Turk. J. Fish.& Aquat. Sci. 25(5), TRJFAS27385
https://doi.org/10.4194/TRJFAS27385

RESEARCH PAPER

Turkish Journal of FISHERIES and AQUATIC SCIENCES

Genome-wide SNP Detection in the Horse Mackerel (*Trachurus trachurus, T. mediterraneus, T. picturatus*) of Turkish Waters Using Illumina and Nanopore Sequencing

Zehra Duygu Düzgüneş[1,2,*], Yahya Tayhan[3], Serkan Gül[1], Yusuf Bektaş[4]

[1]Recep Tayyip Erdoğan University, Faculty of Arts and Sciences, Department of Biology, Rize, Türkiye.
[2]Central Fisheries Research Institute, Department of Breeding and Genetics, Trabzon, Türkiye.
[3]Tokat Gaziosmanpaşa University, Almus Vocational School, Department of Forestry, Tokat, Türkiye.
[4]Karadeniz Technical University, Faculty of Science, Department of Molecular Biology and Genetics, Trabzon, Türkiye.

How to Cite

Duzgunes, Z.D., Tayhan, Y., Gul, S., Bektas, Y. (2025). Genome-wide SNP Detection in the Horse Mackerel (*Trachurus trachurus, T. mediterraneus, T. picturatus*) of Turkish Waters Using Illumina and Nanopore Sequencing. *Turkish Journal of Fisheries and Aquatic Sciences*, *25(5)*, *TRJFAS27385*. https://doi.org/10.4194/TRJFAS27385

Article History
Received 27 November 2024
Accepted 02 January 2025
First Online 10 January 2025

Corresponding Author
E-mail: zehraduygu_83@hotmail.com

Keywords
SNP
Next-generation sequencing
Oxford nanopore sequencing
Trachurus species
Whole genom sequencing

Abstract

The horse mackerel, distributed in the Pacific, Atlantic, Indian Ocean, and Mediterranean waters, is economically valuable due to its high nutritional content and fishing capacity. Global market demand for this species has been increasing over the years. This rising demand causes fishing pressure, endangering the species' sustainability. Especially the horse mackerel species (*Trachurus trachurus, Trachurus mediterraneus and Trachurus picturatus*), which are found as mixed stocks in Turkish waters and under fishing pressure, should be considered as different stocks and evaluated with distinct management strategies. However, due to the difficulty of distinguishing these three *Trachurus* species based on morphological characteristics, conflicting reports have emerged, particularly regarding their distribution areas, and species identification is not always precise. In our study, we sequenced the genomes of these species using next-generation sequencing (Illumina) and third-generation sequencing technologies (Oxford Nanopore), and identified SNP (single nucleotide polymorphism) differences to distinguish the species. We identified a total of 2311386 SNPs in the *Trachurus trachurus* genome, of which 1313612 were transitions and 997774 were transversions. 5099766 SNPs in the *Trachurus picturatus* genome, of which 2927066 transitions and 2172700 transversions. Similarly, 5007863 SNPs in the *Trachurus mediterraneus* genome, of which 2940560 transitions and 2067303 transversions. These SNP datasets provide a robust genetic marker toolkit for accurately distinguishing *Trachurus* species and evaluating genetic diversity within and between populations. The findings of this study contribute essential genetic resources for defining the stock structures of these species, informing sustainable fisheries management strategies, and guiding conservation initiatives. By addressing the challenges of species identification and genetic diversity assessment, this research lays the groundwork for mitigating the impact of overfishing and ensuring the sustainability of horse mackerel populations in Turkish and global waters.

Introduction

Horse mackerels, a small pelagic migratory species belonging to the family Carangidae (Perciformes), are widely distributed throughout coastal and oceanic waters of temperate, tropical and subtropical seas (Bektaş and Beldüz 2008). The horse mackerel, caught in Turkish waters, is an economically important species for Türkiye, as it is worldwide, and constitutes approximately 5.2% of the seafood harvested in Türkiye. The genus *Trachurus* is represented by three species in Turkish waters: the Atlantic Horse Mackerel, *Trachurus trachurus* (Linnaeus, 1758), the Mediterranean Horse Mackerel, *Trachurus mediterraneus* (Steindachner, 1868), and the Blue Jack Horse Mackerel, *Trachurus picturatus* (Bowdich, 1825) (Bostancı 2009). In terms of fishing, only two species are commercially important, especially in the Black Sea and the north eastern Mediterranean - *Trachurus trachurus* and *Trachurus mediterraneus*- (Bilecenoğlu et al., 2014; Melnikova and Kuzminova 2020). The Mediterranean General Fisheries Commission (2019) reports that these two species were

among the overfished stocks. In fact, according to the FAO annual report (2022), it is thought that the stocks of *T. trachurus* in the Mediterranean, Alboran and Black Seas have been fully exploited. Three Horse mackerel species (*T. trachurus, T. mediterraneus and T. picturatus*), which are distributed in Turkish waters and whose stocks are decreasing, have been the subject of many studies on bioecological fishing and genetics (Yankova et al*.,* 2010; Aydın and Karadurmuş, 2012; Bat et al*.,* 2013; Bektaş and Beldüz 2008; Turan, 2009). However, these studies are insufficient to establish a genetic approach to stock management for the sustainability of horse mackerel populations in Turkish seas.

Because there are no clear restricting gene flow in marine species, genetic differentiation and local adaptation are difficult to determine. However, genome-wide association studies have confirmed that variation in many important traits is controlled by a large number of loci distributed across the genome (Atwell et al*.,* 2010, Jumbo-Lucioni et al*.,* 2010) and that SNPs can determine a large proportion of adaptive differentiation among populations if all SNPs across the genome are considered (Yang et al*.,* 2010). In recent years, advances in genetic techniques have enabled the use of new technologies in adaptation and stock assessment studies and facilitated the discovery of new genetic tools (Jansen et al*.,* 2017; Tan et al*.,* 2018, Zhao et al., 2022; Shi et al*.,* 2023). In particular, NGS technologies have contributed to the development of polymorphic microsatellites and SNP biomarkers in many marine and freshwater fishes, enabling the detection of mutations that cause genome scale variation in non-model fish species (Vera et al*.,* 2013; Yu, 2019; Santos et al*.,* 2015; Villanova et al*.,* 2015; Xiao et al*.,* 2015; Yang, 2020).

A new era has begun in sequencing technologies with the development of third-generation sequencing technologies, which are a superior platform in every respect, enabling the production of complete and accurate genomes with long-read sequences (Bian et al., 2020). Oxford Nanopore, a single-molecule sequencing method, has become one of the most powerful methods for the rapid generation of from long read sequences and has the potential to create high-quality genome assemblies. Nanopore sequencing has been used in several fish species to generate high-quality genome assemblies or to increase the completeness of previous genome drafts (Austin et al*.,* 2017; Ge et al*.,* 2019; Jansen et al*.,* 2017; Tan et al*.,* 2018; Zhao et al*.,* 2022; Bian et al*.,* 2020). Recent studies have focused on taxonomic resolution, particularly at the species level, and the investigation of rare taxa, demonstrating the potential of Nanopore sequencing for species identification (Szoboszlay et al., 2023; Angell et al*.,* 2020).

