RESEARCH ARTICLE

MHC class I-α population differentiation in a commercial fish, the European sea bass (*Dicentrarchus labrax*)

Frances C. Ratcliffe | Carlos Garcia de Leaniz | Sofia Consuegra

Department of Biosciences, College of Science, Swansea University, Swansea, UK

Correspondence
Sofia Consuegra, Department of Biosciences, College of Science, Swansea University, Swansea SA2 8PP, UK.
Email: s.consuegra@swansea.ac.uk

Abstract

Identifying population structuring in highly fecund marine species with high dispersal rates is challenging, but critical for conservation and stock delimitation for fisheries management. European sea bass (*Dicentrarchus labrax*) is a commercial species of fisheries and aquaculture relevance whose stocks are declining in the North Atlantic, despite management measures to protect them and identifying their fine population structure is needed for managing their exploitation. As for other marine fishes, neutral genetic markers indicate that eastern Atlantic sea bass form a panmictic population and is currently managed as arbitrarily divided stocks. The genes of the major histocompatibility complex (MHC) are key components of the adaptive immune system and ideal candidates to assess fine structuring arising from local selective pressures. We used Illumina sequencing to characterise allelic composition and signatures of selection at the MHC class I-α region of six *D. labrax* populations across the Atlantic range. We found high allelic diversity driven by positive selection, corresponding to moderate supertype diversity, with 131 alleles clustering into four to eight supertypes, depending on the Bayesian information criterion threshold applied, and a mean number of 13 alleles per individual. Alleles could not be assigned to particular loci, but private alleles allowed us to detect regional genetic structuring not found previously using neutral markers. Our results suggest that MHC markers can be used to detect cryptic population structuring in marine species where neutral markers fail to identify differentiation. This is particularly critical for fisheries management, and of potential use for selective breeding or identifying escapes from sea farms.

KEYWORDS
aquaculture, fisheries, non-neutral markers

INTRODUCTION

With the decline of many traditional fisheries, accurate fish stock management has never been so critical (Beddington et al., 2007). However, identifying population structuring and genetic differentiation in marine populations using molecular markers can be challenging, particularly for highly fecund species with high dispersal rates and complex life cycles (Hedgecock et al., 2007). The use of large population genomic data sets and markers influenced by selection are increasingly revealing fine population structure and patterns of reproductive isolation even in highly dispersive marine species, with important implications for their conservation and management (Gagnaire et al., 2015). However, in many cases, lack of information about population structure results in a pragmatic approach to management (ICES, 2021) and a mismatch between biological and management units (Reiss et al., 2009), which can exacerbate the decline of stocks. Commercial fisheries are already contributing to the homogenisation of genetic diversity of marine fish species (Gandra et al., 2021) and ignoring fine scale local adaptation can further result in the loss of functionally important biodiversity (Limborg et al., 2012).

European seabass (*Dicentrarchus labrax*) is a traditionally important species in terms of commercial and recreational fisheries, and has more recently become a key species for aquaculture (Vandeputte et al., 2019). The extent of population structuring in European seabass is unclear, and in the Atlantic area has been arbitrarily divided by the International Council for the Exploration of the Sea (ICES) into four stocks: Stock West Ireland/ West Scotland; Stock Irish Sea, Celtic Sea, English Channel, and North Sea; Stock Bay of Biscay; and Stock Iberian Peninsula (De Pontual et al., 2019; European Parliament, Directorate-General for Internal Policies of the Union et al., 2016). The seabass stock found in the English Channel, Celtic and Irish Seas, and North Sea is characterised by slow growth and late maturation and its productivity is greatly influenced by sea water temperature (Walker et al., 2020). This stock has declined markedly since 2010, despite stringent quotas and fishing bans (Walker et al., 2020), mainly due to overfishing and low recruitment since 2008, but also due to the influence of long-term climate change and local environmental factors that affect mostly the early developmental stages (Bento et al., 2016). Tagging studies indicate that there is strong site fidelity (Pawson et al., 2008) and frequent migrations between the English Channel and the Bay of Biscay areas, and between the North Sea and the English Channel, highlighting the need for more accurate information on fine population structuring, necessary for correct stock management (Pawson et al., 2007). For example, the Irish stock is exploited separately from the Atlantic stocks although its distribution spans more than one ICES division (ICES, 2021). Existing studies based on mitochondrial DNA suggest that there is population structuring, not consistently supported by neutral markers. Mitochondrial DNA indicates that Atlantic and Mediterranean seabass populations are differentiated as a consequence of their isolation during the Pleistocene, that resulted in divergence in haplotype frequencies (Lemaire et al., 2005). However, a post-glacial secondary contact is thought to have eroded this divergence at the level of nuclear markers (Duranton et al., 2018), while differential introgression resulted in the creation of genomic islands of differentiation (Tine et al., 2014). Within the Atlantic seabass stock, there are three mitochondrial DNA lineages distributed predominantly in the Bay of Biscay (Atlantic 1; from where the Mediterranean lineage could have originated), European coast (Atlantic 2), and the British Islands and Norway (Atlantic 3) (Coscia & Mariani, 2011). In contrast, there is limited regional structuring based on nuclear markers (microsatellites and SNPs), with the exception of the populations in the south-eastern range of the Atlantic lineage (Portugal and Morocco) that display some Mediterranean influence in their SNP allelic composition (Souche et al., 2015). The allele distribution of markers under selection, such as allozymes or the somatolactin gene, allowed to identify some differentiation between the Bay of Biscay and the southern North Sea (Quéré et al., 2010) and at the local scale (Castilho & McAndrew, 1998), suggesting that markers under selection may prove better candidates to identify finer scale structuring for this species than neutral markers.

The genes of the major histocompatibility complex (MHC) are some the most studied and highly polymorphic genes in vertebrates (Edwards & Hedrick, 1998). MHC genes encode for proteins that present antigens to T-cells, triggering the adaptive immune response (Janeway et al., 2004) and most of their diversity is concentrated in the region that binds antigens from pathogens, the peptide binding region (PBR) (Hedrick & Kim, 2000). Polymorphism within the PBR determines to which pathogens an individual can respond (Radwan et al., 2020). PBR diversity is maintained by pathogen selection (Eizaguirre et al., 2012; Slade & McCallum, 1992) and in some species by mate choice (Consuegra & Garcia de Leaniz, 2008; Milinski, 2006), and can be influenced (directly or indirectly) by local environmental factors such as temperature (Dionne et al., 2007). We developed an Illumina sequencing-based protocol to genotype the PBR of the European seabass MHC class I-α gene, and investigated its potential for detecting fine scale population structuring and signatures of local selection pressures across the species' Atlantic range.

METHODS

Genomic DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit from 62 wild seabass individuals (caught between 2013 and 2015) from six areas of the

North Eastern Atlantic (Biscay: ICES rectangle 20 E8, $n = 6$, Dover: ICES 30 F1, $n = 7$, Irish Sea ICES 34E4, $n = 5$, Celtic Shelf: ICES 28 E1, $n = 22$, North Sea: ICES 38 E9, $n = 3$, and Portugal: ICES 4 E1, $n = 19$). Fish were genotyped using a custom designed primer pair: sbMHC_F2 5′ CTGGAGTCCCAAACTTCCC 3′, sbMHC_R2 5′ AGGTGGACACCTCCAGTTTG 3′, amplifying a 241-bp fragment of the protein coding sequence region of the MHC class I-α gene (Methods S1).

