Molecular Phylogenetics and Evolution 224 (2026) 108694

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Global genome-wide patterns of genetic diversity and population structure in *Coryphaena hippurus*

Píndaro Díaz Jaimes [a,*], Verónica Mendoza-Portillo [a], Silvia Hinojosa-Alvarez [b], Erika Magallón-Gayón [b], Rocío Alejandra Chavez-Santoscoy [b], Jesús Hernández-Pérez [b], Adán Fernando Mar-Silva [a], Maried Ochoa-Zavala [c], Sofia Ortega-Garcia [d], Oscar Lao [e], Marco Arculeo [f]

[a] *Unidad Académica de Ecología y Biodiversidad Acuática, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México, Circuito Interior s/n, Ciudad Universitaria, Ciudad de México C.P. 04510, Mexico*
[b] *Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Campus Monterrey, Ave. Eugenio Garza Sada 2501 Sur, Monterrey 64849, Mexico*
[c] *Laboratorio de Genética Evolutiva de Organismos Tropicales, Escuela Nacional de Estudios Superiores Unidad Morelia, Universidad Nacional Autónoma de México. Morelia, 14 Michoacán, Mexico*
[d] *Departamento de Pesquerías y Biología Marina. Instituto Politécnico Nacional-Centro Interdisciplinario de Ciencias Marinas, La Paz, Baja California Sur, Mexico*
[e] *Group of Algorithms for Population Genomics, Department of Genetics, Institut de Biologia Evolutiva, IBE, (CSIC-Universitat Pompeu Fabra), CMIMA-Pssg. de la Barceloneta 37-49. 08003, Barcelona, Spain*
[f] *University of Palermo, Department of Biological, Chemical and Pharmaceutical Sciences and Technologies (STEBICEF), Via Archirafi 18, 90123 Palermo, Italy*

ARTICLE INFO

Keywords:
Dolphinfish
Population genomics
Evolutionary genomics
Genome-wide

ABSTRACT

The origin and maintenance of marine biodiversity remain poorly understood, particularly in highly connected oceanic environments where extensive dispersal is expected to limit population differentiation. Evolutionary processes driving population divergence or speciation frequently depend on a proper knowledge of species' life history and its interaction with major environmental variables. Using a genome-wide approach, we investigated the global population structure of the cosmopolitan pelagic fish *Coryphaena hippurus* Linnaeus, 1758, generating a dataset of 8.7 million SNPs and complete mitochondrial genomes.

Our results reveal, for the first time, four genetically differentiated populations, corresponding to major oceanic basins: Atlantic, Pacific, Indian and Mediterranean Sea. The Mediterranean population exhibited a distinct genomic signature, likely resulting from historical isolation and restricted contemporary exchange through the strait of Gibraltar. Despite this structure, genome-wide analyses uncovered extensive signals of historical connectivity among basins. *D*-statistics and *f4-ratio* tests detected significant excess allele sharing, particularly between Indo-Pacific and Atlantic populations, while phylogenetic network reconstruction in TreeMix supported multiple migration events, highlighting the role of ancestral gene flow in shaping global genetic patterns.

At finer scales, a limited but significant intra-oceanic structure was detected within both Atlantic and Pacific basins. Seascape genomic analyses revealed that environmental gradients such as salinity, phosphate concentration, light availability (PAR, Kd), and pH are significantly associated with genomic variation, suggesting that ecological factors contribute to population differentiation across heterogeneous marine environments.

1. Introduction

Evolutionary studies for widely distributed pelagic species are challenging due to life history traits such as high reproductive output, fast growth and short generation times, which usually result in high genetic diversity and large effective population sizes. These characteristics can obscure signals of genetic divergence in the marine realm. Genomic approaches have been proposed to overcome limitations associated with sampling large numbers of individuals and capturing sufficient genome-wide variation. As a result, diverse protocols have been developed, including reduced-representation sequencing methods such as RADseq (Davey & Blaxter 2010; Peterson

* Corresponding author.
E-mail address: pindaro@cmarl.unam.mx (P.D. Jaimes).

https://doi.org/10.1016/j.ympev.2026.108694
Received 29 November 2025; Received in revised form 13 July 2026; Accepted 16 July 2026
Available online 18 July 2026

et al. 2012; Bayona-Vásquez et al. 2019).

In several widely distributed pelagic species, including yellowfin tuna *Thunnus albacares* (Grewe et al. 2015; Pecoraro et al. 2018), the wahoo *Acanthocybium solandri* (Haro-Bilbao et al. 2021), the sailfish *Istiophorus platypterus* (da Silva Ferrette et al. 2023) exhibiting high levels of diversity and effective population sizes, RADseq has been successfully used to resolve global and regional population structure using thousands of SNPs. However, for other pelagic species such as the dolphinfish (*Cor. hippurus)*, no global-scale genomic studies are currently available, and local-scale analyses based on reduced-representation sequencing have shown limited power to resolve clear genetic patterns (Maroso et al. 2016; Mar-Silva et al. 2023).

Recently, low coverage genome sequencing has become a valuable alternative to resolve population genetic structure at both global and local scales, due to the large number of variants that can be obtained to detect evolutionary processes driving population divergence. Advances in sequencing technologies have enabled population-level genomic studies at reduced cost through low-pass whole genome sequencing (lpWGS), which typically generates ~ 1–5 × coverage per sample (Lou et al. 2021). This approach increases the number of locations sampled across a speciés distribution while maintaining sufficient power to detect population structure and signals of adaptive variation. Such genome-wide SNP coverage is particularly relevant for widely distributed marine species with large effective population sizes, such as *C. hippurus*, where subtle genetic differentiation may otherwise remain undetected.

The dolphinfish (*C. hippurus*), is a circumtropical pelagic species distributed in tropical and subtropical regions (Oxenford 1999). It is a fast-growing and short-lived fish characterized by high reproductive potential through early sexual maturity (at approximately 4–6 months) (Lasso & Zapata 1999; Oxenford 1999) These features result in high levels of gene diversity for this species which is relevant since the evolutionary perspective as it influences the divergence processes by delaying the accumulation of enough genetic variants for being detected by traditional molecular markers.

Genetic studies have highlighted the difficulties for detecting differences in *C. hippurus* at global or local scales, even between ocean basins exhibiting evident physical barriers to gene flow (Díaz-Jaimes et al. 2010). Global mitochondrial DNA studies failed to detect inter-oceanic genetic differences among oceanic basins, although a clear divergent lineage in the Mediterranean was reported (Díaz-Jaimes et al. 2010). However, genetic differences based on nuclear and/or adaptative loci at local scale, have been reported for studies in the Mediterranean (Maroso et al. 2016, Mendoza-Portillo et al. 2026), and the Eastern Pacific (Ochoa-Zavala et al. 2022; Mar-Silva et al. 2023). These findings suggest the possible presence of cryptic population structure at finer geographic scales, potentially driven by oceanographic barriers, spawning site fidelity, or selection across environmental gradients, highlighting the complexity of population dynamics in this highly mobile species.

Despite significant progress in defining the genetic patterns in *C. hippurus*, important knowledge gaps remain; as previous studies have been geographically restricted, limited to reduced marker sets, or focused exclusively on mitochondrial variation. Addressing these limitations requires integrative approaches that jointly leverage nuclear and mitochondrial genomes. Low-coverage genomes provide a cost-effective framework to resolve population structure, contemporary gene flow and infer demographic history, while complete mitogenomes provide insights into the maternal lineage of a species and long-term connectivity of its populations.

In this study, we analyze globally distributed individuals of *C.hippurus* using both low-coverage whole genome sequencing and complete mitogenomes. Specifically, we aim to 1) characterize global population genetic structure across the species distribution using nuclear and mitochondrial loci, 2) evaluate differences and similitudes in the signals of population history between both genomic markers, and 3) assess the role of oceanographic and ecological factors in shaping genetic differentiation. We further investigate whether gene flow among the Mediterranean Sea, Atlantic, Indian, and Pacific basins is constrained by physical and/or oceanographic barriers, and whether signals of population divergence and environmental adaptation can be detected. This integrative genomic framework provides an opportunity to resolve the evolutionary history and ecological dynamics of *C. hippurus* at a worldwide scale and to establish a broader context for the sustainable management of its fisheries.

2. Methods

2.1. Sample collection and sequencing

We analyzed whole-genome data from 87 individuals of *C. hippurus* collected across seven major marine regions: the Mediterranean Sea; Central, Eastern and Western Pacific Ocean; Western Indian Ocean; Eastern and Western Atlantic Ocean (Fig. 1, Table S1). We extracted DNA from muscle tissue using the DNeasy Blood and Tissue Kit (Qiagen) and quantified it with the Qubit 1X HS DNA Assay Kit (Thermo Fisher Scientific). We prepared libraries following the Illumina DNA Prep protocol, assessed fragment size on a QSep 400 (BiOptic), and measured final concentrations using Qubit. We performed sequencing on an Illumina NovaSeq 6000 platform with 2 × 150 bp paired-end reads. We sequenced each genome to ~ 5 × coverage under a low-pass design, to enable genome-wide variant discovery.

2.2. Nuclear genome dataset

2.2.1. Genomic alignment and variant calling

We assessed raw read quality using FastQC and trimmed reads with Trimmomatic. We aligned filtered reads to the *C. hippurus* reference genome (GenBank accession: GCA_047653855.1; Hinojosa-Alvarez et al. 2025) using BWA-MEM. We performed variant calling and genotype likelihood estimation in ANGSD v0.935 (Korneliussen, Albrechtsen & Nielsen 2014), applying filters for base quality ($\geq$ 20), mapping quality ($\geq$ 30), and removal of ambiguous reads. We estimated genotype likelihoods using the GATK model (−GL 2) and used posterior probabilities to calculate minor allele frequencies (MAF $\geq$ 0.05). We further filtered the resulting VCF with VCFtools v0.1.17 (Danecek et al. 2011), retaining sites with a maximum of 20% missing data.

To minimize the effects of linkage disequilibrium (LD), we performed LD pruning using the function *snpgdsLDpruning* implemented in the SNPRelate package v1.34.1 in R (Zheng et al. 2012), applying a threshold of 0.2. We used the resulting subset of independent SNPs in downstream analyses, using parallel computation with 20 threads.

