2019, preprint: not peer reviewed.

Winer BJ, Brown DR, Michels KM. Statistical principles in experimental design. New York: Mcgraw-Hill; 1971.

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012:**44**(7):821–824. https://doi.org/10.1038/ng.2310.