Libraries were prepared in a two-step PCR protocol, using Platinum™ Hot Start PCR Master Mix (2×; Thermo Fisher Scientific). Initially, 2.5 µl of DNA per individual was amplified in a 25-µl reaction (12.5 µl taq, 0.5 µl each of forward and reverse primer, 9 µl water), using the above primers with Illumina overhang adapter sequences added (Illumina, Inc., 16S Metagenomic protocol). PCR amplification consisted of the following: 2 min at 94°C, 28 cycles of 30 s at 94°C, 30 s at 55°C (locus specific annealing temperature), 1 min 72°C (extension). Subsequently, 2 µl of product from the initial reaction was used as template in a second PCR (12.5 µl taq, 1.25 µl each of a sample-specific combination of Nextera XT adapters (Index Kit v2; Illumina, Inc) and 10 µl water) and eight cycles of this second PCR was run with identical conditions to the first. A 5-µl aliquot of each indexed PCR product was pooled and the final pool was cleaned using Agencourt AMPure XP beads (Beckman Coulter), using a 1:1 ratio of beads to pooled template. The pool was then quantified using quatitative PCR (NEBNext® Library Quant Kit for Illumina®, protocol according to manufacturer's instructions). Blanks (where molecular grade water was added in the place of template) were used to check for any cross-contamination between samples. Pair-end sequencing was carried out on the MiSeq platform, using a V3 reagent kit (Illumina, Inc.). Samples were sequenced on two different MiSeq runs (run 1 = 36 individuals, run 2 = 26 individuals, plus six re-sequenced individuals from run 1 to test repeatability).

Bioinformatics and data processing

De-multiplexed raw pair end sequences were processed using the ampliSAT suite of tools (Sebastian et al., 2016) (available at: http://evobiolab.biol.amu.edu.pl/amplisat/). For full methodology see Methods S1. Any samples with less than 1000 reads coverage were discarded from downstream analysis. Parameters were optimised to ensure maximum reliability of genotyping for the entire dataset. Filtering per amplicon frequency was set at 1% (the minimum per amplicon frequency for variants that appeared in both replicates). The minimum dominant frequency was set at 10%, in order to keep true similar variants, whilst removing high frequency motive specific errors, meaning that only sequences with a frequency below that threshold were clustered with a parental sequence (Biedrzycka et al., 2017; Sebastian et al., 2016).

A maximum of 16 alleles per individual was considered, based on the *D. labrax* MHC class I alleles identified by (Pinto et al., 2013) from cDNA cloning, which indicated a minimum of six and maximum of eight loci per individual. The following manual filtering steps were then applied to the AmpliSAS output: all singletons (variants that appeared in one fish only) and any sequences with <10 reads overall were removed from downstream analysis to control for the presence of sequencing errors (Migalska et al., 2019). Repeatability (number of identical variants identified in both PCRs) in the six re-sequenced samples was then calculated following (Biedrzycka et al., 2017), where the number of identical variants are divided by the total number of alleles called in both replicates in a given individual to ascertain a repeatability proportion.

Polymorphism analysis

Initial screening of aligned variants was carried out in DnaSP (Rozas et al., 2017), using the 'polymorphism data' tool and the Codon-based Z-Test of Selection (MEGAX, Nei-Gojobori method). The seabass amino acid sequences were then aligned with other teleost MHC class I sequences of the U linage described in Grimholt et al. (2015), and a neighbour-joining tree based on *p*-distance and pairwise deletion was constructed using 1000 bootstrapping iterations to assess their relationship.

Recombination and selection analysis

Recombination was assessed using single break point analysis using the '010021' model of nucleotide substitution. Positive and negative selection were assessed using the HyPhy package (datamonkey.org) using three selection models. To detect pervasive positive/diversifying selection, FEL (fixed effects likelihood; Kosakovsky Pond & Frost, 2005) and FUBAR (Fast, Unconstrained Bayesian AppRoximation; Murrell et al., 2013) were used. To detect episodic positive/diversifying selection, MEME (mixed effects model of evolution; Pond et al., 2006) was used (for full methodology see Methods S1). The results of the three models were combined to identify positively selected sites for downstream analysis. A site was only considered to be under positive selection if it was selected by all three models. Recombination sites were compared with positively selected sites to avoid confounding effects, and, as they did not overlap, positively selected sites were analysed without removal of recombination sites.

In addition, alleles were analysed for conservative or radical amino-acid changes using TreeSAAP (Woolley et al., 2003), using a neighbour joining tree with 1000 bootstrap replicates (amino acid sequences). A sliding window value of 1 was used to obtain codon level magnitude of change scores. Significant differences in the top

two magnitude categories (7 and 8) were considered as 'radical' changes.

Supertype identification

The amino acids of positively selected sites as identified by FEL, FUBAR and MEME were assigned five z-descriptors (Doytchinova & Flower, 2005): z1 (hydrophobicity), z2 (steric bulk), z3 (polarity), z4 and z5 (electronic effects) (Sandberg et al., 1998) and translated into a mathematical matrix. Gaps in the alignment of positively selected sites were assigned zeros in the matrix.

Supertype clustering was performed in discriminant analysis of principle components (DAPC) with the 'adegenet' package (Jombart, 2008) in R (Version 4.0.4; R Core Team, 2019). Cluster values between the minimum number of clusters after which Bayesian information criterion (BIC) decreased by a negligible amount, and the lowest BIC value were then evaluated using stepwise DAPC runs, plotted, and inspected visually. The number of PCs was chosen using the function 'optim.a.score' and kept consistent for each DAPC analysis.

Loci identification via allele clustering

Phylogenetic trees were generated using the following methods: (1) neighbour joining (codons), using p-distance with 1000 bootstrap support (MEGAX, Kumar et al., 2018); (2) Ward's (Euclidian distance) using the z-matrix (Doytchinova & Flower, 2005); (3) UPGMA using the z-matrix, with three distances: Euclidian, Cosine and Pearson Correlation algorithms (PAST 4; Hammer et al., 2001). In addition, a Neighbor-Net (1000 bootstrap support) was generated in SplitsTree4 (Huson & Bryant, 2006), because networks may be more appropriate than trees for identifying relationships between alleles where duplication, recombination and gene conversion are common (Biedrzycka et al., 2017).

The performance of all methods in identifying loci was evaluated by testing whether any given individual exhibited more than two variants/allele per cluster. This was based on the assumption that, as a diploid organism (Felip et al., 2001), no individual should exhibit more than two alleles at a given locus.