2.2.2. Population structure analyses

We investigated population structure using complementary approaches. We used NGSAdmix (Skotte, Korneliussen & Albrechtsen 2013) and PCAngsd v1.5 (−admix) (Meisner, Albrechtsen & Hanghøj 2021) to estimate principal components and individual ancestry proportions directly from genotype likelihoods. We performed analyses for K = 2–7 ancestral clusters and selected the optimal model based on entropy minimization.

We also applied Discriminant Analysis of Principal Components (DAPC) as implemented in the adegenet package in R v4.3.2 (Jombart 2008). Prior to analysis, we determined the optimal number of principal components to retain using cross-validation with the function *xvalDapc*. We performed cross-validation with parameters n.pca.max = 50, training.set = 0.9, and n.rep = 30. DAPC maximized among-population variance while minimizing within-population variance, and we extracted assignment probabilities from the posterior membership matrix.

To quantify genetic differentiation among regions, we calculated pairwise Weir and Cockerham's F_{ST} values (Weir & Cockerham 1984)

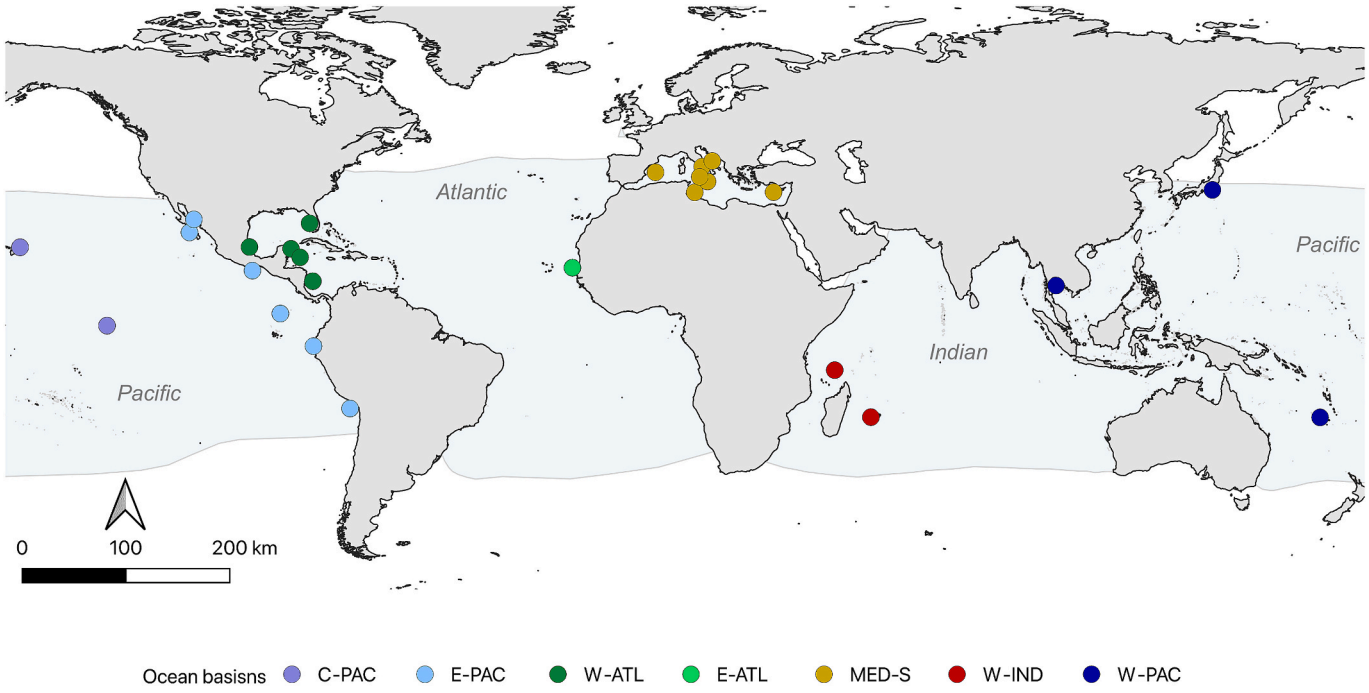

Fig. 1. Sampling sites of *C. hippurus* distributed across major ocean basins: Western Pacific (W-PAC), Central Pacific (C-PAC), Eastern Pacific (E-PAC), Western Atlantic (W-ATL), Eastern Atlantic (E-ATL), Mediterranean Sea (MED-S), and Western Indian Ocean (W-IND).

using the hierfstat v0.5–11 R package (Goudet 2005). We assessed statistical significance using 999 permutations of multilocus genotypes across populations to generate a null distribution of F_{ST} values under panmixia, and we applied false discovery rate (FDR) correction. We visualized the resulting pairwise F_{ST} matrix as a distance matrix using the ape v5.8 R package (Paradis & Schliep 2019). We evaluated the robustness of inferred clusters by estimating bootstrap support values for each pairwise F_{ST} comparison across 999 replicates.

To evaluate the partitioning of genetic variance at different hierarchical levels, we performed analyses of molecular variance (AMOVA) using the poppr package in R (Kamvar, Brooks & Grünwald 2015). We tested two grouping hypotheses: (H1) four groups defined by major ocean basins (C-PAC + W-PAC + E-PAC; W-ATL + E-ATL; MED-S; W-IND), and (H2) three broader groups in which Indian Ocean samples were included within the Pacific (C-PAC + W-PAC + E-PAC + W-IND; W-ATL + E-ATL; MED-S). We assessed significance using 999 permutations.

We calculated population genetic diversity metrics for each regional ocean basin, including observed heterozygosity (*Ho*), expected heterozygosity (*He*), inbreeding coefficient (F_{IS}), and allelic richness (*AR*), using the hierfstat R package. We averaged heterozygosity across loci, and we standardized allelic richness using rarefaction to account for differences in sample size. We estimated genome-wide Tajima's D values for each population using ANGSD to evaluate deviations from neutrality.

2.2.3. Genomic–environmental association (GEA)

We obtained minimum and maximum values of oceanographic variables from the Bio-ORACLE v3 repository (https://bio-oracle.org/) to represent contemporary climatological conditions at the sampling locations. These variables included seafloor topography (terrain ruggedness index, topographic position index, aspect, bathymetry), light and water clarity (diffuse attenuation, photosynthetically available radiation), water column structure and productivity (mixed-layer depth, chlorophyll, primary productivity, iron, silicate, phosphate, nitrate), and oceanographic dynamics and chemistry (sea surface temperature, salinity, dissolved oxygen, pH, and current velocity).

Collectively, these variables capture key physical, chemical, and biological factors shaping the habitat and distribution of *C. hippurus*. To avoid multicollinearity, we standardized all variables using z-transformation and removed highly correlated variables ($|\rho| > 0.7$) based on Spearman correlation and variance inflation factors (VIF) calculated in R (R Core Team 2014).

To assess the influence of environmental variables on genetic variation, we performed Redundancy Analysis (RDA) using the vegan R package v2.6–6 (Oksanen et al. 2017). We used the SNP genotype matrix as the response variable and included non-collinear oceanographic variables as predictors. We performed model selection using forward selection with *ordir2step* based on adjusted R^2. We assessed the significance of each predictor using permutation ANOVA (*anova.cca*, 999 permutations).

2.3. Mitochondrial genome dataset

2.3.1. Mitochondrial genome reconstruction

We reconstructed complete mitochondrial genomes from the raw sequencing data. We used a *C. hippurus* mitochondrial reference genome (GenBank accession: KF719178; Díaz-Jaimes et al. 2015), which we indexed using *samtools faidx* and *bwa index*. We aligned reads to the reference genome using BWA-MEM and retained high-quality alignments (MAPQ $\geq$ 30).

We generated consensus mitochondrial sequences for each individual using *bcftools mpileup*, *bcftools call*, and *bcftools consensus*. We aligned sequences with MAFFT v7 (Katoh et al. 2013) and visually inspected alignments to ensure accuracy.

We were unable to reconstruct complete mitochondrial genomes for all individuals due to variability in coverage and sequence quality within the mitochondrial fraction of the reads. Therefore, we retained a total of 83 out of the 87 individuals for subsequent analyses. We removed the control region due to its high variability and alignment uncertainty, resulting in mitochondrial sequences of 15,620 bp, which we used for downstream population genetic analyses.

2.3.2. Mitochondrial population structure analyses and genetic diversity measures

We inferred population structure from mitochondrial sequence data using the rhierBAPS R package (Tonkin-Hill et al. 2018), with a maximum hierarchical depth of two and exploring multiple levels of clustering (K = 1–10). We determined the optimal number of clusters based on posterior partition probabilities. We visualized cluster assignments by plotting posterior probabilities (Q-matrix) in a Structure-like barplot using pophelperShiny v2.1.1 from the pophelper R package (Francis 2017).

We performed pairwise Φ_{ST} comparisons, neutrality tests, and estimates of standard diversity indices. Specifically, we calculated the number of haplotypes (*H*), haplotype diversity (*Hd*), nucleotide diversity (π), and the number of segregating sites (*S*). We conducted all analyses in R using population genetics packages including adegenet (Jombart 2008), ape (Paradis & Schliep 2019), pegas (Paradis 2010), poppr (Kamvar et al. 2015), and ade4 (Dray & Dufour 2007).

2.3.3. Haplotype network

We reconstructed the mitochondrial haplotype network from the complete mitochondrial genome alignment with the ape package (v5.8–1). We identified haplotypes using the haplotype function implemented in the pegas R package (v1.2), and we inferred the TCS network using the PopART software (Leigh et al. 2015).

2.4. Nuclear and mitochondrial phylogenomic reconstruction

To evaluate the evolutionary relationships and potential genomic discrepances across basins, we performed a comparative phylogenetic analysis. To ensure a direct topological comparison, phylogenetic reconstructions were restricted to the subset of individuals for which both high-quality nuclear SNP data and complete mitochondrial sequences were available. Phylogenetic relationships were inferred using a Maximum Likelihood (ML) approach in IQ-TREE 2 v3.0.1 (Wong et al. 2025). For the mitochondrial dataset, the best-fit model of nucleotide substitution was selected using ModelFinder according to the Bayesian Information Criterion (BIC). Given the nature of SNP nuclear data, we applied the ascertainment bias correction (+ASC) to the substitution model to account for the absence of invariant sites in the alignment. The models were run under 1000 bootstrap iterations.