Genetic markers preferred for genetically identifying stocks have developed rapidly and single nucleotide polymorphisms (SNPs) have become quite popular, replacing allozymes and microsatellites (Schlötterer, 2004; Hauser et al*.,* 2008). Compared to microsatellites, SNPs have rapidly increased in use because they are faster to develop, easier to identify, and the resulting genotypes can be easily transferred between laboratories (Seeb et al*.,* 2011). Whole genome sequencing, one of the next-generation sequencing techniques that enable the discovery of SNP variations, is an essential genetic tool for the protection and sustainability of fish stocks. Using next-generation sequencing, a large number of SNPs are being discovered, representing an effective approach to facilitate ecological, population and conservation genetics, and strengthening the applications of future studies (Elshire et al*.,* 2011; Glaubitz et al*.,* 2014; Narum et al*.,* 2013).

This study aimed to identify single nucleotide polymorphisms (SNPs) at the chromosomal level among three *Trachurus* species through whole genome sequencing using Oxford Nanopore and Illumina next-generation sequencing technologies. By developing SNP markers specific to the SNP regions identified in this study, rapid and precise identification of these three *Trachurus* species will be facilitated in future studies. Besides, the high-quality genomes we obtained can be used as a reference genome for the assessment of genetic diversity within and between populations. Moreover, this study will provide a critical genomic resource that can support studies on genetic evolution and molecular mechanisms not only within the *Trachurus* genus but also for other fish species.

Materials and Methods

Fish Sample Preparation and Genomic DNA Isolation

Samples belonging to three species were collected from purse seine boats during the fishing season at the designated stations (*Trachurus mediterraneus*/ Black Sea; *Trachurus trachurus*/ Mediterrean Sea; *Trachurus picturatus* /Mediterrean Sea). Samples and sampling areas of the three *Trachurus* species are shown in Figure 1 and Figure 2. Caudal fin tissues were immediately clipped from dead fishes on board and preserved in Eppendorf tube include 96% ethanol. The genomic DNA was isolated using Qiagen DNeasy® Blood & Tissue kit and NanoDrop 2000 Spectrofotometer (ThermoFisher Scientific, cat # ND-2000) and %0.8 electrophoresis agarose gel were used for DNA purity. DNA was quantified using Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, Cat. # Q32851) and was used to generate the whole genome sequence data (Figure 3).

DNA Library Preparation and Sequencing

A gene pool was created for each species-containing five samples each- and a nanopore DNA library containing approximately 20 kb DNA size was prepared. For this purpose, DNAs of appropriate size

Figure 1. Samples of *Trachurus mediterraneus, Trachurus trachurus* and *Trachurus picturatus*.

Figure 2. Collection site of fish samples (Black Sea; 1 was represented as 41.171728, 39.750023; Mediterranean Sea; 2 was represented as 36.844541 30.687879, 3 was represented as 36.586091 36.002311).

were selected by agarose gel electrophoresis. The isolated dsDNA was prepared for sequencing using the Oxford Nanopore SQK-NBD114.24 kit according to the procedure recommended by the company. NEBNext Ultra II End Repair/dA-Tailing, NEBNext FFPE DNA Repair Mixes were used to repair DNA damage and breaks. During library preparation, DNA purification was performed with Agencourt AMPure XP beads (A63880, Beckman Coulter). After the Native barcode ligation and Adapter ligation processes, DNA purification was performed again. Each sample for three species was subjected to nanopore sequencing for 48 hours using the R10.4.1 PromethION FLO-PRO114M flow cell.

Through the MinKNOW program, which is the control software of the sequencing device, nanopore data were obtained in the pod5 format, a file format developed by Oxford Nanopore that replaced the old fast5 format. Barcode and adapter sequences were trimmed with ONT Dorado (v.0.5.1, Oxford Nanopore Technologies) software. Basecalling results were converted to BAM and FASTQ file format using the high-accuracy model with the latest version of ONT Dorado software (v.0.5.1, Oxford Nanopore Technologies). The raw data were filtered to remove low-quality reads using FastQC (Galaxy Version 0.74).

For Illumina sequencing library, the genomic DNA was randomly sheared into shorter fragments (2x150bp). The fragments were end-repair, adding A-tailed and ligated with Illumina adapters, respectively. The library was quantified through Qubit and qPCR and then size distribution detected by fragment analyser. Quantified libraries were pooled and sequenced on Illumina platforms NovaSeq PE150. The original raw data from the Illumina platform was converted into Sequence Reads by Novogene Co., Ltd. and provided to us as a FASTQ file containing the sequencing reads. The raw data were filtered to remove low-quality reads using FastQC. Barcode and adapter sequences were trimmed with Trimmomatic V0.32. For short reads, SAM files were converted to BAM files with Minimap2 (v.2.24), and Nanopore and Illumina sequencing data were combined for *Trachurus mediterraneus*.

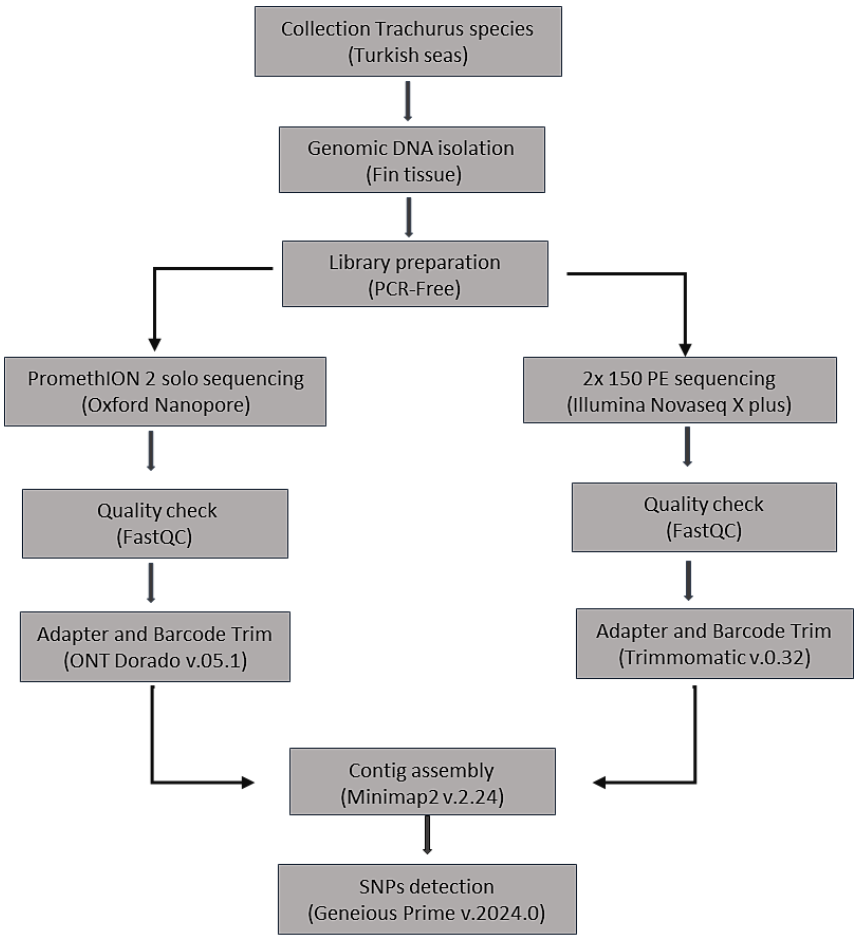

Figure 3. Schematic diagram of the whole genome sequencing methodology and SNP identification.