Population analysis

Fish from the two regions with the largest sample size, Portugal (n = 19, ICES rectangle 4 E1) and Celtic Shelf (n = 22, ICES rectangle 28 E1), were examined to identify private alleles to each region and potential population structuring. To further explore population structure between these two populations, we used DAPC (Jombart, 2008; Jombart & Collins, 2015), using a presence/absence matrix where each column represented an allele, and each row an individual fish. Presence/absence of alleles was then recoded as frequencies, where alleles were represented as a proportion of the total alleles for each individual. This matrix was then used to calculate pairwise Jost's D (pairwise D_{est}, [Jost, 2008], 'mmod' package in R [Winter, 2012]). Jost's D is appropriate for genes that exist in multiple copies with unknown locus affiliation of alleles (Jost, 2008; Lighten et al., 2017). We used the function 'confusionMatrix' in the 'caret' package (Kuhn, 2009) in R, to assess the accuracy of the assignment of individuals to their stock based on DAPC.

RESULTS

MHC class I α polymorphism in seabass

A total of 62 individual fish were sequenced with sufficient depth for downstream analysis (>1000 reads). Samples were capped at a maximum of 5000 reads for AmpliSAT processing reasons and, after processing (prior to removal of singletons), had an average coverage of 3140 reads (STDEV = 1346 reads). Singleton removal reduced the number of variants/alleles from 332 to 133, and removal of two variants <200 bp in length resulted in 131 variants retained in the dataset for selection analysis. None of the variants was an exact match to already published alleles (Pinto et al., 2013); however, SBmhc1_111 has only one nucleotide difference to HQ290120.1/1-825 clone_15 and HQ290116.1/1-825 clone_1 (which are identical in the target region). Repeatability (number of identical variants identified in both PCRs) in the six re-sequenced samples was estimated as 0.89 (based on those samples as processed with the whole dataset). Individual copy number varied between seven and 16 alleles (mean = 12.6, mode = 13).

The 131 different nucleotide sequences identified corresponded to 123 unique amino acid sequences (Figure 1). An indel of two amino acids resulted in sequences of two lengths, 240 bp and 234 bp. From the 240 sites, 137 were variable and there were a total 228 mutations (Eta). Nucleotide diversity per site (Pi) was: 0.13216. The codon-based Z-test of selection indicated the presence of positive selection (p = 0.03, Z = 1.85).

Recombination and selection analysis

Strong support for recombination was found at position 109, inferred from both Akaike information criterion and BIC with 100% support. MEME analysis indicated that there was possible episodic positive/diversifying selection at 23 codon sites (Table S1, Figure 1). FEL analysis indicated that there was pervasive positive/diversifying selection at 19 sites and pervasive negative/purifying selection at six sites (Table S1). FUBAR found evidence
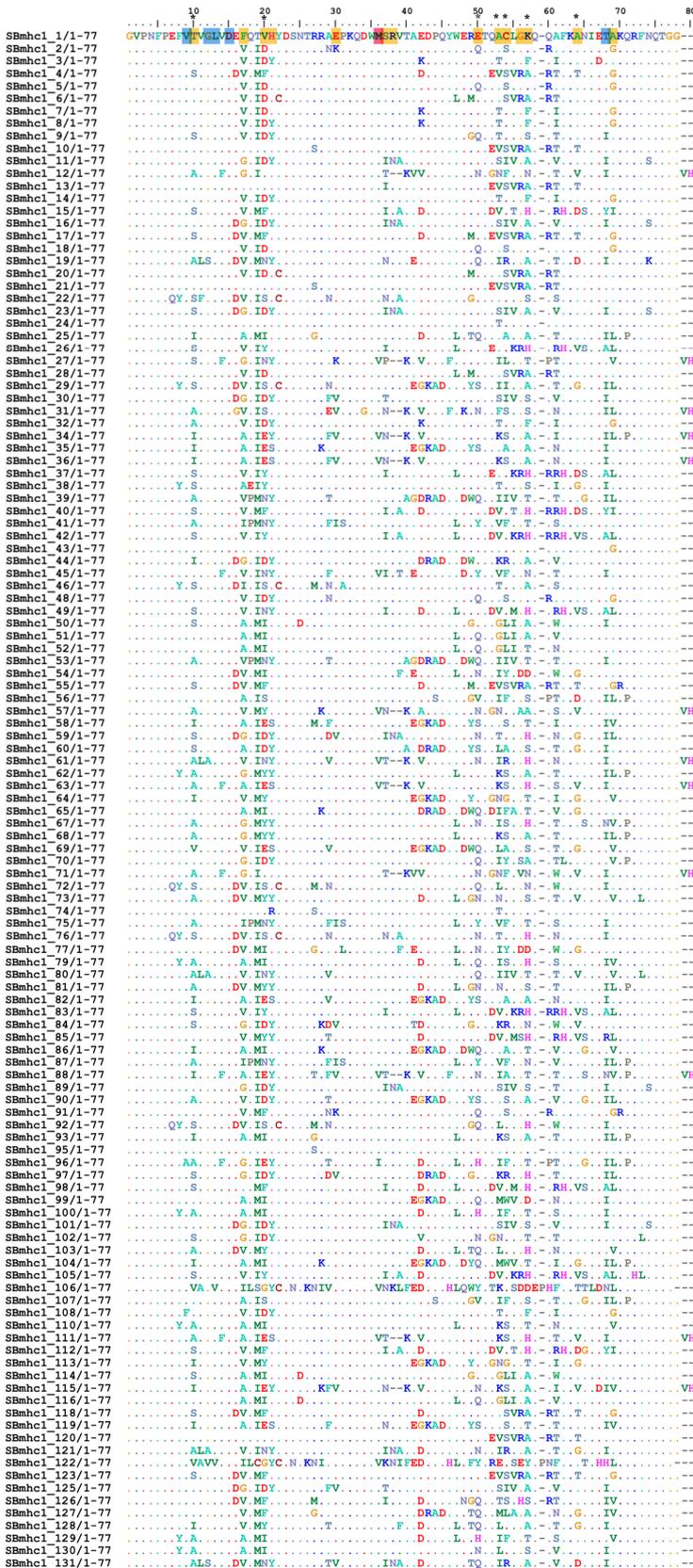
FIGURE 1 Codon alignment (Clustal W) of all alleles sequenced in this study. Yellow highlighting indicates pervasive/episodic positive/diversifying selection detected by mixed effects model of evolution (MEME), fixed effects likelihood (FEL), and Fast, Unconstrained Bayesian AppRoximation (FUBAR). Blue highlighting indicates negative episodic negative/purifying selection detected by both FEL and FUBAR. Pink highlighting indicates breakpoint (nucleotide 109) detected by single break point (all analyses performed using datamonkey.org tools). Asterisks represent codons which align to residues under selection at the peptide binding region of the human HLA-A2 gene (Grimholt et al., 2015)

of episodic positive/diversifying selection at 20 sites episodic negative/purifying selection at eight sites. In total, 14 sites were identified as being under positive selection by all three models (Table S1; Figure 1) and a further five

sites that were identified as under negative selection by both FUBAR and FEL (Table S1). Of the positively selected sites, seven coincided with residues contributing to the peptide binding pockets of the human HLA-A2 gene,

based on the alignment of the sequences with those from (Grimholt et al., 2015) (Figure 1).