To evaluate topological incongruence between the mitochondrial and nuclear genomes, we performed a co-phylogenetic mirror plot analysis using the R package phytools (Revell, 2012). This approach was expanded into an integrated multi-panel diagram to align individual tree tips side-by-side with their respective nuclear (NGSAdmix) and mitochondrial (rhierBAPS) assignment profiles. The degree of congruence between the two phylogenetic trees was quantified using the Robinson-Foulds (RF) distance (Robinson and Foulds, 1981) and a Mantel test based on patristic distance matrices evaluated in ade4 and pegas R packages. Finally, we identified samples showing the largest differences between nuclear and mitochondrial distances relative to other individuals, these were classified as discordant samples.

2.4.1. Inter-basin gene flow and patterns of ancestral migration

To further investigate the biological drivers of the observed discordance, we conducted *D* and *f4-ratio* statistics using Dsuite v0.5 (Malinsky et al., 2021) to identify specific signals of inter-basin gene flow. Given the strong phylogenetic divergence of the Mediterranean lineage in the phylogenies, this population was used as a reference for allele polarization in *D-statistics* analyses. Accordingly, we evaluated multiple population triplets (P1, P2, P3), using the Mediterranean group as an outgroup-like baseline to detect non-random patterns of allele sharing among oceanic basins. Significance was assessed using block-jackknife Z-scores, and values of $|Z| > 3$ were considered evidence of significant excess allele sharing. Additionally, we calculated the *f4-ratio* to approximate the fraction of the genome potentially derived from ancestral admixture events. Because all sampled groups belong to the same species, results are interpreted as relative patterns of allele sharing rather than evidence of strictly directional gene flow. *P*-values were corrected using FDR.

To infer patterns of population splits and gene flow, we used TreeMix (Pickrell and Pritchard, 2012). For that, we fitted models allowing 0–7 migrations events, with Mediterranenan group (MED-S) as an outgroup. SNPs were grouped into blocks of 1,000 (−k 1000) to account for linkage disequilibrium. Model fit was assessed by comparing the log-likelihood across different values of m (m = 0–7), and the optimal number of migration edges was determined based on the point at which increases likelihood stabilized. The proportion of explained variance and residual covariance patterns, and inferred migration edges were interpreted as evidence of gene flow among ocean basins.

3. Results

3.1. Nuclear genome dataset

After quality filtering, 87 individuals and 8,751,028 SNPs were retained for downstream analyses. The initial dataset contained approximately 20% missing data, which was reduced to less than 2% following LD pruning.

3.2. Population structure analyses and genetic diversity measures

Admixture analysis (Fig. 2) revealed a clear population structure with K = 4, supported by the lowest mean admixture entropy (Supplementary Fig. S1). Populations were broadly grouped by ocean basins. Mediterranean Sea (MED-S) individuals were primarily assigned to cluster 1 or cluster 4, although few individuals showed ancestry components from clusters 2 and 3.

In the eastern Atlantic (E-ATL), individuals exhibited mixed ancestry primarily derived from cluster 1, also common in MED-S, and cluster 3, which was predominant in the Pacific populations. Several individuals displayed approximately equal proportions of clusters 1 and 3.

In the western Atlantic (W-ATL), individuals showed a broader range of ancestry patterns. Some individuals displayed ancestry proportions similar to those observed in E-ATL, whereas others showed combinations of clusters 2 and 4 (Fig. 2). Pacific populations (W-PAC, C-PAC, and E-PAC) displayed individuals assigned almost 100% to either cluster 1, cluster 2 and 3, suggesting higher connectivity within the basin and the presence of hidden population substructure within each considered region. The Western Indian Ocean (W-IND) population exhibited ancestry components shared with both Atlantic and Pacific populations; however, no individuals showed mixed ancestry proportions. Overall, ancestry coefficients varied considerably among individuals within basins, suggesting that additional hierarchical structure may exist beyond the four major clusters recovered by NGSadmix.

Discriminant Analysis of Principal Components (DAPC, Fig. 3) further supported the basin-level structure, with the first two discriminant functions explaining 55.6% and 33.2% of the variance, respectively. Atlantic and Mediterranean populations formed distinct clusters, with E-ATL and W-ATL closely grouped but are separated from MED-S. The W-IND population also formed an intermediary cluster, positioned closer to the Atlantic and Pacific groups. In contrast, Pacific populations formed an overlapping cluster, indicating low or any genetic differentiation within the basin.

Pairwise F_{ST} estimates among basins (Table 1) ranged from very low to moderate values (0.00—–0.053). The highest differentiation was observed for the Mediterranean (MED-S) and both Eastern (E-PAC, F_{ST} = 0.053, $P < 0.05$) and Central Pacific (C-PAC, F_{ST} = 0.047, $P < 0.05$), suggesting restricted gene flow between these regions. In contrast, within the Pacific and Atlantic basins estimates showed minimal differences; the lowest values were detected between Western (W-PAC) and Central Pacific (C-PAC) with a F_{ST} = 0.002 ($P < 0.05$). Similarly low

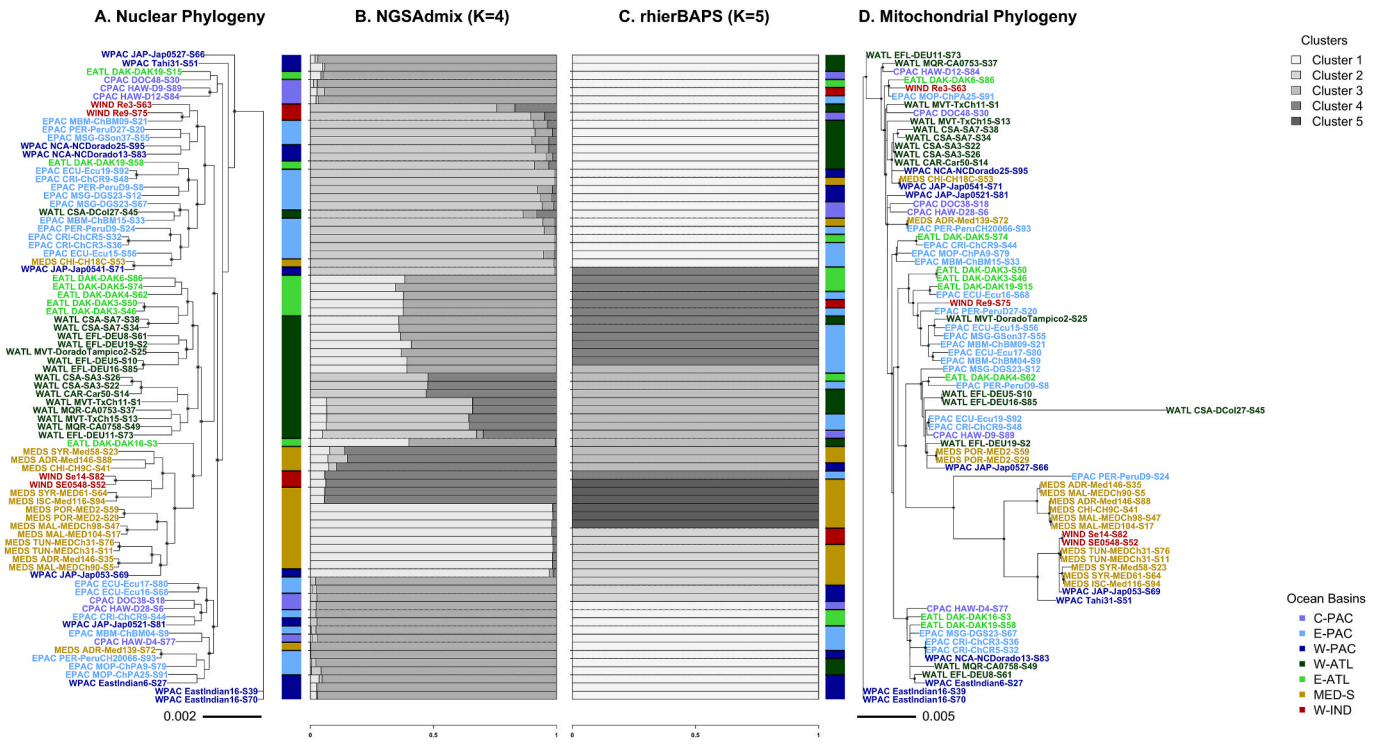

Fig. 2. Individual-level comparison of nuclear phylogeny (A), nuclear ancestry assignments (B: NGSadmix, K = 4), mitochondrial clustering (C: rhierBAPS, K = 5), and mitochondrial phylogeny (D) among individuals from different ocean basins in dolphinfish (*Coryphaena hippurus*).

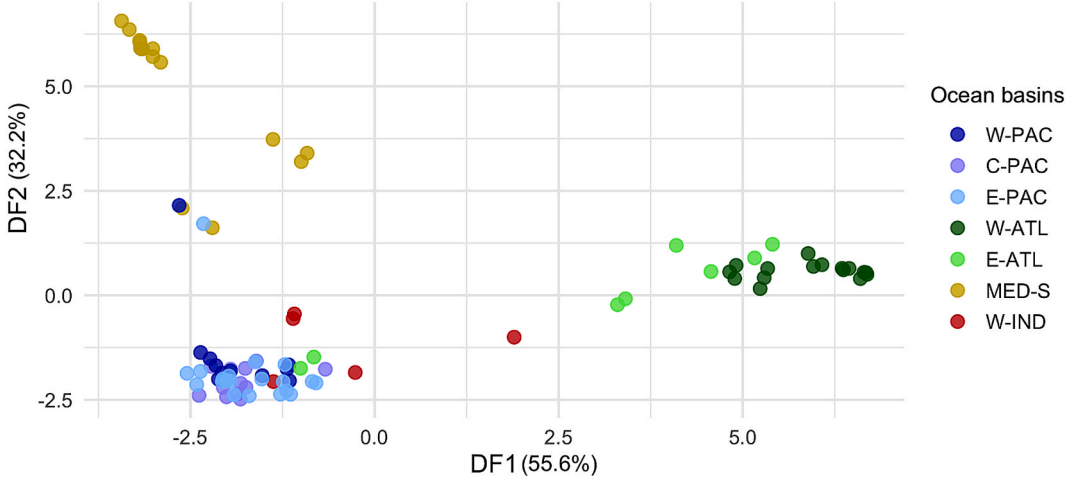

Fig. 3. Discriminant Analysis of Principal Components (DAPC) showing genetic clustering by ocean basin. Each point represents an individual, colored by an ocean basin. The first two discriminant functions (DF1 and DF2) explain 87.8% of the among-group variation, highlighting clear genetic structure across regions.