Genome Assembly and SNP Detection Using Reference Genome

All reads were aligned to the *Trachurus trachurus* reference genome downloaded from the NCBI database (Genome assembly fTraTra1.2, Genbank ID: GCA_905171665.1) using Minimap2 (v.2.24) and consensuses were created. All assembled reads were imported into Geneious Prime (v.2024.0) for further analysis. Variants were detected with the "Find Variations/SNPs" command in Geneious Prime (v.2024.0). While running this command, the minimum coverage value to find variants was taken as five and the minimum variant frequency was taken as 0.70.

Result

Genome Sequencing

We sequenced the genome of *T. trachurus, T. picturatus* and *T. mediterraneus* using Oxford Nanopore sequencing. Three Flow Cells yielded 26.91 Gb of data total, of which 18.6 Gb was collected after filtering. Each one *Trachurus* species N50 length was 2.29 kb, 2.33kb and 1.24 kb, respectively (Table 1). The quality Q-score was a minimum value of 9. Reads that fall below the value of 9 classified as failed reads (Figure 4). For the *Trachurus mediterraneus* for which we could not obtain sufficient data from the Oxford Nanopore Technology, sequencing was performed using the Illumina NovaSeq X Plus technology. We obtained a total of 8.2 Gb raw data with 90.6% of Q30 and 43% of GC content (Figure 5). A total of 26.8 Gb of filtered sequences were generated from both technologies.

Genome Assembly

Reads were aligned to the *Trachurus trachurus* reference genome (Genome assembly fTraTra1.2, Genbank ID: GCA_905171665.1), after raw reads quality controlled. After the final assembly, we determined that *Trachurus trachurus* was a total sequence length 1051.7 Mb with maximum sequence length 230.8 kb, *Trachurus picturatus* was a total sequence length 1080.9 Mb with maximum sequence length 205.1 kb and *Trachurus mediterraneus* was a total sequence length 838.7 Mb with maximum sequence length 107.6 kb. The majority, 99.5 % of the assembly sequence was assigned to 24 chromosomal-level scaffolds for *Trachurus trachurus*. 94.53% and 98.9% of the assembly sequence was assigned to 24 chromosomal-level scaffolds for *Trachurus mediterraneus* and *Trachurus picturatus,* respectively (Table 2)

Table 1. Statistic of the sequencing data

Species	Method	Sequencing Platform	Library Size bp	Clean data Gb	Total Read kb	Passed Read%	N50kb
Trachurus trachurus	Nanopore	PromethION	20000	8.9	6564.03	92.1%	2.29
Trachurus picturatus	Nanopore	PromethION	20000	8.9	6892.35	93.2%	2.33
Trachurus mediterraneus	Nanopore	PromethION	1000	0.8	796.18	76%	1.24
Trachurus mediterraneus	Illumina	NovaSeq X Plus Series (PE150)	150	8.2	54834.60	98.37%	-

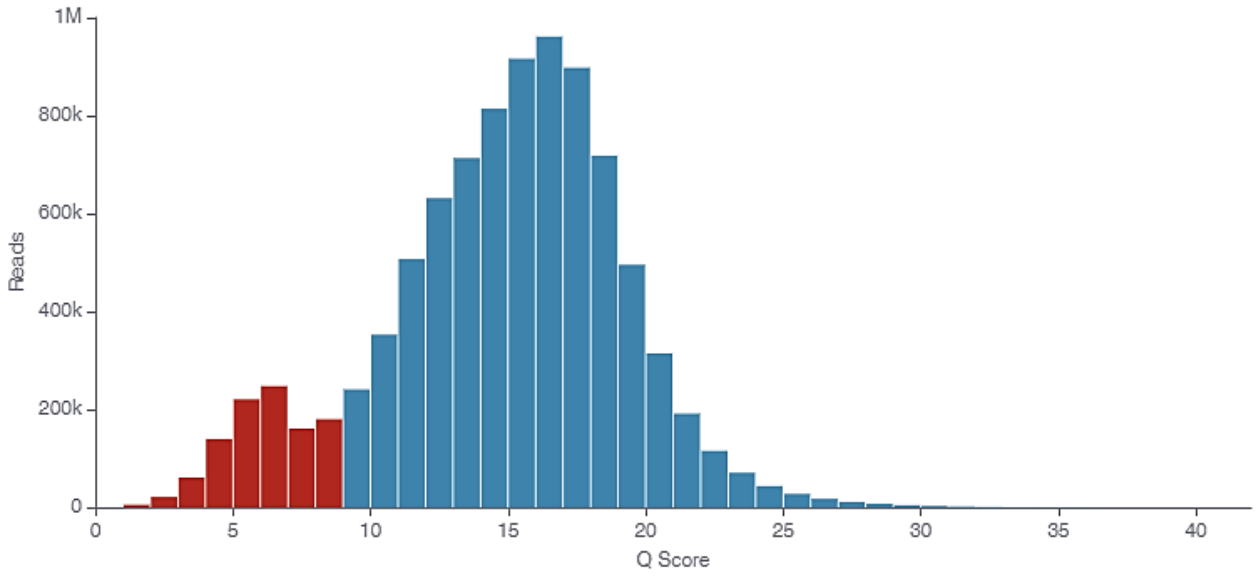

Figure 4. The quality score in Oxford Nanaopore sequencing.

SNP Discovery

To determine the distribution of SNPs in the *Trachurus* species in Türkiye seas, we made use of *Trachurus trachurus* in to the GeneBank accession assembly fTraTra1.2 Genbank ID: GCA_905171665.1. We identified a total of 2311386 SNPs in *Trachurus trachurus* genome of which 1313612 were transition and 997774 were transversion; 5099766 SNPs in *Trachurus picturatus* genome of which 2927066 were transition and 2172700 were transversion; 5007863 SNPs in *Trachurus mediterraneus* genome of which 2940560 were transition and 2067303 were transversion. The total number of insertions detected for *Trachurus trachurus, Trachurus picturatus* and *Trachurus mediterraneus* were 458990, 1126362, 905097 and the total number of deletions were 591904, 1352994, 1040946, respectively (Table 3).

The relationship between the number of SNPs within chromosomes and the total chromosome size in terms of Mb is shown in Figure 6 The correlation coefficient (r) between number of SNPs and chromosome size 0.92 in *Trachurus mediterraneus*. 0.91, 0.84 in *Trachurus picturatus* and *Trachurus trachurus,* respectively. The detected SNPs present an even distribution across the chromosomes on three *Trachurus* species genome assembly.