Amino acid properties under selection (TreeSAAP)

TreeSAAP analysis identified significant deviations from neutral expectations (categories 7 and 8, Table S2, Figure 2), within the whole sequenced fragment, for the following physicochemical amino acid properties: 'power to be at the c-terminus' (energy potential of the c-terminus of an α helix to interact with other residues); 'power to be at the middle of the α-helix' (energy potential of the middle of an α helix to interact with other residues); 'mean r.m.s. fluctuation displacement' (the ability of a residue to change position in three-dimensional space); 'compressibility' (the contribution of a residue to the local density of protein secondary structures); 'solvent accessible reduction ratio' (the reduction ratio of cross peak intensity when residues are or are not irradiated with aliphatic protons); and 'thermodynamic transfer hydrophobicity' (the difference in the solubility of amino acids in water and ethanol;

Figure 2, Table S2). This indicates that selection pressure on these properties have driven changes at this region (Woolley et al., 2003).

Supertype and loci clustering

BIC values indicated that there were four main supertype clusters (Figure 3) before BIC values decreased by a negligible amount (Figure S1). We also explored a finer scale clustering, up to eight clusters, which represented the lowest BIC value (Figure 3). Assignment of potential alleles to individuals indicated that there could be between nine and 18 loci, depending on the method used (Figures S2, S4–S8; Table S4) but none of the clustering methods resulted in all individuals in the dataset having only two alleles in a cluster (Table S4), even when fine clustering (18 clusters) was applied. Seabass sequences clustered closer to class I-α lineage I than to the other lineages (Figure S3). Two sequences, SBmhc1_106 and SBmhc1_122, clustered with lineage I sequences from salmonids (Atlantic salmon, brown trout, and rainbow trout), and close to Medaka sequences, also from lineage I (Figure S3).
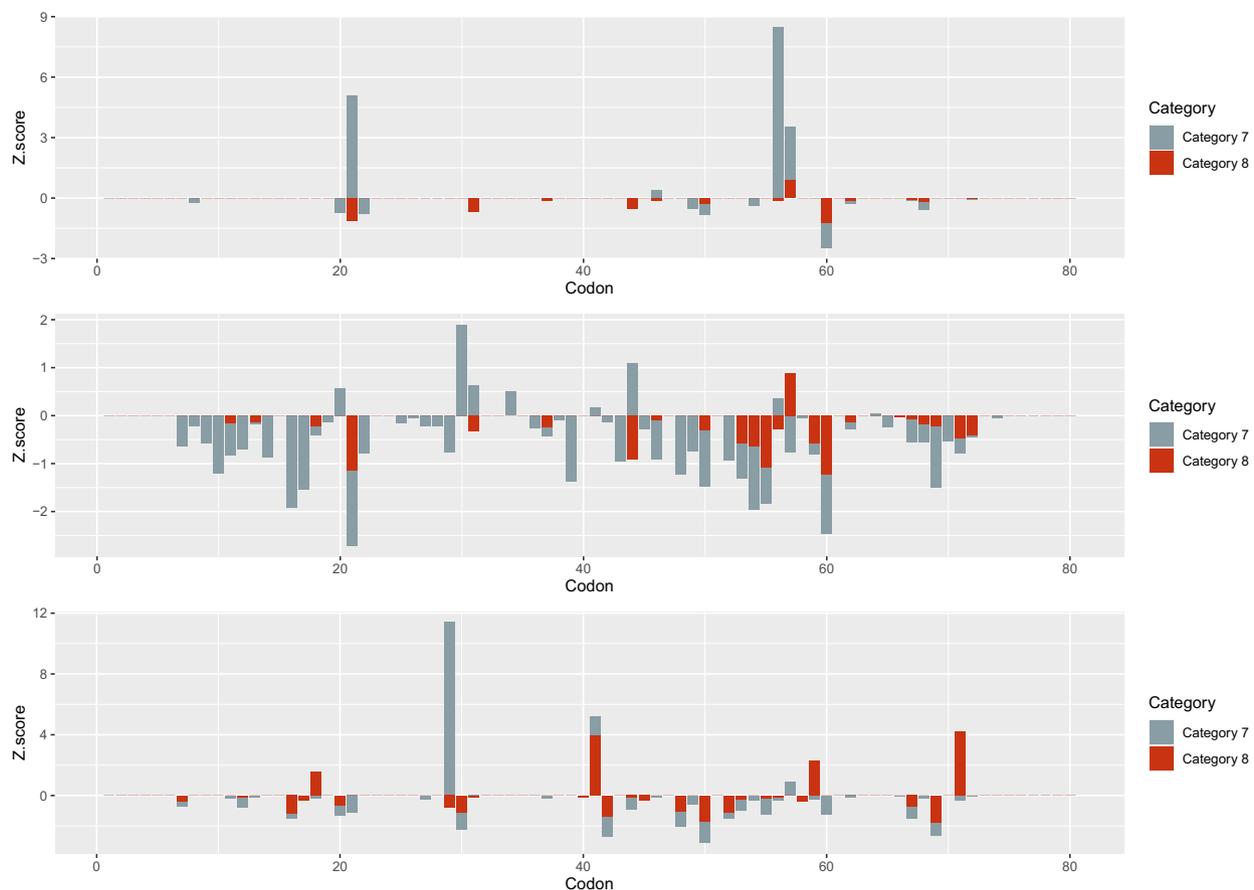

FIGURE 2 Z-scores relating to phenotypic amino acid properties/traits (Power to be at the middle of the α-helix, power to be at the c-terminus and compressibility). Categories represent magnitude of amino acid property change at nonsynonymous residues on a site-by-site basis. Class 8 represents changes of the highest magnitude, followed by class 7. Low and 0 z score values represent areas of conservation in terms of amino acid properties, in contrast to high values, which represent areas of selection
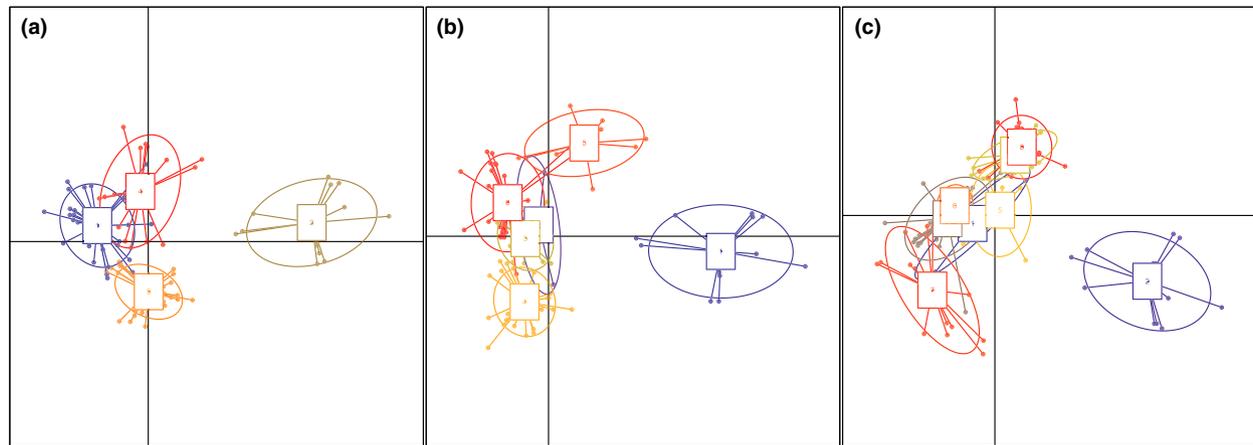
FIGURE 3 (a–c) Clustering of major histocompatibility complex alleles in *Dicentrarchus labrax*, based on the physicochemical properties of amino acids (Z scores) of positively selected sites, using discriminant analysis of principal components. Each point represents the positioning of each major histocompatibility complex allele within corresponding number of discriminant functions (Eigenvalues insets) and the circles represent the supertypes