Table 1
Pairwise F_{ST} comparisons of nuclear (below the diagonal) and mitochondrial (above the diagonal) genomes across different regions of the ocean basins. See Fig. 1 for acronyms. Significance codes: 0.001 '***', 0.01 '**', 0.05 '*'.

	W-PAC	C-PAC	E-PAC	W-ATL	E-ATL	MED-S	W-IND
W-PAC	–-	0.081	0.062	0.051**	0.069	0.013	0.027
C-PAC	0.002**	–-	0.271	0.672	0.253	0.001	0.188
E-PAC	0.007*	0.019**	–-	0.051	0.581**	0.001	0.131**
W-ATL	0.020**	0.031***	0.027***	–-	0.455*	0.001	0.026
E-ATL	0.008**	0.007**	0.027***	0.010*	–-	0.434***	0.163
MED-S	0.037***	0.047**	0.053***	0.034**	0.035*	–-	0.159**
W-IND	0.012*	0.037*	0.015**	0.015**	0.026*	0.021**	–-

differentiation was observed between Eastern and Western Atlantic regions ($F_{ST} = 0.010$, $P < 0.05$).

AMOVA revealed that most of the genetic variation was found within populations, with 94.87 % and 94.70 % of total variance for the four-group (H1) and three-group (H2) hypotheses, respectively (F_{ST} = 0.949 and 0.947; $P < 0.001$, Supplementary Table S2). Under H1, the variation among groups was significant and accounted for 3.76 % of the total variance (F_{CT} = 0.038; P = 0.001), whereas variation among populations within groups contributed 1.37 % (F_{SC} = 0.014; P = 0.002). Under H2, in which the Indian Ocean population was grouped with the Pacific, among-group variance was slightly lower at 3.50 % but significant (F_{CT} = 0.035; P = 0.001, Table Supplementary S2), and among populations within groups contributed 1.80 % of total variance (F_{SC} = 0.018; P = 0.002).

Genome-wide Tajima's D estimates revealed that E-PAC showed the highest value (0.69), followed by MED-S (0.40). Whereas C-PAC (0.12) and W-IND (0.11) were close to zero. In contrast, negative values were detected in some populations, including C-PAC (–0.13) and E-ATL (–0.14) (Supplementary Table S1).

Genetic diversity across all ocean basins revealed consistently high levels of observed and expected heterozygosity (*Ho* = 0.966–0.969; *He* = 0.560–0.568), indicating substantial standing genetic variation within *C. hippurus* populations (Supplementary Table S1). Allelic richness was also comparable among regions (AR = 2.12–2.18), reflecting a uniformly diverse nuclear gene pool. Inbreeding coefficients were negative across all basins (F_{IS} ranging from –0.731 to –0.711).

3.3. Genomic–environmental association (GEA)

The redundancy analysis (RDA) indicated that eleven of nineteen uncorrelated environmental predictors significantly explained genetic variation (best RDA model = genetic ~ SAL + PO. + Kd + PAR + pH + Si + TRI + ASP + VEL + NO. + SST; F = 1.58, P = 0.001), accounting for 18.8% of the total variance, while 81.2% was attributed to unconstrained components. Among individual predictors, salinity (SAL, F = 2.57, P = 0.001, Supplementary Table S3), phosphate (PO, F = 1.90, P = 0.001), light attenuation coefficient (Kd; F = 1.61, P = 0.001), photosynthetically available radiation (PAR; F = 1.56, P = 0.001), and pH (F = 1.63, P = 0.001) showed the strongest associations with genetic variation. Additional significant correlation was also detected for silicate (SIL, F = 1.52, P = 0.002), terrain ruggedness index (TRI; F = 1.47, P = 0.003), topographic aspect (ASP; F = 1.32, P = 0.015), current velocity (VEL; F = 1.24, P = 0.042), nitrate (NO, F = 1.31, P = 0.021), and sea surface temperature (SST; F = 1.27, P = 0.038, Supplementary Table S3).

The first two RDA axes explained 19.2% and 12.2% of the constrained genetic variance, respectively (Fig. 4), showing that the environmental data explained an important proportion of population variation. The MED-S population indicated a strong association with salinity. The Atlantic populations (W-ATL and E-ATL) grouped together at the negative end of the RDA, suggesting that the variation in these populations is strongly correlated to sea surface temperature, pH, diffuse attenuation, and silicates. Pacific populations (W-PAC, C-PAC, and E-PAC) were associated with primary productivity and current velocity. In contrast, the W-IND basin showed intermediate positioning across environmental gradients.

3.4. Mitochondrial genome dataset

Mitochondrial population structure analyses and genetic diversity measures

Hierarchical clustering using rhierBAPS identified five clusters (K = 5) as the optimal model based on assignment probabilities and model likelihoods (log(ml × prior) = –35.516; Supplementary Table S4). Cluster membership did not show a clear geographic structure, as individuals from most regions (W-PAC, C-PAC, E-PAC, and W-ATL) were distributed across multiple clusters (Fig. 2C). A slight tendency toward grouping was observed in Mediterranean samples (MED-S), which showed partial clustering.

Pairwise Φ_{ST} values ranged from 0.024 (W-ATL–W-PAC) to 0.434 (E-

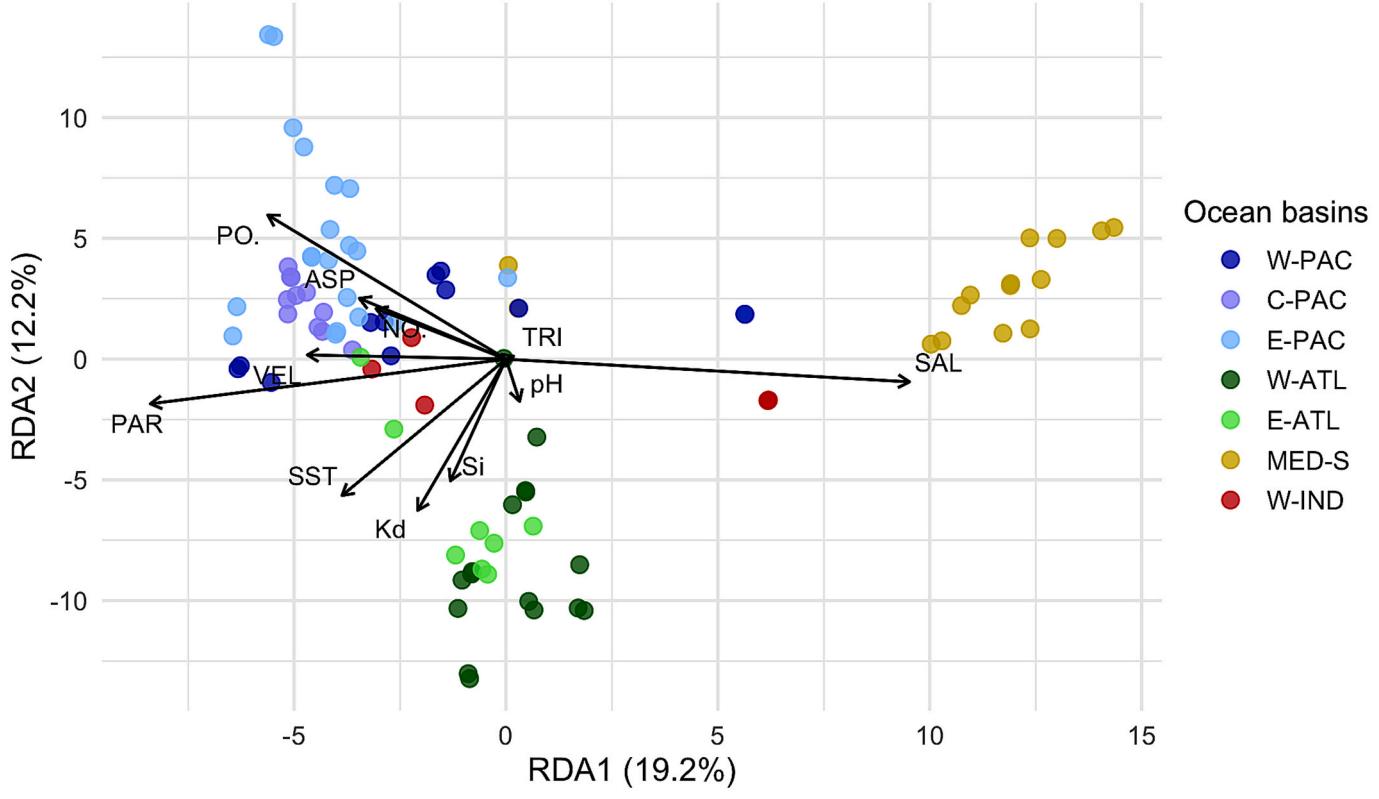

Fig. 4. Redundancy analysis of genome-wide variations constrained by environmental data.

ATL–MED-S). Comparisons involving MED-S showed the highest values (e.g., MED-S–E-ATL = 0.434; MED-S–E-PAC = 0.432; MED-S–W-ATL = 0.404), all statistically significant ($P \leq 0.014$). The W-IND population also showed significant differentiation from E-PAC ($\Phi_{ST} = 0.131$; $P = 0.012$) and W-ATL ($\Phi_{ST} = 0.130$; $P = 0.026$). All other pairwise comparisons were not significant (Table 1).

AMOVA results indicated low but highly significant variance among groups ($\Phi_{CT} = 0.0204$; $P = 0.0001$). In contrast, variance among populations within groups ($\Phi_{SC} = 0.4033$; $P = 0.9984$) and overall population differentiation ($\Phi_{ST} = 0.4154$; $P = 0.9852$) were not significant.