Discussion

The morphological identification of *Trachurus* species is quite difficult due to overlapping physical characteristics, leading to frequent misidentification and the use of incorrect species names in many studies (Smith et al., 2008). This misidentification issue has in several studies (Kalayci et al*.,* 2010; Aydın and Karadurmuş, 2012; Durmuş, 2019), where contradictory statements regarding species distribution and biology are prevalent. These discrepancies underscore the limitations of traditional morphological methods and point to the need for more accurate, molecular-based species identification to avoid confusion. The inability to reliably distinguish between *Trachurus* species has significant consequences for the management of fish stocks. Misidentification can lead to inappropriate stock assessments, which, in turn, results in the application of incorrect management strategies. This issue is particularly critical for the horse mackerel species in the Turkish waters, where differences in reproductive lengths (GFCM 2019) and conservation status (with *Trachurus trachurus* listed as vulnerable on the IUCN Red List; Natale et al., 2011) suggest distinct ecological needs and management requirements. Furthermore, the practice of grouping three *Trachurus* species under a single stock is problematic, as it does not account for the species-specific variations in biology and behavior, thereby compromising efforts to implement effective and sustainable fisheries management. The genetic

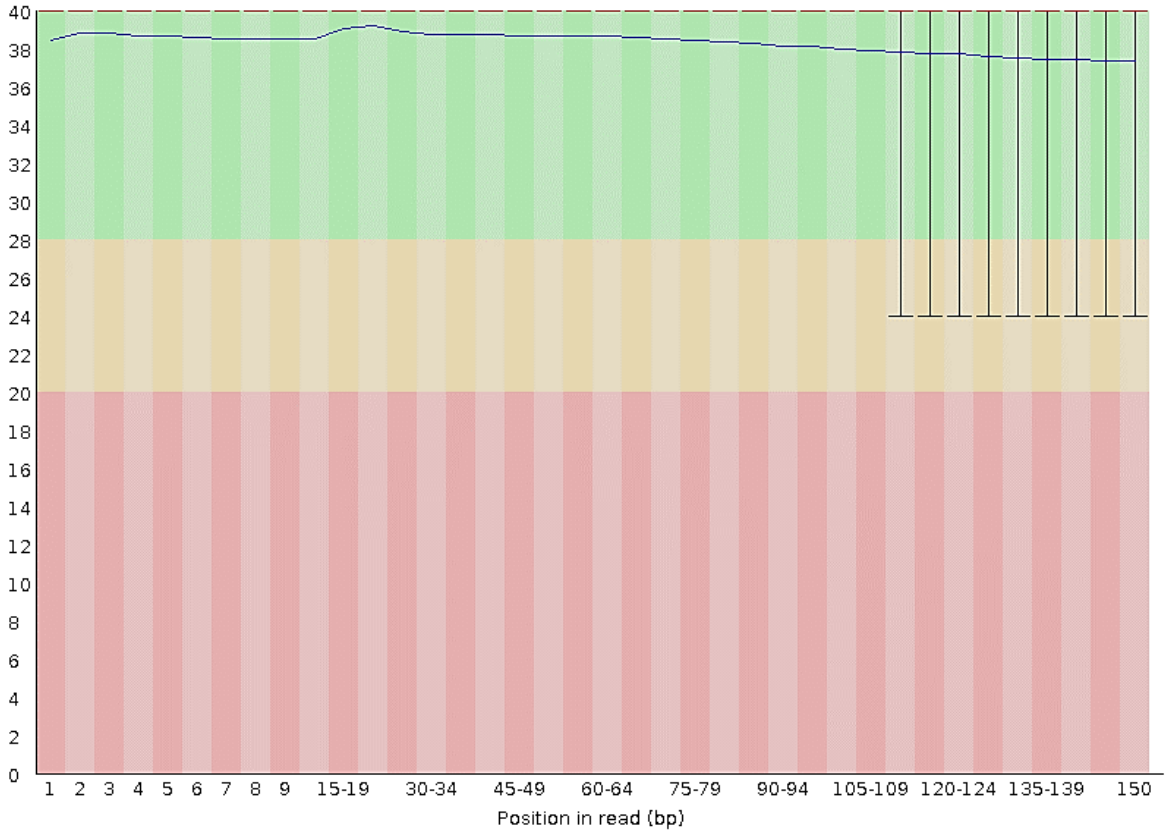

Figure 5. The quality score in Illumina sequencing.

Table 2. Statistics of genome assembly of *Trachurus trachurus*, *Trachurus picturatus* and *Trachurus mediterraneus*

	Trachurus trachurus		***Trachurus mediterraneus***		***Trachurus picturatus***	
	Chromosome	Unplaced scaffold	Chromosome	Unplaced scaffold	Chromosome	Unplaced scaffold
Sequence lenght Mb	1026.66	25.07	827.53	11.17	1057.99	22.91
Ref seq lenght(Mb)	790.72	10.5	790.72	10.5	790.72	10.5
Number of sequence	1716158	1198936	53825885	919677	19128819	896708
Identical sites%	47.15	16.37	75.83	50.18	44.9	37.535
Of ref seq %	99.5	94.56	94.53	56.94	98.9	84.61
GC %	41.9	40.86	42.8	44.31	42	41.74
Max seq lenght(kb)	230.8	119.9	107.6	14.6	205.1	29.5

Sequence lenght : Residue lenght of the sequence, Ref seq lenght : The lenght of the reference sequence in a contig, Number of sequence: The number of sequences in an alignmentor set of sequences, Identical sites: The percentage of columns in the alignment for which all sequence are identical, Of ref seq: The percentage of the reference sequence in a contig that is covered by one or more reads, GC : The percentage of A,C,G,T,U,S,W nucleotides (excluding reference sequence)that are either G,C or S, Max seq lenght :The maximum lenght of non-reference sequences in a sequence list or alignment.

Table 3. Single nucleotide polymorphism (SNP) in *Trachurus trachurus*, *Trachurus picturatus* and *Trachurus mediterraneus*

SNPs		***Trachurus trachurus***		***Trachurus picturatus***		***Trachurus mediterraneus***	
		Chomosommal-level scaffold	Unplaced Scaffold	Chomosommal-level scaffold	Unplaced Scaffold	Chomosommal-level scaffold	Unplaced Scaffold
SNP (transition)	A -> G	350153	2761	745887	5905	726727	2434
	C -> T	302566	2623	709625	5404	739686	2473
	G -> A	300632	2772	704115	6253	740045	2487
	T -> C	349587	2518	744721	5156	724303	2405
SNP (transversion)	A -> C	135092	1069	285610	2275	266343	1051
	A -> T	142715	1289	309337	2649	283982	1097
	C -> A	124053	1111	270157	2361	268725	968
	C -> G	92279	935	210940	1855	215095	989
	G -> C	91848	864	210545	1811	213508	1014
	G -> T	125117	1075	274165	2066	270338	951
	T -> A	142648	1342	308121	2509	274685	1103
	T -> G	135170	1167	285859	2440	266404	1050
insertion		450850	8140	1113196	13166	894374	10723
deletion		584091	7813	1340920	12074	1031103	9843

diversity within *Trachurus* species, particularly in the context of mismanagement due to species confusion, poses a long-term threat to the conservation of genetic resources (FAO, 2022). As such, the continued practice of lumping these species together in stock assessments will likely lead to overexploitation, and compromise sustainability of these populations. In light of these challenges, it is essential to develop and apply new genetic approaches for species identification and stock assessment. The use of molecular markers, such as SNPs, can contribute to species identification and determine mixing ratios in areas where Trachurus species are mixed, allowing for accurate stock assessments and species-specific management strategies. Many studies have suggested that SNP panels or SNP assays could contribute to fisheries management strategies.(Elliott et al., 2018; Beacham et al., 2020; Attard et al., 2022) Supportively, Garner et al. (2016) have highlighted that conservation practitioners benefit from genomic research than is often apparent and that increased collaboration between academics and management agencies is needed. The use of next-generation sequencing (NGS) has enabled researchers to obtain more data than traditional genetic markers and has unlocked many studies (Kumar et al., 2016). Third-generation DNA sequencing (TGS) technologies in particular have provided new opportunities for genome sequencing studies by providing an innovative, efficient and less costly basis for existing sequencing technologies (Yang et al., 2020). Using third-generation sequencing technology, the Oxford Nanopore platform has emerged as a promising platform for studying fish