Population analysis

We found 30 unique alleles (representing 29 unique amino acid sequences) in the Celtic Shelf population that were not represented in the Portuguese population and 22 alleles (21 unique amino acid sequences) unique to the Portuguese population (Figure 4a), including the two alleles with an insertion at codon 59 in the alignment (SBmhc1_106 and SBmhc1_122). Private alleles were present across the phylogenetic tree, but Celtic shelf alleles were more frequent in recently diverged clusters than those from Portugal (Figure 4a). Most unique alleles appeared in only one or two individuals in the respective populations (Table S4), with the exception of allele SBmhc1_25 and SBmhc1_38, which appeared in six and three Portuguese individuals respectively, and SBmhc1_54, present in three Celtic Shelf individuals. There was no difference in the number of alleles per fish between the two populations (mean of Celtic population = 12.64 alleles per individual, mean of Portuguese population = 12.79 alleles per individual; Welch two-sample t-test $t = -0.26442$, df = 38.999, $p = 0.7928$). DAPC analysis indicated low differentiation in allele frequencies as a whole (Figure 4b) between the two regions ($D_{est} = 0.05$), with most of the variance described by the first principal component. Based on all the alleles, we were able to assign 75% of the fish (31/41) to their region of origin (Celtic Shelf or Portugal). The accuracy of the assignment was significantly greater than chance (aAccuracy = 0.756; 95% CI: 0.597, 0.8764; $p = 0.003$).

DISCUSSION

The high variability of MHC genes is known to be driven by natural and sexual selection and, for this reason, they represent good candidates to assess fine-scale structuring potentially derived from local selective pressures

(Larson et al., 2019). We found that the MHC class I-α gene in the Atlantic seabass displays high allelic diversity, evidence for selection and, importantly, potential for detecting fine scale structuring within stocks largely considered panmictic (ICES, 2021). Our study represents the first population study of MHC class I in seabass and indicates that, as for other fishes, the *D. labrax* MHC class I-α locus has undergone strong positive selection and recombination, while maintaining intermediate copy numbers and relatively few supertypes. This is similar to the diversity of MHC class II in guppies, where supertypes are maintained by balancing selection and consist of groups of alleles with similar functionality that are subject to positive selection (Lighten et al., 2017).

Most seabass MHC-I sequences clustered close to the lineage I of the U-type classical alleles from different fish species, two of them grouping with salmonid alleles (Grimholt et al., 2015). Close clustering between seabass and salmonid alleles might indicate trans-species polymorphism (the occurrence of similar alleles in related species) typical of MHC genes (Klein et al., 1998), but this must be interpreted with caution as it is based on the α-1 domain only and the bootstrapping support was relatively low. We found an average of 13 copies per individual in *D. labrax*, which is similar to other teleost species with U lineage genes, such as the African moonfish, *Selene dorsalis* and Sand roller *Percopsis transmontana* (Malmstrøm et al., 2016). In general, the teleost MHC class I genes are highly polymorphic, with copy number varying between families and species (Malmstrøm et al., 2016). For instance, Atlantic salmon (*Salmo salar*) has a single MHC I locus (Grimholt et al., 2003), while three-spined stickleback, *Gasterosteus aculeatus* possess between three and nine alleles per individual, which are difficult to assign to loci probably due to high allele similarity (Aeschlimann et al., 2003). In contrast, most gadoids have lost classical MHC II genes and possess a largely expanded number of MHC I copies; for example,
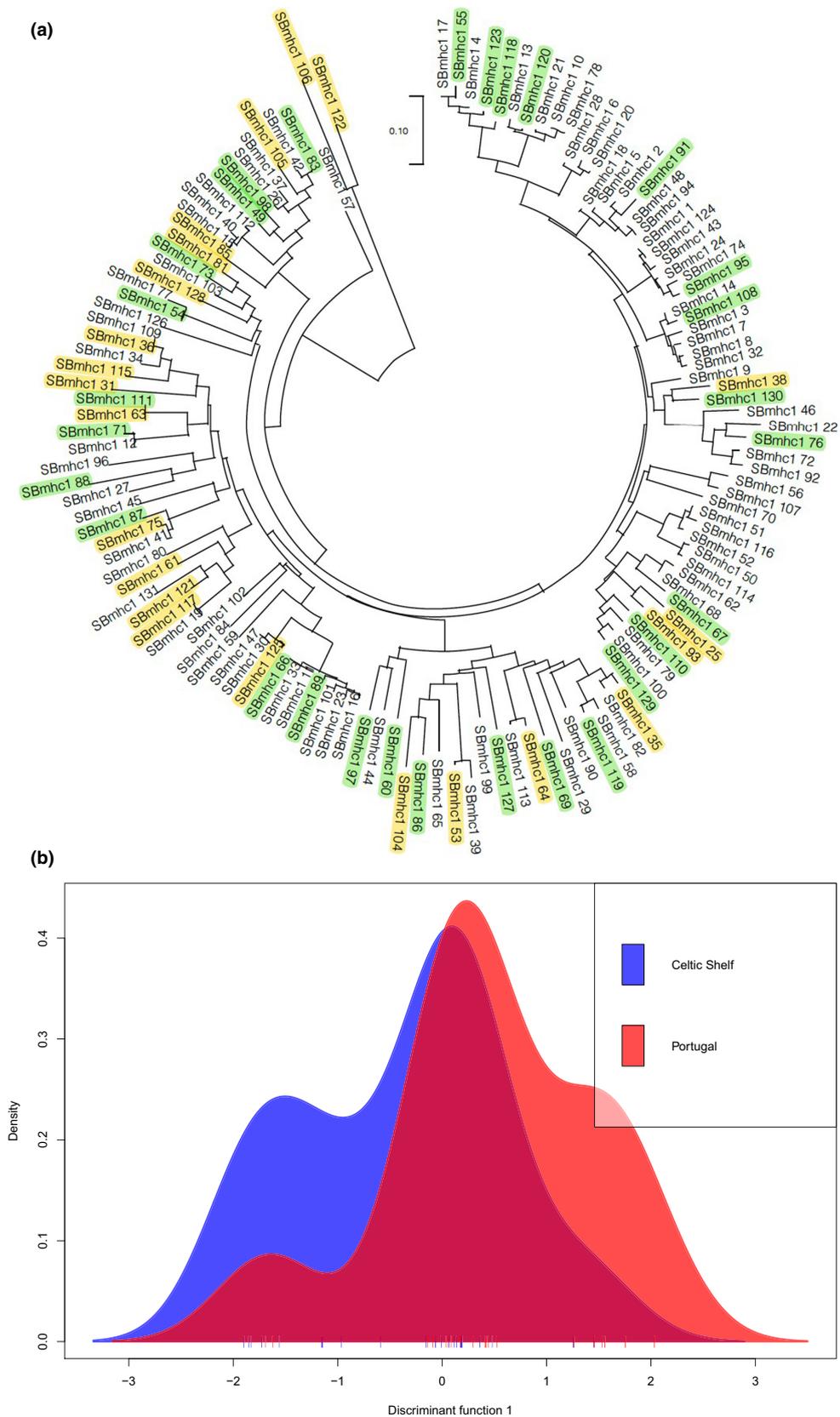
FIGURE 4 (a) Neighbour-joining tree phylogenetic tree of the *Dicentrarchus labrax* major histocompatibility complex class I-α alleles identified in this study. Celtic Shelf population alleles highlighted in green and Portugal population alleles highlighted in yellow. (b) Individual densities of allele frequencies (presence/absence for each individual) for the Portuguese and Celtic Shelf populations, plotted against the first discriminant component used in analysis of principal components (DAPC)