A total of 70 haplotypes were identified. The number of haplotypes (H) ranged from six in C-PAC, W-IND, and W-PAC to 18 in E-PAC. Mitogenome analysis revealed high haplotype diversity across all locations ($Hd = 0.917 – 1.000$) in contrast to the low nucleotide diversity (π), which ranged from 0.002 in C-PAC to 0.005 in W-IND. The number of segregating sites (S) varied considerably, from 102 in C-PAC to 325 in W-ATL (Table S1). Tajima's D values ranged from −1.935 in W-ATL to 0.803 in MED-S. W-IND (0.320) and MED-S (0.803) locations showed positive values (Supplementary Table S1).

3.5. Mito-Nuclear discordance

Nuclear and mitochondrial phylogenies exhibited topological incongruence. The normalized Robinson–Foulds distance between trees was 0.80, indicating that a large proportion of branching patterns differed between genomic compartments.

The nuclear phylogeny supports the DAPC analysis separating the W-ATL and the E-ATL in two different lineages; excepting of few individuals that occupied intermediate positions between two clades (Fig. 2). In contrast, the mitochondrial phylogeny showed weaker resolution between E-ATL and W-ATL lineages, with partial haplotype sharing across Atlantic sub-basins. Mediterranean individuals (MED-S) were mostly recovered within a deeply divergent and monophyletic clade in both nuclear and mitochondrial datasets. In contrast, Pacific samples displayed additional hierarchical subdivision, with Western, Central, and Eastern Pacific lineages forming distinct subclades. The Western Indian Ocean lineage occupied an intermediate position between Atlantic and Pacific groups.

At the individual level, several discrepancies between geographic origin and phylogenetic placement were observed. Most Mediterranean samples formed a cohesive nuclear lineage; however, two individuals (MED-CHI-CHI18C-S53 and MED-ADR-MED139-S72) occupied relatively divergent positions within the tree. Likewise, two Western Indian Ocean specimens (WIND-Se14-S82 and WIND-SE0548-S52, Fig. 2) clustered close to Mediterranean-associated lineages despite their geographic origin. Similar patterns were also observed in some Atlantic and Pacific specimens occupying intermediate positions among major clades. These individuals did not exhibit unique ancestry profiles in NGSadmix and were not associated with exclusive mitochondrial lineages, suggesting that the observed discrepancies represent localized departures from the predominant basin-level structure rather than distinct evolutionary units.

In contrast, the mitochondrial phylogeny showed weaker geographic structure With Western and Eastern Atlantic haplotypes exhibiting partial overlap, and extensive lineage mixing was observed across Pacific regions. Notably, many Eastern Pacific individuals were nested within clades containing Western Atlantic haplotypes, a pattern that was completely absent in the high-resolution nuclear tree (Fig. 2). Despite the overall monophyly of the Mediterranean assemblage, none of the individuals occupying atypical positions in the nuclear phylogeny formed exclusive mitochondrial lineages. Instead, they remained embedded within broader mitochondrial clades, indicating that individual-level discordance was distributed across multiple ocean basins rather than being restricted to a single region.

Despite this topological discordance, patristic distance matrices were moderately correlated (Mantel r = 0.38, $P = 0.001$), suggesting partial but incomplete concordance in pairwise divergence patterns.

3.5.1. Inter-basin gene flow and patterns of ancestral migration

D-statistics analyses revealed multiple signals of excess allele sharing among ocean basins (Table S5). Several population trios showed significant deviations from the null expectation of no admixture ($|Z| > 3$), with the strongest signals involving comparisons between Indo-Pacific and Atlantic populations (Fig. S3). Significant positive *D*-values were detected in tests involving the Eastern Atlantic and Western Indian Ocean ($D = 0.15$, $Z = 6.70$, $P < 0.001$), as well as between Western Pacific and Eastern Atlantic populations ($D = 0.17$, $Z = 18.52$, $P < 0.001$). Comparisons involving Western Atlantic and Western Indian Ocean populations also showed significant signals ($D = 0.15$, $Z = 5.73$, $P < 0.001$; Fig. S3).

Additional significant signals were observed across multiple combinations of Pacific and Atlantic populations, although with lower effect sizes (e.g., $D \approx 0.02–0.05$), suggesting more moderate levels of allele sharing. In contrast, some comparisons, particularly among Pacific populations, did not show significant deviations from the null expectation, indicating more limited evidence of gene flow in these cases (Table S5). Estimates of the *f4-ratio* support heterogeneous ancestry sharing across regions, with values ranging from near zero to approximately 0.25, suggesting variable contributions of migrant ancestry among population pairs.

Statistical signals were corroborated by maximum likelihood phylogenetic networks in TreeMix (Fig. 5). The optimal model, reached at $m = 3$ migration events, showed a substantial increase in log-likelihood compared to the null model (*m0*, Table S6.), indicating that a strictly bifurcating tree is insufficient to explain the observed covariance. Consistent with the high *D*-values, the TreeMix graph identified a high-intensity migration event ($w \approx 0.5$) from the Pacific lineage toward the Western Indian Ocean (W-IND), confirming a major ancestral contribution to this region (Fig. 5). Furthermore, the network revealed a significant *trans*-Pacific connection between Western (W-PAC) and Eastern Pacific (E-PAC) lineages, and a notable gene flow event towards the Eastern Atlantic. Despite these additions, the residual covariance matrix highlighted remaining unresolved affinities, particularly between E-ATL and E-PAC (Fig. S4), suggesting that the colonization history of these basins likely involved rapid divergence or additional pulses of gene flow not fully captured by the primary graph.

4. Discussion

Results from this high-resolution genomic study, reveal for the first time, the global-scale genetic differentiation in dolphinfish based on both nuclear and mitochondrial genomes. Four genetically distinct lineages were identified (Pacific, Atlantic, Indian, and Mediterranean) however, these lineages are not strictly isolated. Instead, *D*-statistics, *f4-ratio* estimates, and phylogenetic networks indicate a complex history of partial divergence with historical connectivity among oceans. This contrasts with previous studies based on individual mtDNA sequences (e.g. ND1) which were unable to resolve these broad-scale differences (Díaz-Jaimes et al. 2010).

A key finding of our study is the pronounced mito-nuclear discordance, reflecting that mitochondrial and nuclear genomes capture distinct evolutionary timescales. This topological incongruence highlights that mitochondrial data alone do not fully resolve the complex evolutionary history of this species. Ancestral migration signals inferred from nuclear data indicate that the Western Indian Ocean (W-IND) has acted as a key source of genetic variants toward both the Pacific and Atlantic basins. This pattern is consistent with phylogeographic signatures reported in other cosmopolitan pelagic taxa, which often exhibit divergent Indo-Pacific and Atlantic lineages connected by historical gene-flow corridors (Toonen et al. 2016; Haro-Bilbao et al. 2021; da Silva Ferrette et al. 2023). Consistent with this complex history, individual-level comparisons revealed localized mismatches between

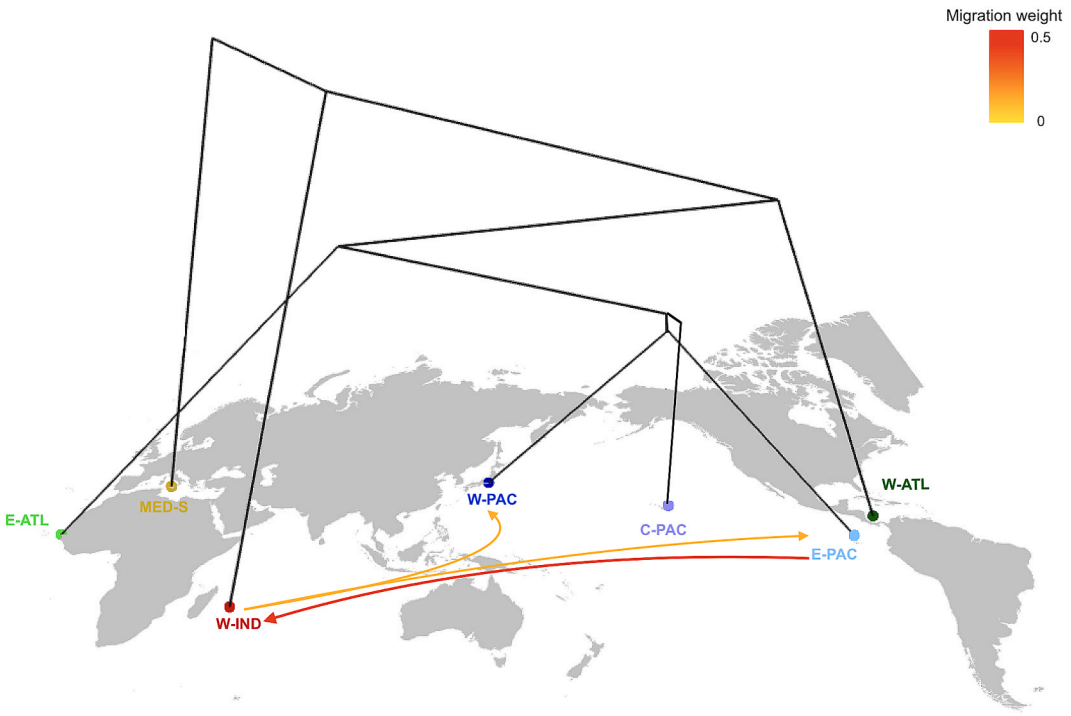

Fig. 5. Maximum likelihood phylogenetic network of global populations. The graph shows the relationships among basins with $m = 3$ migration events. Arrows represent gene flow, with the color scale indicating the migration weight (w). The underlying map shows the hypothesized spatial trajectories of these historical gene flow events.

ancestry profiles, phylogenetic placement, and geographic origin across multiple basins. Although the number of individuals available from several localities was limited, the analysis of more than 8.7 million nuclear SNPs provided strong resolution of basin-scale relationships. Therefore, the few individuals occupying atypical positions are unlikely to represent cryptic forms or complete reproductive isolation. Instead, they most likely reflect incomplete lineage sorting and retention of ancestral polymorphisms associated with historical connectivity among ocean basins, a pattern frequently observed in highly dispersive marine species with large effective population sizes (Toews and Brelsford, 2012). Comparable patterns have been reported in other highly mobile marine predators. For example, phylogeographic analyses of the great white shark (*Carcharodon carcharias*) suggested that the Mediterranean population originated through ancient colonization events and retained distinctive genetic signatures despite historical connectivity with adjacent oceans (Leone et al. 2020).