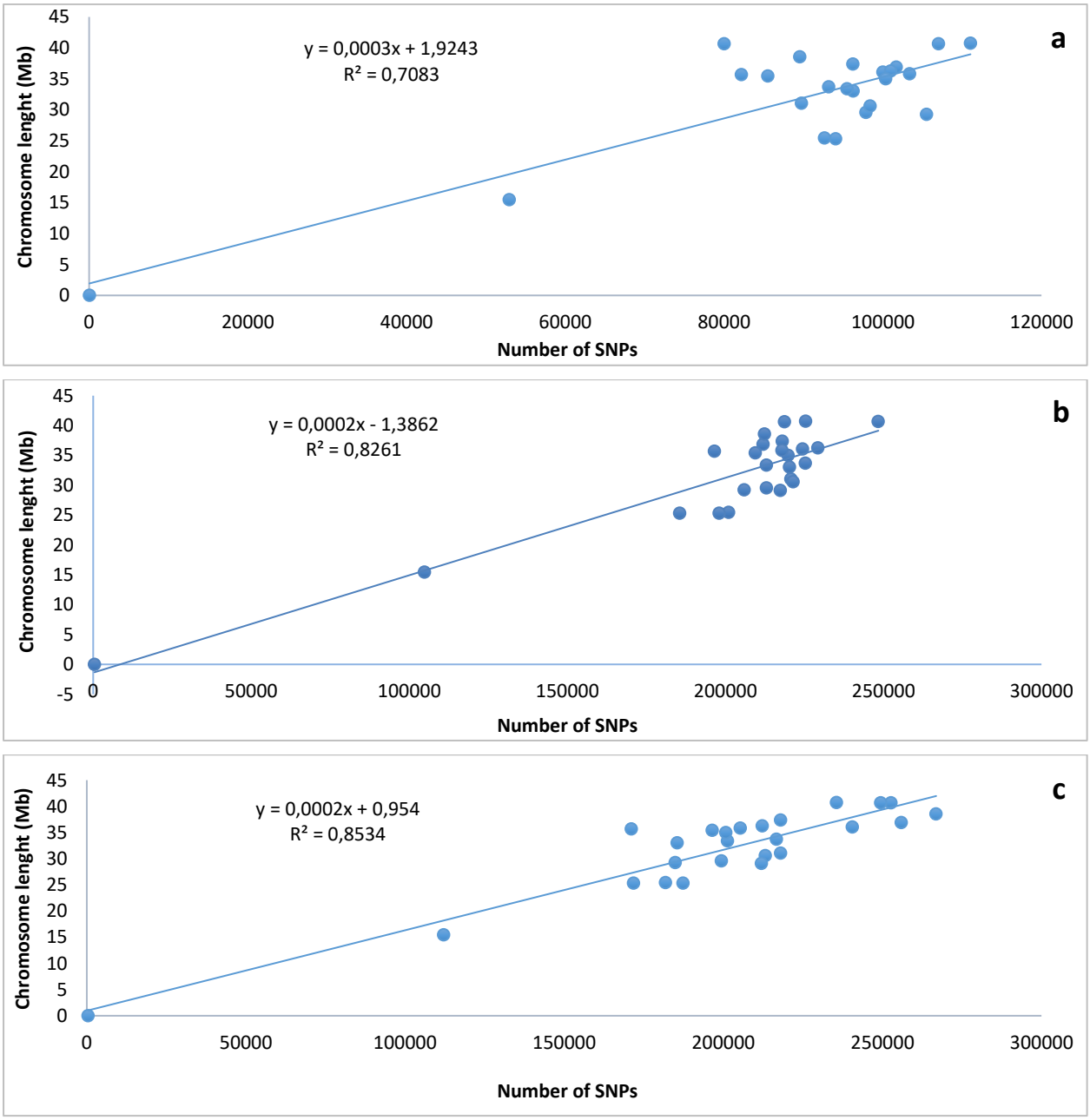

Figure 6. Relationship between the number of SNPs and chromosome length. a) *Trachurus trachurus*, b) *Trachurus picturatus*, c)*Trachurus mediterraneus*

species due to its ability to provide long read lengths and rapid genome sequencing without the need for additional DNA amplification or enzymatic incorporation of modified nucleotides (Steinbock and Radenovic, 2015; Bian et al*.,* 2020). Studies have also demonstrated the potential of nanopore sequencing for species identification, especially when focused on species-level taxonomic resolution, the search for rare taxa or accurate estimation of their genetic diversity (Angell et al., 2020; Szoboszlay et al., 2023). For instance, genomic information for the conservation of the endangered silky shark (*Carcharhinus falciformis*) was rapidly obtained using Nanopore sequencing and appropriate strategies were developed (Baeza and León, 2022). Similarly, researchers have leveraged Nanopore sequencing and Hi-C technology to assemble a chromosome-level genome for *Thamnaconus septentrionalis*, a species experiencing severe population declines. This study produced 50.95 Gb of Nanopore sequencing data, ultimately yielding a high-quality 474.31 Mb de novo reference genome (Bian et al., 2020). Additionally, for brown marbled grouper (*Epinephelus fuscoguttatus*), an important fish in the coral reef ecosystem and marine aquaculture industry, a total of 91.7 Gb and 107.86 Gb of raw data were obtained using Oxford Nanopore technology and Illumina Novaseq technology, respectively. In this study, a high-quality chromosome-level genome with a total length of 1047 Mb was assembled and a de novo reference genome was obtained (Yang et al*.,* 2021). Moreover, Ge et al*.* (2019), assembled the contiguous genome of red-spotted grouper (*Epinephelus akaara*) by de novo assembly of long sequence reads generated by Nanopore sequencing technology (GridION) and Hi-C. They obtained 54.7 Gb of clean reads with Illumina technology and 111.05 Gb of data using Oxford Nanopore sequencing with nineteen Flow Cells. After filtering, a total of 106.29 Gb of data was obtained with read length and N50 length of 18.35 kb and 26 kb, respectively. Next-generation sequencing studies were also conducted to support conservation and management research on endangered species. For this purpose, in Murray cod, a total of 70.6 Gb of nucleotide sequence was obtained with the Illumina sequence and 804 Mb (N50: 4.438 bp, longest read: 129.945 bp) with Oxford Nanopore MinION technology (Austin, 2017). Genner et al. (2022), assembled the Denovo genome of Atlantic horse mackerel (Genome assembly fTraTra1.2, Genbank ID: GCA_905171665.1) using three different sequencing platforms. 98.68% of the genome sequence was linked to 24 chromosomal pseudomolecules, resulting in a total genome of 801 megabases in length. In our study, 26.91 Gb of data total, of which 17.61 Gb was collected after filtering in Oxford Nanopore PromethION for three *Trachurus* species and a total of 8.2 Gb raw data the in Illumina NovaSeq X Plus for *Trachurus mediterraneus*. After raw reads, controlled quality and reads were aligned to the *Trachurus trachurus* reference genome. We assembled genome of total sequence length 1051 Mb with maximum sequence length 230 kb for *Trachurus trachurus*, total sequence length 838 Mb with maximum sequence length 107 kb for *Trachurus mediterraneus* and total sequence length 1080 Mb with maximum sequence length 205 kb for *Trachurus picturatus*. For *Trachurus mediterraneus*, sequencing with the Illumina NovaSeq X Plus platform was employed to address limitations encountered with PromethION, including pore death in the flow cell and suboptimal data yield. PromethION, while advantageous for its long-read sequencing capability, is susceptible to reduced sequencing throughput due to flow cell degradation during extended runs, which can compromise the quantity and quality of the generated data. To overcome this limitation, the integration of data from Illumina's NovaSeq X Plus was implemented. Illumina's platform is well established for its high accuracy and reliability in short-read sequencing, making it an ideal complement to PromethION's long-read approach. By combining these technologies, we achieved enhanced coverage and resolution in the genetic analysis of *T. mediterraneus.*