Atlantic cod (*Gadus morhua*) which has approximately 100 (Star et al., 2011). The observed diversity in *D. labrax* was, therefore, within the range expected for a species with a classical MHC II and fitted well with the previously predicted six to more than eight MHC-I loci for the species (Pinto et al., 2013), although none of them were an exact match to those previously published, likely due to the different origin of the populations and the high variability of the MHC genes. However, as our analysis was based on DNA sequences and not on expressed alleles (mRNA), we could not identify the exact number of loci and it is possible that some of the sequences could represent pseudogenes.

Positive selection was identified at 14 polymorphic sites across the 80 codons in this study, some of them coinciding with PBR sites previously defined in fish based on the alignment with human sequences (Grimholt et al., 2015) (Figure S3). This level of selection is comparable to other fishes (Wegner, 2008) for example sockeye salmon (*Oncorhynchus nerka*) where four codons are under positive selection in a 32-codon fragment (McClelland et al., 2011). High rates of positive selection are almost always observed in MHC alleles, driven by parasite-mediated balancing selection or sexual selection (Bernatchez & Landry, 2003). In addition, recombination can also act as a significant force to rapidly increase diversity in MHC alleles (Consuegra et al., 2005), and we found strong support for a breakpoint at position 109 bp, possibly contributing to the observed allelic diversity.

While codon-based selection models give an indication of the areas of the gene under selection, the ability to bind antigens from specific pathogens is determined by changes in the shape of the cleft of the PBR (Dionne et al., 2007). Nucleotide differences, or even amino acid differences, may not directly translate to protein binding differences, and additional alleles may not necessarily confer the ability to respond to a greater range of pathogens, unless the changes in the protein also change the binding properties of that allele (Ellison et al., 2012; Lighten et al., 2017). The analysis of functional differences based on binding properties analysis indicated that the energy potential of the C-terminus, the position of amino acids at the middle of an α helix (where they can interact with other residues), and the ability for a residue to change position in three-dimensional space were under selection, supporting the role of selective pressures (Doytchinova & Flower, 2005; Lighten et al., 2017). Alleles within a supertype are predicted to bind similar antigenic 'supermotifs' (Phillips et al., 2018). We identified between four and eight clusters, although the precise number of clusters is not clear. Clustering can be challenging when loci share identical alleles, there are null alleles or, as found here, copy number varies between haplotypes (Huang et al., 2019). Even with an upper estimate of eight clusters, we found relatively low supertype diversity in comparison to other fishes, despite a relatively diverse allele pool of 131 alleles. For instance,

in guppies, 66 alleles were clustered to 13 supertypes (Smallbone et al., 2021). This could be the result of convergent evolution resulting in a relatively small number of overlapping peptide-binding motifs, as for the highly diverse human class I alleles that may be clustered into as few as nine supertypes (Sidney et al., 2008).

Despite loci remaining undefined, we observed allelic differentiation in the two seabass populations we compared, confirmed by DAPC analysis, although private alleles were detected in relatively few individuals within each region. Overall, the differentiation between regions was small but, based on the MHC sequences identified here, we were able to assign 75% of the fish to their region of origin (Celtic Shelf or Portugal) with high confidence. Previous studies of Eastern Atlantic seabass using microsatellite and SNP analysis failed to identify consistent population structuring (Souche et al., 2015), and tagging studies have also shown that *D. labrax* are capable of swimming large distances, e.g. 1200 km within 2 months of tags being deployed, and migrate considerable distances offshore to spawn (Pawson et al., 2007). It had therefore been assumed that Eastern Atlantic seabass formed one single panmictic population. Seabass seems to display a shallow population structuring based on neutral markers but, as for other marine species (Milano et al., 2014), a finer level of structuring resulting from local selective pressures can be detected when markers under selection are used. The difference in private alleles between the Celtic shelf and Portuguese uncovered population structuring not observed using neutral markers (Souche et al., 2015), and indicates the potential for MHC-I to detect fine-scale population variation. Markers in the MHC region can be particularly good for showing fine scale differentiation, as changes can accumulate relatively quickly and are affected by local selection (Consuegra et al., 2005), and genome-wide high-throughput methods (Griot et al., 2021; Penaloza et al., 2021) can help identify additional markers under selection (Carreras et al., 2017). Our study provides the first detailed analysis of MHC class I in seabass, a species that supports important commercial and sport fisheries, as well as aquaculture. Our approach provides a quick tool to screen the highly variable MHC region of Atlantic seabass for fisheries management purposes, that can be adapted to other marine species. Moreover, given the increasing importance of seabass aquaculture (Vandeputte et al., 2019), and the potential risk that escapes pose to wild populations (Arechavala-Lopez et al., 2018), the same approach employed here could be used for genotyping farmed species, potentially for selective breeding for disease resistance (Pawluk et al., 2019) and/or to monitor the incidence of escapees from fish farms (Monzón-Argüello et al., 2013).

Environmental complexity has the potential to create refugia for marine species that may result in local adaptation and create cryptic population structuring, essential for the long-term persistence of exploited stocks

(Midway et al., 2018). In many marine species with high fecundity and dispersal rates, this fine population structure can only be identified by markers under natural selection (André et al., 2011; Jorde et al., 2018; Lamichhaney et al., 2012). Based on our initial results from sea bass, we propose that MHC markers can be used for the management of marine species with cryptic population structure, for which preserving their fine population structuring is needed to maintain their functional biodiversity (Limborg et al., 2012).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

ETHICAL APPROVAL

All work was carried out with approval from Swansea University Ethics Committee (Reference number SU-Ethics-Student-280621/4326).

DATA AVAILABILITY STATEMENT

Sequences have been submitted to GenBank and are accessible under Accession numbers MZ466411-MZ466541.