4.1. Global population structure from an historical perspective

Genetic differences among ocean basins were more clearly resolved in nuclear than mitochondrial data. These findings highlight the need of high-resolution genomic approaches for detecting population structure in widely distributed species characterized by high reproductive potential, rapid growth, high dispersal capacity, and short generation times (Pecoraro et al. 2018; Haro-Bilbao et al. 2021; da Silva Ferrette et al. 2023; Mar-Silva et al. 2023).

In contrast, mitochondrial analyses showed more limited resolution. Although AMOVA separated populations into ocean basins, including the Mediterranean Sea, previous studies based on ND1 mtDNA sequences failed to detect genetic differentiation among basins, except for the deep divergence observed in the Mediterranean (Díaz-Jaimes et al. 2010). The use of the whole mtDNA genome confirmed also the isolation of the Mediterranean population which could be due to eustatic events that occurred during the Glacial periods in the Pleistocene (Domingues et al. 2007; Ludt and Rocha 2015).

Genetic differentiation among major ocean basins was clearly resolved in the nuclear dataset, indicating partial restriction to gene flow across large geographic and oceanographic scales. However, this differentiation was not complete, as several individuals showed mixed nuclear ancestry, while the mitochondrial haplotype network revealed haplotype sharing among basins. Together, these patterns support a scenario of partial divergence with connectivity among ocean basins rather than strict isolation. Such a pattern of incomplete isolation has also been described in other highly migratory marine species, where historical colonization events and intermittent connectivity generate genetic discontinuities that are stronger in genetic datasets than expected from present-day geography alone (Leone et al. 2020).

Restrictions to gene flow among ocean basins arise from both physical barriers (e.g., land masses) and oceanographic features. However, in highly migratory pelagic species with large dispersal potential, these barriers are often semi-permeable, limiting but not preventing gene flow among basins (Reid et al. 2016). The Panama Isthmus represents the most prominent physical barrier for tropical marine species, driving divergence between the Eastern Pacific and Western Atlantic regions (Haug and Tiedemann 1998). While previous studies failed to detect differentiation across this boundary in species such as *C. hippurus* (Díaz-Jaimes et al. 2010) and yellowfin tuna (Ely et al. 2005), our results reveal strong nuclear differentiation based on 8.7 million SNPs. In contrast, no differentiation was detected in the mitochondrial genome.

This pattern suggests that major oceanographic barriers such as the Panama Isthmus and the Eastern Pacific Barrier (EPB) constrain contemporary nuclear gene flow but do not represent absolute barriers. Instead, connectivity in *C. hippurus* reflects a combination of historical isolation and variable gene flow mediated by oceanographic conditions. In this context, the mitochondrial genome may retain ancestral polymorphisms or reflect historical connectivity predating the complete sorting of maternal lineages.

Significant differences based on nuclear data were observed between the Indian Ocean and both Atlantic regions, whereas mtDNA data showed more limited differentiation, restricted to comparisons with the

western Atlantic. This pattern is consistent with a scenario in which differentiation among basins exists but is not complete. Interruptions to gene flow in tropical species have been inferred to occur through the Southern tip of Africa during glacial periods. Secondary contact has also been detected during inter-glacial periods, resulting in complex genetic patterns (Toonen et al. 2016). This region is considered a soft barrier for tropical species that can move between both oceans through South Africa (Teske et al. 2011) during the summer season when the cold Agulhas current diminishes in intensity. The absence of strong mitochondrial differentiation may reflect this intermittent connectivity. Similar patterns have been reported in other pelagic species; for example, genomic analyses in yellowfin tuna (*Thunnus albacares*) revealed predominant gene flow from the Indian Ocean into the Atlantic basin (Barth et al. 2017).

In addition to connectivity across the Atlantic and Indian Oceans, differentiation between the Indian and Pacific basins further reflects the role of regional barriers in shaping population structure. Significant differences between the Indian Ocean and western Pacific were detected with nuclear dataset (Table 1), suggesting reduced connectivity across the Indo-Australian Archipelago. This region, which includes approximately 22,000 islands and the Sunda Shelf biogeographic barrier, represents a major discontinuity for marine species. Historical reductions in sea level during Pleistocene glacial periods exposed large portions of the continental shelf (Sundaland), limiting dispersal between the eastern Indian Ocean and the western Pacific (Voris 2000; Husson et al. 2020; Sarr et al. 2019). These eustatic changes have been associated with reduced gene exchange in many tropical marine taxa (Cheng & Faidi 2025). Despite this, the Sunda Shelf appears to function as a semi-permeable barrier, restricting but not completely preventing gene flow. This is consistent with the additional genetic differences observed between the Indian Ocean and central and eastern Pacific regions, reflecting limited but ongoing connectivity across the Indo-Pacific.

The Mediterranean Sea is a semi closed basin connected to the Eastern Atlantic by the strait of Gibraltar (~12.9 km wide and ~ 286 m deep), which restricts water exchange and consequently limits gene flow for species relying on larval dispersal (Baschek et al. 2001). This connection has been interrupted by drastic sea level reductions during the Messinian salinity crisis about 3.5 million years ago also and the glacial cycles in the Pleistocene (Krijgsman et al. 1999; Krijgsman et al. 2024). Together, these processes have established the Mediterranean as an independent biogeographic province with distinct oceanographic conditions, resulting in a well-documented phylogeographic break in many marine species (Patarnello, Volckaert & Castilho 2007). Significant differentiation was also observed in both mitochondrial and nuclear dataset supporting the existence of a genetically differentiated population. The Mediterranean population appears to have experienced a higher degree of isolation relative to other ocean basins, likely driven by both historical and contemporary factors reflected in both genomic datasets.

4.2. Genetic-environmental associations as a contemporary view

Our results showed that oceanographic and environmental gradients contribute to shaping the genetic structure of the pelagic fish *C. hippurus* across major ocean basins. Despite the species' high dispersal capability and broad distribution (Palko et al. 1982; Oxenford, 1999), the RDA revealed significant associations between genetic variation and large-scale environmental variables as salinity, phosphate, light availability (PAR, Kd), and pH emerging as the strongest predictors (Supplementary Table S3). This pattern is consistent with growing evidence on highly dispersal pelagic fishes exhibiting genetic subdivision driven by oceanographic regimes (Toonen et al. 2016; Ochoa-Zavala et al. 2022).

The strong influence of salinity along RDA1 underscores the role of water mass properties in promoting genetic differences between populations, particularly between the Atlantic and Mediterranean Sea, as well as between Mediterranean with Indian with Pacific basins. Distinct of salinity regimes and current systems may act as semi-permeable barriers to gene flow or promote adaptive variation (Gaither & Rocha, 2013). The Mediterranean oceanography differs from the Atlantic and Indo-Pacific basins for its semi-enclosed hydrography, elevated salinity, and restricted water exchange (Patarnello et al. 2007). These features classify this ocean as a marine phylogeographic break. Several studies have reported how environmental gradients shape population genetic structure between the Atlantic and Mediterranean (e.g. Mendoza-Portillo et al. 2026).

For instance, genome wide analysis in the Atlantic bluefin tuna (*T. thynnus)* identified distinct Atlantic and Mediterranean spawning stocks despite extensive adult mixing, with divergence maintained by differential thermal regimes and spawning phenology between basins (Riccioni et al. 2013). Similarly, the European sea bass (*Dicentrarchus labrax)*, exhibits strong genome-wide differentiation between Atlantic and Mediterranean lineages, driven by local adaptation to contrasting salinity and temperature regimes (Tine et al. 2014). Likewise, the gilthead seabream (*Sparus aurata*) displays regional structure associated with environmental heterogeneity across coastal gradients, where variable thermal and salinity conditions promote local selection and limited effective dispersal (Chaoui et al. 2012). In *C. hippurus*, the elevated salinity and reduced productivity (Fig. 4), characterizing the Mediterranean seems to act as environmental filters maintaining a distinct genomic signature in comparison with the rest of the ocean basins.

Nutrient availability and light-related variables were also significant predictors of genetic variation, highlighting the influence of large-scale productivity gradients on dispersal and local adaptation in *C. hippurus*. In the RDA (Fig. 4), phosphate concentration (PHO), photosynthetically available radiation (PAR), and light attenuation (Kd) were among the strongest environmental drivers associated with the Pacific and Atlantic clusters, emphasizing ecological contrasts between these basins.

Elevated phosphate levels associated with upwelling systems such as the Humboldt Current enhance primary productivity and trophic transfer efficiency, influencing pelagic habitat structure (Graco et al. 2017; Massing et al. 2022). In parallel, PAR and Kd reflect light availability and water column properties that shape vertical habitat structure (Morel et al. 2007). Together, these variables capture key oceanographic gradients that can.

The Eastern Pacific is one of the most productive marine regions globally, characterized by strong seasonal upwelling and supporting diverse communities of plankton, fishes, marine mammals, and birds (Fiedler & Lavín 2017). The pronounced temporal and spatial variability of upwelling events generate dynamic foraging habitats that influence seasonal movements, reproductive timing, and population turnover in *C. hippurus* (Gozzer 2015). Abundance and catch per unit effort coincide with peak upwelling periods, when increased primary productivity suggests that reproductive activity and larval survival of dolphinfish are closely associated with nutrient availability. Population genetic studies in this region have supported the role of productivity regimes in shaping the evolutionary history of *C. hippurus*. Mar-Silva et al. (2023) identified genomic differences between northern and southern populations, consistent with strong latitudinal gradients in temperature and upwelling intensity along the Humboldt system. Similar environment-linked patterns have been described in other large pelagic fishes inhabiting the same region. Yellowfin tuna (*T. albacares*) exhibits subtle genomic structure between northern and southern populations associated with gradients in upwelling intensity and sea-surface temperature (Pecoraro et al. 2018), while swordfish (*Xiphias gladius*) populations show genetic breaks coinciding with the Humboldt upwelling front and the oxygen-minimum zone (Alvarado Bremer et al. 2005).