Next-generation sequencing technologies enable the discovery of single nucleotide polymorphisms (SNPs) as well as variations at the gene level for genetic diversity and phylogenetic analyses (Wenne, 2023). Single nucleotide polymorphic (SNP) markers are used in many fields such as population studies, ecological and conservation genetics; because they are distributed throughout, the genome and have a high variation rate (Helyar et al., 2012, Vera et al., 2013, Mastrochirico-Filho et al*.,* 2016, Saint-Pe et al*.,* 2019). With recent advances in sequencing technology, a large number of single nucleotide polymorphisms (SNPs) can identify local adaptations and relationships between environments through genotype-environment association analyses. (Shi et al., 2023) In a previous study, Mollah et al*.* (2019), isolated a total of 792939 SNPs by constructing a reference genome to identify genes associated with ecological adaptation and migratory behavior for the economically important *T. ilisha* belonging to the family Clupeidae. In addition, Louro et al. (2018), identified a total of 2.3 million filtered heterozygous SNPs in European sardine using the Illumina HiSeq X platform for conservation and fisheries management. Moreover, Yanez et al. (2020) identified to investigate phenotype–genotype associations 29.9 million non-redundant/filterless SNPs in 326 nil tilapia (Oreochromis niloticus) individuals with the illumina Hiseq. In our study, for three species 12.4 milion SNPs were identified in unplaced scaffolds and 24 chromosomes. We also identified isolated SNPs for *Trachurus trachurus, Trachurus picturatus, Trachurus mediterraneus* with transversion:transition ratio of 1:1.3, 1:1.3, 1:1.4, respectively. In our study, SNPs with a minimum variant frequency of 0.70 were considered. This threshold was set because, although PromethION has the capability for rapid detection of genetic

variations and excels in long-read sequencing, it also exhibits a relatively high error rate (~5-10%), which can lead to an increased number of false positives in SNP identification. On the other hand, the Illumina platform, with its high accuracy in short-read sequencing (~99.9%), is widely regarded as a standard for SNP detection. For *Trachurus mediterraneus*, we observed that combining PromethION's long reads with Illumina's short reads resulted in increased variant frequencies. This finding highlights the advantage of using a multi-platform approach, as it reduces error rates and enhances the reliability of SNP identification.

Recent advancements in whole-genome sequencing have revolutionized SNP discovery by enabling the use of reference genomes, which simplify genome assembly and improve the accuracy of identified SNP markers. However, a single reference genome is now recognized as inadequate to represent the full spectrum of genetic diversity within a species. The construction of a pangenome—a comprehensive genomic resource that integrates the collective genetic variation of multiple individuals—has become crucial for understanding the extent and nature of genomic diversity within and across populations (Hurgobin and Edwards, 2017). In this study, we constructed a pangenome for *Trachurus trachurus* to capture the genomic diversity within this species, while reference genomes were generated for other species in the *Trachurus* genus. The *T. trachurus* pangenome provides a framework for identifying population-specific or rare SNPs that may not be represented in a single reference genome, thereby reducing biases in downstream analyses. This dataset represents a significant contribution to genomic studies of *Trachurus* species, offering a valuable resource for future investigations into genetic diversity, population structure, and adaptive evolution. By integrating a pangenomic approach with species-specific reference genomes, we address the limitations of traditional single-genome methods and pave the way for more accurate and representative genomic studies.

Conclusion

In this study, we present the first comprehensive genomic assemblies for Trachurus mediterraneus and Trachurus picturatus from the Turkish seas using the reference genome of Trachurus trachurus. By identifying single nucleotide polymorphisms (SNPs) across these genomes, we have generated a robust set of genetic markers that are highly informative for discriminating between these species. The SNP markers identified in this study will play a crucial role in defining genetic stock structures, which is essential for sustainable fisheries management. The ability to monitor population diversity and resolve mixed stock compositions using these markers will greatly enhance efforts to prevent overfishing, support biodiversity and ensure the long-term sustainability of these commercially important species. Furthermore, these markers can be applied to genetic monitoring programs that track population trends and assess the effects of anthropogenic pressures on genetic diversity. From a phylogenetic perspective, the SNPs we identified will provide valuable information on the evolutionary history and adaptive strategies of *Trachurus* species and contribute to a broader consideration of ecological and evolutionary patterns of the Carangidae family. In terms of conservation and management, this study provides essential genomic data that are critical for developing strategies to conserve and sustain horse mackerel populations, which are facing increasing threats due to fishing pressures and global market demand. Accurately determining genetic stock structures using the SNP sets we generated in this study will inform evidence-based decision-making for monitoring population dynamics, fisheries regulations, habitat conservation and population restoration efforts, and ultimately contribute to the resilience and sustainability of these species in the face of global environmental challenges.

Ethical Statement

All experiments in the present study were approved by, Animal Ethics Committee of Central Fisheries Research Institute (HADYEK/325.04.02-23), Türkiye.

Funding Information

This study was supported by General Directorate of Agricultural Research and Policies of Türkiye within "NATIONAL FISHERY GENETICS DATABESE" project which its number is (TAGEM/HAYSÜD/Ü/23/A6/P3/6022).

Author Contribution

Z.D.D designed the study, collected of the sample, realized to the analysis, and drafted the manuscript.

Y.T. participated in the design of the study, contributed to the bioinformatics analysis and writing.

S.G. participated in the design of the study and writing

Y.B. participated in the design of the study and writing.

All authors have reviewed and approved the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial or non-financial, professional, or personal conflicts that could have appeared to influence the work reported in this paper.

Acknowledgements

This study consists of a part of the doctoral thesis titled " Development of SNP Markers with Genome Sequencing Method for Trachurus Species (*Trachurus mediterraneus, Trachurus trachurus, Trachurus picturatus*, Perciformes: Carangidae) Distributed in Turkish Seas". We thank Dr. Merve Karakuş and Eyüp Çakmak for their help in collecting the samples.