ORCID

Sofia Consuegra https://orcid.
org/0000-0003-4403-2509

REFERENCES

Aeschlimann, P.B., Haberli, M.A., Reusch, T.B.H., Boehm, T. & Milinski, M. (2003) Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. *Behavioral Ecology and Sociobiology*, 54, 119–126.

André, C., Larsson, L.C., Laikre, L., Bekkevold, D., Brigham, J., Carvalho, G.R. et al. (2011) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity*, 106, 270–280.

Arechavala-Lopez, P., Toledo-Guedes, K., Izquierdo-Gomez, D., Šegvić-Bubić, T. & Sanchez-Jerez, P. (2018) Implications of sea bream and sea bass escapes for sustainable aquaculture management: a review of interactions, risks and consequences. *Reviews in Fisheries Science & Aquaculture*, 26, 214–234.

Beddington, J.R., Agnew, D.J. & Clark, C.W. (2007) Current problems in the management of marine fisheries. *Science*, 316, 1713–1716.

Bento, E.G., Grilo, T.F., Nyitrai, D., Dolbeth, M., Pardal, M.Â. & Martinho, F. (2016) Climate influence on juvenile European sea bass (*Dicentrarchus labrax*, L.) populations in an estuarine nursery: a decadal overview. *Marine Environmental Research*, 122, 93–104.

Bernatchez, L. & Landry, C. (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16, 363–377.

Biedrzycka, A., O'Connor, E., Sebastian, A., Migalska, M., Radwan, J., Zając, T. et al. (2017) Extreme MHC class I diversity in the sedge warbler (Acrocephalus schoenobaenus); selection patterns and allelic divergence suggest that different genes have different functions. *BMC Evolutionary Biology*, 17, 1–12.

Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E. et al. (2017) Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Scientific Reports*, 7, 1–12.

Castilho, R. & McAndrew, B. (1998) Population structure of seabass in Portugal: evidence from allozymes. *Journal of Fish Biology*, 53, 1038–1049.

Consuegra, S. & Garcia de Leaniz, C. (2008) MHC-mediated mate choice increases parasite resistance in salmon. *Proceedings of the Royal Society B: Biological Sciences*, 275, 1397–1403.

Consuegra, S., Megens, H.J., Schaschl, H., Leon, K., Stet, R.J.M. & Jordan, W.C. (2005) Rapid evolution of the MH class I locus results in different allelic compositions in recently diverged populations of Atlantic salmon. *Molecular Biology and Evolution*, 22, 1095–1106.

Coscia, I. & Mariani, S. (2011) Phylogeography and population structure of European sea bass in the north-east Atlantic. *Biological Journal of the Linnean Society*, 104, 364–377.

De Pontual, H., Lalire, M., Fablet, R., Laspougeas, C., Garren, F., Martin, S. et al. (2019) New insights into behavioural ecology of European seabass off the West Coast of France: implications at local and population scales. *ICES Journal of Marine Science*, 76, 501–515.

Dionne, M., Miller, K.M., Dodson, J.J., Caron, F. & Bernatchez, L. (2007) Clinal variation in MHC diversity with temperature: evidence for the role of host–pathogen interaction on local adaptation in Atlantic salmon. *Evolution*, 61, 2154–2164.

Doytchinova, I.A. & Flower, D.R. (2005) In silico identification of supertypes for class II MHCs. *Journal of Immunology*, 174, 7085–7095.

Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F. & Gagnaire, P.-A. (2018) The origin and remolding of genomic islands of differentiation in the European sea bass. *Nature Communications*, 9, 1–11.

Edwards, S.V. & Hedrick, P.W. (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution*, 13, 305–311.

Eizaguirre, C., Lenz, T.L., Kalbe, M. & Milinski, M. (2012) Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature Communications*, 3, 621.

Ellison, A., Allainguillaume, J., Girdwood, S., Pachebat, J., Peat, K.M., Wright, P. et al. (2012) Maintaining functional major histocompatibility complex diversity under inbreeding: the case of a selfing vertebrate. *Proceedings of the Royal Society B: Biological Sciences*, 279, 5004–5013.

European Parliament, Directorate-General for Internal Policies of the Union, Biseau, A., Le Goff, R. & Drogou, M. (2016). European sea bass stocks and management measures within the European Union, European Parliament. https://data.europa.eu/doi/10.2861/70576

Felip, A., Piferrer, F., Carrillo, M. & Zanuy, S. (2001) Comparison of the gonadal development and plasma levels of sex steroid hormones in diploid and triploid sea bass, *Dicentrarchus labrax* L. *Journal of Experimental Zoology*, 290, 384–395.

Gagnaire, P.A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F. et al. (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, 8, 769–786.

Gandra, M., Assis, J., Martins, M.R. & Abecasis, D. (2021) Reduced global genetic differentiation of exploited marine fish species. *Molecular Biology and Evolution*, 38, 1402–1412.

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A. et al. (2003) MHC polymorphism and disease

resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics*, 55, 210–219.

Grimholt, U., Tsukamoto, K., Azuma, T., Leong, J., Koop, B.F. & Dijkstra, J.M. (2015) A comprehensive analysis of teleost MHC class I sequences. *BMC Evolutionary Biology*, 15, 1–17.

Griot, R., Allal, F., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A. et al. (2021) 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*, 530, 735930.

Hammer, O., Harper, D. & Ryan, P. (2001) PAST: Paleontological statistics software package for education and data analysis. *Paleontologia Electronica*, 4, 9.

Hedgecock, D., Barber, P.H. & Edmands, S. (2007) Genetic approaches to measuring connectivity. *Oceanography*, 20, 70–79.

Hedrick, P. & Kim, T. (2000) *Genetics of complex polymorphisms: parasites and maintenance of MHC variation*. Cambridge, MA: Harvard University Press.

Huang, K., Zhang, P., Dunn, D.W., Wang, T., Mi, R. & Li, B. (2019) Assigning alleles to different loci in amplifications of duplicated loci. *Molecular Ecology Resources*, 19, 1240–1253.

Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267.

ICES (2021) Working group for the Celtic Seas Ecoregion (WGCSE). In: Lundy, M. & Nimmegeers, S. (Eds.). *ICES scientific reports*. p. 1082. https://doi.org/10.17895/ices.pub.8139

Janeway, C.A., Travers, P., Walport, D. & Shlomchik, M.J. (2004) *Immunobiology: The immune system in health and disease*. Garland Science.

Jombart, T. (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.

Jombart, T. & Collins, C. (2015) *A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0*. London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling.

Jorde, P.E., Synnes, A.E., Espeland, S.H., Sodeland, M. & Knutsen, H. (2018) Can we rely on selected genetic markers for population identification? Evidence from coastal Atlantic cod. *Ecology and Evolution*, 8, 12547–12558.

Jost, L. (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, 17, 4015–4026.

Klein, J., Sato, A., Nagl, S. & O'hUigín, C. (1998) Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics*, 29, 1–21.

Kosakovsky, P.S. & Frost, S. (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, 22, 1208–1222.

Kuhn, M. (2009) The caret package. *Journal of Statistical Software*, 28.