Major oceanographic features such as the Indo-Pacific Barrier, the Isthmus of Panama, and the Benguela upwelling system have been described as important breaks influencing dispersal in marine species (Randall 1998; Lessios 2008; Bowen et al. 2016). In this context, our results highlight the western Indian Ocean (W-IND) as a transitional region connecting the Indian and Pacific basins. This group forms a

distinct cluster in both nuclear and mitochondrial datasets, while still exhibiting shared ancestry with Pacific populations, consistent with partial connectivity across the Indo-Pacific. Environmental associations further support this interpretation, as genetic variation in W-IND is linked to gradients in salinity, phosphate, and pH (Fig. 4), reflecting the heterogeneous conditions characteristic of this region (Crandall et al. 2019).

The Atlantic Ocean is characterized by broad-scale circulation patterns that promote extensive connectivity among pelagic species. The western basin, influenced by the Gulf Stream and Caribbean Current, is connected to the eastern Atlantic through the North and South Equatorial Currents, facilitating *trans*-Atlantic dispersal of larvae and juveniles (Toonen et al. 2016). In *C. hippurus*, this high connectivity is reflected in the very low differentiation observed between western and eastern Atlantic populations (Table 1), as well as in the high levels of heterozygosity, consistent with strong gene flow and large effective population sizes maintaining genomic cohesion across the basin.

This pattern is also consistent with previous studies based on mitochondrial markers, which reported low differentiation across Atlantic populations, with the Mediterranean representing the main exception (Díaz-Jaimes et al. 2010; Maggio et al. 2019). Despite this overall connectivity, the RDA (Fig. 4) revealed subtle associations between allele frequencies and environmental gradients such as temperature, pH, and diffuse attenuation (Kd), suggesting that weak but recurrent selective processes may operate along latitudinal transitions between tropical and subtropical waters. Similar fine-scale signals of selection under high connectivity have been reported in other pelagic fishes, including wahoo (Haro-Bilbao et al. 2021), yellowfin tuna (*T. albacares*) (Pecoraro et al. 2018), and sailfish (da Silva Ferrette et al. 2023).

The use of genome wide low coverage data resolved a multilayered evolutionary history of *C.hippurus* showing a clear example of how marine populations are shaped by ancient vicariant events, dispersal through ocean circulation and local environmental variables across species distributional range. Future work will benefit from integrating oceanographic simulation, functional genomics, and targeted tests of loci under selection that permits each population to adapt and persist in rapidly changing ocean conditions.

Data statement

The raw data used in this study is openly available in Figshare under the following DOI: 10.6084/m9.figshare.32971490 and in the NCBI databases under BioProject accession PRJNA1146410 and BioSample accession SAMN43101139.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Grammarly and ChatGPt to review the correct use of English. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Píndaro Díaz Jaimes: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Verónica Mendoza-Portillo:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Silvia Hinojosa-Alvarez:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Erika Magallón-Gayón:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Rocío Alejandra Chavez-Santoscoy:** Writing – review & editing, Writing – original draft. **Jesús Hernández-Pérez:** Writing – review & editing, Writing – original draft, Validation. **Adán Fernando Mar-Silva:** Writing – review & editing, Writing – original draft, Validation. **Maried Ochoa-Zavala:** Writing – review & editing, Writing – original draft, Validation. **Sofia Ortega-Garcia:** Writing – review & editing, Visualization, Validation. **Oscar Lao:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Conceptualization. **Marco Arculeo:** Supervision, Resources, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Mexico, through the project "Dilucidando los componentes evolutivos de la divergencia genética en el ambiente marino: el dorado Coryphaena hippurus como modelo de estudio" (grant no. CF-2023-G-493).

Verónica Mendoza-Portillo thanks SECIHTI for a postdoctoral fellowship EPM_2023(1).

We thank the 'Chihuil' cluster from Instituto de Ciencias del Mar y Limnología -UNAM for facilitating the bioinformatic analysis.

We appreciate the constructive feedback provided by the anonymous reviewers on this Manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2026.108694.

Data availability

The raw data used in this study is openly available in Figshare under the following DOI: https://doi.org/10.6084/m9.figshare.32971490

References

Alvarado Bremer, J.R., Viñas, J., Mejuto, J., Ely, B., Pla, C., 2005. Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. Mol. Phylogenet. Evol. 36 (1), 169–187.

Barth, J.M.I., Damerau, M., Matschiner, M., Jentoft, S., Hanel, R., 2017. Genomic differentiation and demographic histories of atlantic and indo-pacific yellowfin tuna (*Thunnus albacares*) populations. Genome Biol. Evol. 9 (4), 1084–1098.

Baschek, B., Send, U., Lafuente, J.G., Candela, J., 2001. Transport estimates in the Strait of Gibraltar with a tidal inverse model. J. Geophys. Res. Oceans 106 (C12), 31033–31044.

Bayona-Vásquez, N.J., Glenn, T.C., Kieran, T.J., Pierson, T.W., Hoffberg, S.L., Scott, P.A., Bentley, K.E., Finger, J.W., Louha, S., Troendle, N., Diaz-Jaimes, P., Mauricio, R., Faircloth, B.C., 2019. 2019, 'Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD)'. PeerJ 10, e7724.

Chaoui, L., Gagnaire, P.A., Guinand, B., Quignard, J.P., Tsigenopoulos, C., Kara, M.H., Bonhomme, F., 2012. Microsatellite length variation in candidate genes correlates with habitat in the gilthead sea bream *Sparus aurata*. Mol. Ecol. 21 (22), 5497–5511.

Cheng, S., Faidi, M.A., 2025. Palaeodrainages of the Sunda Shelf detailed in new maps. J. Palaeogeogr. 14 (1), 186–202.

Crandall, E.D., Riginos, C., Bird, C.E., Liggins, L., Treml, E., Beger, M., Barber, P.H., Connolly, S.R., Cowman, P.F., DiBattista, J.D., Eble, J.A., Magnuson, S.F., Horne, J. B., Kochzius, M., Lessios, H.A., Liu, S.Y.V., Ludt, W.B., Madduppa, H., Pandolfi, J.M., Toonen, R.J., Gaither, M.R., 2019. The molecular biogeography of the Indo-Pacific: testing hypotheses with multispecies genetic patterns. Glob. Ecol. Biogeogr. 28 (7), 943–960.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27 (15), 2156–2158.

Davey, J.L., Blaxter, M.W., 2010. RADSeq: next-generation population genetics. Brief. Funct. Genomics 9 (5–6), 416–423.

Díaz-Jaimes, P., Bayona-Vasquez, N., Hinojosa-Alvarez, S., Uribe-Alcocer, M., Marcet-Houben, M., 2015. The complete mitogenome of the common dolphinfish (*Coryphaena hippurus*). Mitochondrial DNA 26 (6).

Díaz-Jaimes, P., Uribe-Alcocer, M., Rocha-Olivares, A., García-de-León, F.J., Nortmoon, P., Durand, J.D., 2010. Global phylogeography of the dolphinfish (*Coryphaena hippurus*): the influence of large effective population size and recent dispersal on the divergence of a marine pelagic cosmopolitan species. Mol. Phylogenet. Evol. 57 (3), 1209–1218.

Domingues, V.S., Santos, R.S., Brito, A., Alexandrou, M., Almada, V.C., 2007. Mitochondrial and nuclear markers reveal isolation by distance and effects of Pleistocene glaciations in the northeastern Atlantic and Mediterranean populations of the white seabream (*Diplodus sargus*, L.). J. Exp. Mar. Biol. Ecol. 346 (1–2), 102–113. https://doi.org/10.1016/J.JEMBE.2007.03.002.

Dray, S., Dufour, A.B., 2007. The ade4 package: Implementing the duality diagram for ecologists. J. Stat. Softw. 22 (4).

Ely, B., Viñas, J., Alvarado Bremer, J.R., Black, D., Lucas, L., Covello, K., Labrie, A.V., Thelen, E., 2005. Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (*Thunnus albacares*) and the skipjack tuna (Katsuwonus pelamis). BMC Evol. Biol. 5 (1), 19.

Fiedler, P.C., Lavín, M.F., 2017. Oceanographic conditions of the eastern tropical pacific. Coral Reefs of the World 8, 59–83.

Francis, R.M., 2017. *pophelper: an R package and web app to analyse and visualize population structure*. Mol. Ecol. Resour. vol. 17.

Gaither, M.R. & Rocha, L.A., 2013, 'Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: Evidence for the centre of overlap Grewe.

Goudet, J., 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5 (1).

Gozzer, 2015. R.,2015, 'Diversidad genética y estructura poblacional del perico (*Coryphaena hippurus,* Linnaeus1758) en el Pacífico Peruano' – PhD thesis, Tesis de licenciatura en Ingenieria Pesquera. Lima: Universidad Nacional.

Graco, M.I., Purca, S., Dewitte, B., Castro, C.G., Morón, O., Ledesma, J., Flores, G., Gutiérrez, D., 2017. The OMZ and nutrient features as a signature of interannual and low-frequency variability in the peruvian upwelling system. Biogeosciences 14 (20), 4601–4617. https://doi.org/10.5194/BG-14-4601-2017.

Grewe P.M., Feutry, P., Hill, P.L., Gunasekera, R.M., Schaefer, K.M., Itano, D.G., Fuller, D.W., Foster, S.D. & Davies, C.R., 2015, 'Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource', *Scientific Reports 2015 5:1*, 5(1), 16916.

Haro-Bilbao, I., Riginos, C., Baldwin, J.D., Zischke, M., Tibbetts, I.R., Thia, J.A., 2021. Global connections with some genomic differentiation occur between Indo-Pacific and Atlantic Ocean wahoo, a large circumtropical pelagic fish. J. Biogeogr. 48 (8), 2053–2067.

Haug, G.H., Tiedemann, R., 1998. Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. Nature 393 (6686), 673–676.

Hinojosa-Alvarez, S., Mendoza-Portillo, V., Chavez-Santoscoy, R.A., Hernández-Pérez, J., Felix-Ceniceros, A., Magallón-Gayón, E., Mar-Silva, A.F., Ochoa-Zavala, M., Díaz-Jaimes, P., 2025. The draft genome assembly of the cosmopolitan pelagic fish dolphinfish *Coryphaena hippurus*. G3 Genes|genomes|genetics 15 (5), 59.