References

Angell, I.L., Nilsen, M., Carlsen, K.C.L., Carlsen, K.H., Hedlin, G., Jonassen, C. M., & Rudi, K. (2020). De novo species identification using 16S rRNA gene nanopore sequencing. PeerJ, 8, e10029.

Attard, C. R., Sandoval-Castillo, J., Gilligan, D. M., Unmack, P. J., Faulks, L. K., & Beheregaray, L. B. (2022). Genomics outperforms genetics to manage mistakes in fisheries stocking of threatened species. Biodiversity and Conservation, 31(3), 895-908.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., & Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature, 465(7298), 627-631.

Austin, C.M., Tan, M.H., Harrisson, K.A., Lee, Y.P., Croft, L.J., Sunnucks, P., Gan, H. M. (2017). De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. GigaScience, 6(8), gix063. https://doi.org/10.1093/gigas cienc e/gix063.

Aydın, M., & Karadurmuş, U. (2012). Age, growth, length-weight relationship and reproduction of the Atlantic horse mackerel (*Trachurus trachurus* Linnaeus, 1758) in Ordu (Black Sea). Ordu Üniversitesi Bilim ve Teknoloji Dergisi, 2(2), 68-77.

Baeza, J.A., & García-De León, F.J. (2022). Are we there yet? Benchmarking low-coverage nanopore long-read sequencing for the assembling of mitochondrial genomes using the vulnerable silky shark *Carcharhinus falciformis*. BMC genomics, 23(1), 320.

Bat, L., Sezgin, M., Şahin, F., Özdemir, Z.B., & Ürkmez, D. (2013). "Sinop City Fishery of the Black Sea" Mar. Sci. 3: 55-64.

Beacham, T. D., Wallace, C., Jonsen, K., McIntosh, B., Candy, J. R., Rondeau, E. B., & Withler, R. E. (2020). Accurate estimation of conservation unit contribution to coho salmon mixed-stock fisheries in British Columbia, Canada, using direct DNA sequencing for single nucleotide polymorphisms. Canadian Journal of Fisheries and Aquatic Sciences, 77(8), 1302-1315.

Bektas Y.,& Belduz A.O. (2008). "Molecular phylogeny of Turkish *Trachurus* species (*Perciformes: Carangidae*) inferred from mitochondrial DNA analyses" J. Fish Biol. 73:1228-1248.

Bian, L., Li, F. H., Ge, J. L., Wang, P. F., Chang, Q., Zhang, S. N., ... & Liu, X. T. (2020). Chromosome-level genome assembly of the greenfin horse-faced filefish (*Thamnaconus septentrionalis*) using Oxford Nanopore PromethION sequencing and Hi-C technology. 20. https://doi. org/10.1111/1755-0998.13183, 1069-1079.

Bilecenoğlu M., Kaya M., Cihangir B., & Çiçek E. (2014). "An updated checklist of the marine fishes of Turkey", Turk. J. Zool. 38: 901-929.

Bostanci, D. (2009). Sarıkuyruk İstavrit, *Trachurus mediterraneus* (Steindachner, 1868)'un Otolit Özellikleri ve Bazı Populasyon Parametreleri. Firat University Journal of Science, 21(1).

Bowdich, S. L. (1825). Fishes of Madeira. Pp. 121-125 and 233-238.

Di Natale, A., Bariche, M., Bizsel, C., & Massuti, E. (2011). *Trachurus trachurus* (Mediterranean assessment). The IUCN Red List of Threatened Species 2011: e.T198647A9066421. Accessed on 16 October 2024.

Durmuş, M. (2019). Evaluation of nutritional and mineral-heavy metal contents of horse mackerel (*Trachurus trachurus*) in the Middle Black Sea in terms of human health. Biological Trace Element Research, 190, 208-216.

Elliott, L., & Russello, M. A. (2018). SNP panels for differentiating advanced-generation hybrid classes in recently diverged stocks: A sensitivity analysis to inform monitoring of sockeye salmon re-stocking programs. Fisheries Research, 208, 339-345.

Elshire, R., Glaubitz, J., Sun, Q., Poland, J., Kawamoto, K., Buckler, E., Mitchell, S., Robust, A. (2011). Simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5):e19379

FAO. (2022). Meeting summary of the FAO/CECAF Working Group on the Assessment of Small Pelagic Fish off Northwest Africa. Nouakchott, Mauritania, 5-9 December 2022. CECAF/SSCIX/2022/4a.

Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M., ... & Luikart, G. (2016). Genomics in conservation: case studies and bridging the gap between data and application. Trends in ecology & evolution, 31(2), 81-83.

Ge, H., Lin, K., Shen, M., Wu, S., Wang, Y., & Zhang, Z. … Zheng, L. (2019). De novo assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. Molecular Ecology Resources, 19(6), 1461–1469. https://doi. org/10.1111/1755-0998.13064

General Fisheries Commission for the Mediterranean. Studies and Reviews; Rome Iss. 99, (2019). 1-3, 5-17, 19-83, 85-151,153-259, I, IV, IX.

Genner, M., Collins, R., of Life, W. S. I. T., & Darwin Tree of Life Consortium. (2022). The genome sequence of the Atlantic horse mackerel, *Trachurus trachurus* (Linnaeus 1758). Wellcome Open Research, 7.

Glaubitz, J., Casstevens, T., Lu, F., Harriman, J., Elshire, R., Sun, Q., & Buckler, E. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9(2):e90346

Hauser, L., & Seeb, J.E. (2008). Advances in molecular technology and their impact on fisheries genetics. Fish. Fish. 9(4): 473–486. https://doi. org/10.1111/j.1467-2979.2008.00306.x.

Helyar, S. J., Limborg, M. T., Bekkevold, D., Babbucci, M., Van Houdt, J., Maes, G. E., ... & Panitz, F. (2012). SNP discovery using next generation transcriptomic sequencing in Atlantic herring (Clupea harengus). PLoS ONE 2012, 7, e42089. https://doi.org/10.1371/journal.pone.0042089

Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Penagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K., et a. (2014). Insights into the maize pan-genome and pan-transcriptome. Plant Cell, 26, 121–135.

https://www.ou.edu/mideast/country/turkey.htm.

Hurgobin, B., & Edwards, D. (2017). SNP discovery using a pangenome: has the single reference approach become obsolete? Biology, 6(1), 21.

Jansen, H.J., Liem, M., Jong-Raadsen, S.A., Dufour, S., Weltzien, F.A., Swinkels, W., & Henkel, C.V. (2017). Rapid de novo assembly of the European eel genome from nanopore sequencing reads. Scientific reports, 7(1), 7213.

Jumbo-Lucioni, P., Ayroles, J. F., Chambers, M. M., Jordan, K. W., Leips, J., Mackay, T. F., & De Luca, M. (2010). Systems genetics analysis of body weight and energy metabolism traits in Drosophila melanogaster. BMC genomics, 11(1), 297.

Kalaycı, F., Samsun, S., Şahin, C., Samsun, O., & Gözler, A.M. (2010). Comparison of biological characteristics of the horse Mackerel (*Trachurus trachurus* L. 1758) which caught of different fishing gears in the Southern Black Sea (Turkey).