Lamichhaney, S., Barrio, A.M., Rafati, N., Sundstrom, G., Rubin, C.-J., Gilbert, E.R. et al. (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, 109, 19345–19350.

Larson, W.A., Dann, T.H., Limborg, M.T., McKinney, G.J., Seeb, J.E. & Seeb, L.W. (2019) Parallel signatures of selection at genomic islands of divergence and the major histocompatibility complex in ecotypes of sockeye salmon across Alaska. *Molecular Ecology*, 28, 2254–2271.

Lemaire, C., Versini, J.J. & Bonhomme, F. (2005) Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal of Evolutionary Biology*, 18, 70–80.

Lighten, J., Papadopulos, A.S., Mohammed, R.S., Ward, B.J., Paterson, I.G., Baillie, L. et al. (2017) Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. *Nature Communications*, 8, 1–11.

Limborg, M.T., Helyar, S.J., De Bruyn, M., Taylor, M.I., Nielsen, E.E., Ogden, R. et al. (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, 21, 3686–3703.

Malmstrøm, M., Matschiner, M., Tørresen, O.K., Star, B., Snipen, L.G., Hansen, T.F. et al. (2016) Evolution of the immune system influences speciation rates in teleost fishes. *Nature Genetics*, 48, 1204–1210.

McClelland, E.K., Ming, T.J., Tabata, A. & Miller, K.M. (2011) Sequence analysis of MHC class I α2 from sockeye salmon (*Oncorhynchus nerka*). *Fish & Shellfish Immunology*, 31, 507–510.

Midway, S.R., White, J.W. & Scharf, F.S. (2018) The potential for cryptic population structure to sustain a heavily exploited marine flatfish stock. *Marine and Coastal Fisheries*, 10, 411–423.

Migalska, M., Sebastian, A. & Radwan, J. (2019) Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proceedings of the National Academy of Sciences*, 116, 5021–5026.

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G.R. et al. (2014) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, 23, 118–135.

Milinski, M. (2006) The major histocompatibility comlex, sexual selection, and mate choice. *Annual Review of Ecology, Evolution, and Systematics*, 37, 159–186.

Monzón-Argüello, C., de Leaniz, C.G., Gajardo, G. & Consuegra, S. (2013) Less can be more: loss of MHC functional diversity can reflect adaptation to novel conditions during fish invasions. *Ecology and Evolution*, 3, 3359–3368.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L. et al. (2013) FUBAR: a fast, unconstrained Bayesian AppRoximation for inferring selection. *Molecular Biology and Evolution*, 30(5), 1196–1205.

Pawluk, R.J., Garcia de Leaniz, C. & Consuegra, S. (2019) Sea lice loads correlate with the diversity at the Major Histocompatibility Complex-related loci in farmed Atlantic salmon, Salmo salar. *Journal of Fish Diseases*, 42, 1091.

Pawson, M., Brown, M., Leballeur, J. & Pickett, G. (2008) Will philopatry in sea bass, *Dicentrarchus labrax*, facilitate the use of catch-restricted areas for management of recreational fisheries? *Fisheries Research*, 93, 240–243.

Pawson, M., Pickett, G., Leballeur, J., Brown, M. & Fritsch, M. (2007) Migrations, fishery interactions, and management units of sea bass (*Dicentrarchus labrax*) in Northwest Europe. *Journal of ICES Marine Science*, 64, 332–345.

Penaloza, C., Manousaki, T., Franch, R., Tsakogiannis, A., Sonesson, A.K., Aslam, M. et al. (2021) Development and testing of a combined species SNP array for the European seabass (*Dicentrarchus labrax*) and gilthead seabream (*Sparus aurata*). *Genomics*, 113, 2096–2107.

Phillips, K.P., Cable, J., Mohammed, R.S., Herdegen-Radwan, M., Raubic, J., Przemsycka, K.J. et al. (2018) Immunogenetic novelty confers a selective advantage in host–pathogen coevolution. *Proceedings of the National Academy of Sciences*, 115, 1552–1557.

Pinto, R.D., Randelli, E., Buonocore, F., Pereira, P.J. & dos Santos, N.M. (2013) Molecular cloning and characterization of sea bass (*Dicentrarchus labrax* L.) MHC class I heavy chain and β2-microglobulin. *Developmental & Comparative Immunology*, 39, 234–254.

Pond, S.L.K., Posada, D., Gravenor, M.B., Woelk, C.H. & Frost, S.D.W. (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution*, 23, 1891–1901.

Quéré, N., Guinand, B., Kuhl, H., Reinhardt, R., Bonhomme, F. & Desmarais, E. (2010) Genomic sequences and genetic differentiation at associated tandem repeat markers in growth hormone,

somatolactin and insulin-like growth factor-1 genes of the sea bass, *Dicentrarchus labrax*. *Aquatic Living Resources*, 23, 285–296.

R Core Team (2019) *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Radwan, J., Babik, W., Kaufman, J., Lenz, T.L. & Winternitz, J. (2020) Advances in the evolutionary understanding of MHC polymorphism. *Trends in Genetics*, 36, 298–311.

Reiss, H., Hoarau, G., Dickey-Collas, M. & Wolff, W.J. (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, 10, 361–395.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E. et al. (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34, 3299–3302.

Sebastian, A., Herdegen, M., Migalska, M. & Radwan, J. (2016) amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*, 16, 498–510.

Sidney, J., Peters, B., Frahm, N., Brander, C. & Sette, A. (2008) HLA class I supertypes: a revised and updated classification. *BMC Immunology*, 9, 1–15.

Slade, R.W. & McCallum, H.I. (1992) Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, 132, 861–864.

Smallbone, W., Ellison, A., Poulton, S., van Oosterhout, C. & Cable, J. (2021) Depletion of MHC supertype during domestication can compromise immunocompetence. *Molecular Ecology*, 30, 736–746.

Souche, E.L., Hellemans, B., Babbucci, M., MacAoidh, E., Guinand, B., Bargelloni, L. et al. (2015) Range-wide population structure of European sea bass *Dicentrarchus labrax*. *Biological Journal of the Linnean Society*, 116, 86–105.

Star, B., Nederbragt, A.J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T.F. et al. (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, 477, 207–210.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T. et al. (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, 5, 1–10.

Vandeputte, M., Gagnaire, P.A. & Allal, F. (2019) The European sea bass: a key marine fish model in the wild and in aquaculture. *Animal Genetics*, 50, 195–206.

Walker, N.D., Boyd, R., Watson, J., Kotz, M., Radford, Z., Readdy, L. et al. (2020) A spatially explicit individual-based model to support management of commercial and recreational fisheries for European sea bass *Dicentrarchus labrax*. *Ecological Modelling*, 431, 109179.

Wegner, K. (2008) Historical and contemporary selection of teleost MHC genes: did we leave the past behind? *Journal of Fish Biology*, 73, 2110–2132.

Winter, D.J. (2012) MMOD: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, 12, 1158–1160.

Woolley, S., Johnson, J., Smith, M., Crandall, K. & McClellan, D. (2003) TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics*, 19, 671–672.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.