Husson, L., Boucher, F.C., Sarr, A.C., Sepulchre, P., Cahyarini, S.Y., 2020. Evidence of Sundaland's subsidence requires revisiting its biogeography. J. Biogeogr. 47 (4), 843–853.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24 (11), 1403–1405.

Kamvar, Z.N., Brooks, J.C., Grünwald, N.J., 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. 6 (JUN).

Katoh K, Standley DM. 2013. 'MAFFT multiple sequence alignment software version 7: improvements in performance and usability', *Molecular Biology Evolution,* Apr;30(4): 772-80. doi: 10.1093/molbev/mst010. Epub2013 Jan 16. PMID: 23329690; PMCID: PMC3603318.

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinf. 15 (1), 1–13.

Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J. & Wilson, D.S., 1999, 'Chronology, causes and progression of the Messinian salinity crisis', *Nature 1999 400:6745*, 400 (6745), 652–655.

Krijgsman, W., Rohling, E.J., Palcu, D.V., Raad, F., Amarathunga, U., Flecker, R., Florindo, F., Roberts, A.P., Sierro, F.J., Aloisi, G., 2024. Causes and consequences of the Messinian salinity crisis. Nat. Rev. Earth Environ. 5 (5), 335–350. https://doi.org/10.1038/s43017-024-00533-1.

Lasso, J., Zapata, L., 1999. Fisheries and biology of *Coryphaena hippurus* (Pisces: Coryphaenidae) in the pacific coast of colombia and panama. Sci. Mar. 63 (3–4), 387–399.

Leigh, J.W., Bryant, D., Nakagawa, S., 2015. POPART: full-feature software for haplotype network construction. Methods in Ecology & Evolution 6 (9), 1110–1116.

Leone, A., Puncher, G.N., Ferretti, F., Sperone, E., Tripepi, S., Micarelli, P., Tinti, F., 2020. Pliocene colonization of the Mediterranean by Great White Shark inferred from fossil records, historical jaws, phylogeographic and divergence time analyses. J. Biogeogr. 47 (5), 1119–1129.

Lessios, H.A., 2008, 'The great American schism: Divergence of marine organisms after the rise of the Central american Isthmus', *Annual Review of Ecology, Evolution, and Systematics*, 39(Volume 39, 2008), 63–91.

Lou, R.N., Jacobs, A., Wilder, A.P., Therkildsen, N.O., 2021. A beginner's guide to low-coverage whole genome sequencing for population genomics. Mol. Ecol. 30 (23), 5966–5993.

Ludt, W.B., Rocha, L.A., 2015. Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. J. Biogeogr. 42 (1), 25–38. https://doi.org/10.1111/JBI.12416.

Maggio, T., Allegra, A., Andaloro, F., Pedro Barreiros, J., Battaglia, P., Butler, C.M., Cuttitta, A., Fontes, M.R.J., Freitas, R., Gatt, M., Karakulak, F.S., Macias, D., Nicosia, A., Oxenford, H.A., Saber, S., Vasco Rodrigues, N., Yildiz, T., Sinopoli, M., 2019. Historical separation and present-day structure of common dolphinfish (*Coryphaena hippurus*) populations in the Atlantic Ocean and Mediterranean Sea. ICES J. Mar. Sci. 76 (4), 1028–1038.

Malinsky, M., Matschiner, M., Svardal, H., 2021. Dsuite - Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21 (2), 584–595. https://doi.org/10.1111/1755-0998.13265.

Mar-Silva, A.F., Díaz-Jaimes, P., Ochoa-Zavala, M., Ortega-Garcia, S., Sotil, G., Alegre, A., Escarabay, T., Elias, E., Carvajal-Rodríguez, J.M., 2023. A genomic approach for the identification of population management units for the dolphinfish *(Coryphaena hippurus)* in the eastern Pacific. Front. Mar. Sci. 10, 1294509.

Maroso, F., Franch, R., Dalla Rovere, G., Arculeo, M., Bargelloni, L., 2016. RAD SNP markers as a tool for conservation of dolphinfish Coryphaena hippurus in the Mediterranean Sea: Identification of subtle genetic structure and assessment of populations sex-ratios. Mar. Genomics 28, 57–62.

Massing, J.C., Schukat, A., Auel, H., Auch, D., Kittu, L., Pinedo Arteaga, E.L., Correa Acosta, J., Hagen, W., 2022. Toward a solution of the "Peruvian Puzzle": pelagic food-web structure and trophic interactions in the northern humboldt current upwelling system off peru. Front. Mar. Sci. 8, 759603. https://doi.org/10.3389/FMARS.2021.759603/XML.

Meisner, J., Albrechtsen, A., Hanghøj, K., 2021. Detecting selection in low-coverage high-throughput sequencing data using principal component analysis. BMC Bioinf. 22 (1).

Mendoza-Portillo, V., Jaimes, P.D., Mar-Silva, A.F., Arculeo, M., Ochoa-Zavala, M., Vecchioni, L., Adams, D.H., 2026. Seascape genomics uncovers genomic patterns shaped by temperature and salinity outlier loci in dolphinfish (*Coryphaena hippurus Linnaeus, 1758*) across the atlantic ocean and mediterranean sea. Aquat. Conserv. Mar. Freshwat. Ecosyst. 36 (2), e70334.

Morel, A., Claustre, H., Antoine, D., & Gentili, B. 2007. 'Natural variability of bio-optical properties in Case 1 waters: attenuation and reflectance within the visible and near-UV spectral domains, as observed in South Pacific and Mediterranean waters', *Biogeosciences,* 4, 913–925. www.biogeosciences.net/4/913/2007/.

Ochoa-Zavala, M., Diaz-Jaimes, P., Ortega-García, S., Galván-Magaña, F., 2022. Genetic divergence at species boundaries of the dolphinfish (*Coryphaenahippurus*) in the Tropical Eastern Pacific. PeerJ 10, e14389.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L. & Solymos, P., 2017, 'others (2019) vegan: community ecology package. R package version 2.5-6', *Community Ecol package, version*.

Oxenford, H.A., 1999. Biology of the dolphinfish (*Coryphaena hippurus*) in the western central Atlantic: a review. Sci. Mar. 63 (3–4), 277–301.

Palko, B.J., Beardsley, G.L., Richards, W.J., 1982. Synopsis of the biological data on dolphin-fishes, *Coryphaena hippurus* Linnaeus and *Coryphaena equiselis* Linnaeus. FAO Fisheries Synopsis Rep 130 (1).

Paradis, E., Schliep, K., 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35 (3).

Paradis, E., 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26 (3), 419–420.

Patarnello, T., Volckaert, F.A.M.J., Castilho, R., 2007. Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? Mol. Ecol. 16 (21), 4426–4444.

Pecoraro, C., Babbucci, M., Franch, R., Rico, C., Papetti, C., Chassot, E., Bodin, N., Cariani, A., Bargelloni, L. & Tinti, F., 2018, 'The population genomics of yellowfin tuna (*Thunnus albacares*) at global geographic scale challenges current stock delineation', *Scientific Reports 2018 8:1*, 8(1), 1–10.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo snp discovery and genotyping in model and non-model species. PLoS One 7 (5), e37135.

Pickrell, J., Pritchard, J., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8 (11), e1002967. https://doi.org/10.1038/npre.2012.6956.1.

R Core Team, 2014, 'R Core Team (2014). R: A language and environment for statistical computing.', *R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/*.

Randall, J.E., 1998, 'Zoogeography of Shore Fishes of the Indo-Pacific Region', vol. 37.

Reid, K., Hoareau, T., Graves, J., et al., 2016. Secondary contact and asymmetrical gene flow in a cosmopolitan marine fish across the Benguela upwelling zone. Heredity 117, 307–315. https://doi.org/10.1038/hdy.2016.51.

Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3 (2), 217–223. https://doi.org/10.1111/J.2041-210X.2011.00169.X.

Riccioni, G., Stagioni, M., Landi, M., Ferrara, G., Barbujani, G., Tinti, F., 2013. Genetic structure of bluefin tuna in the mediterranean sea correlates with environmental variables. PLoS One 8 (11), e80105.

Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53 (1–2), 131–147. https://doi.org/10.1016/0025-5564(81)90043-2.

Sarr, A.C., Husson, L., Sepulchre, P., Pastier, A.M., Pedoja, K., Elliot, M., Arias-Ruiz, C., Solihuddin, T., Aribowo, S., Susilohadi, 2019. Subsiding sundaland. Geology 47 (2), 119–122.

da Silva Ferrette, B.L., Coimbra, R.T.F., Winter, S., Jong, M.J.D., Williams, S.M., Coelho, R., Rosa, D., Rotundo, M.M., Arocha, F., Mourato, B.L., Mendonça, F.F.,

Janke, A., 2023. Seascape Genomics and Phylogeography of the Sailfish (*Istiophorus platypterus*). Genome Biol. Evol. 15 (4).
Skotte, L., Korneliussen, T.S., Albrechtsen, A., 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics 195 (3), 693–702.
Teske, P.R., Heyden, S. Von Der, Mcquaid, C.D., Barker, N.P. & Teske, P., 2011, 'A review of marine phylogeography in southern Africa', *S Afr J Sci*, 107(5).
Tine, M., Kuhl, H., Gagnaire, P.A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F.A.M., Bargelloni, L., Power, D.M., Bonhomme, F., Canario, A. V.M. & Reinhardt, R., 2014, 'European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation', *Nature Communications 2014 5:1*, 5(1), 5770-.
Toews, D.P., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21 (16), 3907–3930.
Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W., Corander, J., 2018. RhierBAPs: an R implementation of the population clustering algorithm hierbaps. Wellcome Open Research 3.
Toonen, R.J., Bowen, B.W., Iacchei, M. & Briggs, J.C., 2016, 'Biogeography, Marine', *Encyclopedia of Evolutionary Biology*, 166–178.
Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. J. Biogeogr. 27 (5), 1153–1167.
Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38 (6).
Wong, T. K. F., Ly-Trong, N., Ren, H., Baños, H., Roger, A. J., Susko, E., Bielow, C., Maio, N. de, Goldman, N., Hahn, M. W., Huttley, G., Lanfear, R., & Minh, B. Q. 2025., 'IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models'. 10.32942/X2P62N.
Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28 (24).