Kumar, S., Banks, T.W., & Cloutier, S. (2016). SNP discovery through next-generation sequencing and its applications. Crop Breeding, 209-244.

Li, Y. H., Zhou, G., Ma, J., Jiang, W., Jin, L. G., Zhang, Z., ... & Qiu, L. J. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature biotechnology*, 32(10), 1045-1052.

Linnaeus, C. (1789). Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species; cum characteribus, differentiis, synonymis, locis (Vol. 1). apud JB Delamolliere.

Louro, B., De Moro, G., Garcia, C., Cox, C. J., Veríssimo, A., Sabatino, S. J., ... & Canario, A. V. (2019). A haplotype-resolved draft genome of the European sardine (*Sardina pilchardus)*. GigaScience, 8(5), giz059.

Mastrochirico-Filho, V.A., Hata, M.E., Sato, L.S., Jorge, P.H., Foresti, F., Rodriguez, M.V., Martínez, P., Porto-Foresti, F., & Hashimoto, D.T. (2016). SNP discovery from liver transcriptome in the fish *Piaractus mesopotamicus*. Conserv. Genet. Resour. 8, 109–114.

Melnikova, E., & Kuzminova, N. (2020). Growth of horse mackerel populations in the western parts of the Black, Marmara, and Aegean seas. Fisheries & Aquatic Life, 28(1), 11-17.

Mollah, M.B.R., Khan, M.G.Q., Islam, M.S., & Alam, M.S. (2019). First draft genome assembly and identification of SNPs from hilsa shad (*Tenualosa ilisha*) of the Bay of Bengal. F1000Research, 8.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., & Hohenlohe, P.A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. 22(11): 2841–2847.
https://doi. org/10.1111/mec.12350. PMID: 23711105

Saint-Pe, K., Leitwein, M., Tissot, L., Poulet, N., Guinand, B., Berrebi, P., Marselli, G., Lascaux, J.M., & Gagnaire, P.A. (2019). Blanchet, S. Development of a large SNPs resource and a low-density SNP array for brown trout (*Salmo trutta*) population genetics. BMC Genom. 20, 582.

Santos, B.S., Romana-Eguia, M.R.R., Basiao, Z.U., & Ikeda, M., (2015). Development andcharacterization of nine novel microsatellite markers for the milkfish *Chanoschanos*. Conserv. Genet. Resour. 7, 451–453.

Schlötterer, C. (2004). The evolution of molecular markers—just a matter of fashion? Nat. Rev. Genet. 5(1): 63–69. doi:10.1038/nrg1249. PMID: 14666112

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., & Seeb, L.W. (2011). Singlenucleotide polymorphism (SNP) discovery and applications of SNP genotypingin nonmodel organisms. Mol. Ecol. Resour. 11, 1–8

Shi, Y., Bouska, K. L., McKinney, G. J., Dokai, W., Bartels, A., McPhee, M. V., & Larson, W. A. (2023). Gene flow influences the genomic architecture of local adaptation in six riverine fish species. Molecular Ecology, 32(7), 1549-1566.

Steinbock, L.J., & Radenovic, A. (2015). The emergence of nanopores in next-generation sequencing. Nanotechnology, 26(7), 074003.

Steindachner, F. (1868). Ichthyologischer Bericht über eine nach Spanien und Portugal unternommene Reise. Kaiserl. Akademie der Wissenschaften.

Szoboszlay, M., Schramm, L., Pinzauti, D., Scerri, J., Sandionigi, A., & Biazzo, M. (2023). Nanopore is preferable over Illumina for 16S amplicon sequencing of the gut microbiota when species-level taxonomic classification, accurate estimation of richness, or focus on rare taxa is required. Microorganisms, 11(3), 804.

Tan, M.H., Austin, C.M., Hammer, M.P., Lee, Y.P., Croft, L.J., & Gan, H.M. (2018). Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. Gigascience, 7(3), gix137.

Turan, B.C., Gurlek, M., Yaglioglu, D., & Ozturk, B. (2009). Genetic differentiation of Mediterranean horse mackerel (*Trachurus mediterraneus*) populations as revealed by mtDNA PCR-RFLP analysis. Journal of Applied Ichthyology, 25(2), 142-147.

Vera, M., Alvarez-Dios, J.A., Fernandez, C., Bouza, C., Vilas, R., & Martinez, P. (2013). Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. *International Journal of Molecular Sciences*, *14*(3), 5694-5711.

Villanova, G.V., Vera, M., Díaz, J., Martinez, P., Calcaterra, N.B., & Arranz, S.E. (2015). Isolation and characterization of 20 polymorphic microsatellite loci in themigratory freshwater fish *Leporinus obtusidens* (*Characiformes: Anostomidae*) using 454 shotgun pyrosequencing. J. Fish Biol. 86, 1209–1217

Ward, R. D. (2000). Genetics in fisheries management. Hydrobiologia, 420(1), 191-201.

Wenne, R. (2023). Single nucleotide polymorphism markers with applications in conservation and exploitation of aquatic natural populations. Animals, 13(6), 1089.

Xiao, S., Han, Z., Wang, P., Han, F., Liu, Y., Li, J., & Wang, Z.Y. (2015). Functional markerdetection and analysis on a comprehensive transcriptome of large yellowcroaker by next generation sequencing. PLoS One 10 (4), e0124432.

Yáñez López, J., Yoshida, G., Barría, A., Palma Véjares, R., Travisany, D., Díaz Madrid, D., & Maass Sepúlveda, A. (2020). High-throughput single nucleotide polymorphism (SNP) discovery and validation through whole-genome resequencing in Nile Tilapia (*Oreochromis niloticus*).

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. Nature genetics, 42(7), 565-569.

Yang, Y., Wu, L., Wu, X. I., Li, B., Huang, W., Weng, Z., Lin, Z., Song, L., Guo, Y., Meng, Z., Liu, X., & Xia, J. (2020). Identification of candidate growth-related SNPs and genes using GWAS in brown-marbled grouper

(*Epinephelus fuscoguttatus*). Marine Biotechnology, 22(2), 153–166. https://doi.org/10.1007/s1012 6-019- 09940-8

Yankova M., Raykov V., Gerdzhikov D., Bogomilova P. (2010). Growth and length-weight relationships of the horse mackerel, *Trachurus mediterraneus* ponticus (Aleev, 1956), in the Bulgarian Black Sea Coast – Turk. J. Zool. 34: 85-92.

Yu, A., Shi, Y., & Yan, Y. (2020). Development and characterization of 50 SNP markers in *Coilia ectenes*. Conservation genetics resources, 12(2), 177-181.

Zhao, S., Yang, X., Pang, B., Zhang, L., Wang, Q., He, S., & Zhang, H. (2022). A chromosome-level genome assembly of the redfin culter (*Chanodichthys Erythropterus*). Scientific Data, 9(1), 535.

Zhao, X., Huang, Y., Bian, C., You, X., Zhang, X., Chen, J., & Shi, Q. (2022). Whole genome sequencing of the fast-swimming Southern bluefin tuna (*Thunnus maccoyii*). Frontiers in Genetics, 13, 1020017.