scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Highly contiguous chromosome-level assembly of the rock goby (*Gobius paganellus*) genome

Paolo Franchini[1], Giulia Gentile[1], Martin Pippel[2], Amando Macali[1] & Daniele Canestrelli[1] ✉

The gobies represent the most species-rich family of marine fishes and occupy a remarkable range of ecological niches. Despite their abundance and key role in coastal ecosystems, Mediterranean gobies remain underrepresented in genomic resources. The rock goby (*Gobius paganellus*), a common inhabitant of rocky intertidal and subtidal habitats throughout the Northeast Atlantic and Mediterranean Sea, exemplifies ecological and physiological resilience to highly variable environments. Here, we present a chromosome-level genome assembly for *G. paganellus*, generated using a combination of PacBio HiFi and Hi-C sequencing. The final assembly spans 813 Mb, with more than 99.9% of the sequence anchored to 23 pseudochromosomes (Scaffold N50: 36.4 Mb; Contig N50: 20.3 Mb), and shows high completeness, as indicated by a 98.8% BUSCO score. We characterised the genome's repeat landscape and identified 23,493 protein-coding genes, over 96% of which showed homology to known proteins. This high-quality genomic resource provides a foundation for future population genomics-based research into this species and holds the potential to foster comparative genomic analysis across gobies and teleosts more broadly.

Background & Summary

The Gobiidae, also known as gobies, constitute the most species-rich family of marine fishes, with over 2,000 described species distributed across marine, estuarine, and freshwater ecosystems worldwide[1]. These fish exhibit extraordinary ecological diversity and adaptability, with members occupying habitats ranging from tropical coral reefs to cold temperate estuaries, and from freshwater rivers to terrestrial burrows. Gobies are also among the most abundant fish in many coastal environments, often dominating benthic communities in terms of both biomass and species richness[2].

Their high reproductive capacity and fine-scale habitat specialisation make gobies valuable model systems for studying life-history evolution, dispersal, and adaptation to dynamic or extreme environments[3]. Moreover, their repeated transitions among marine, brackish, freshwater, and even semi-terrestrial habitats provide unique opportunities to explore the genomic underpinnings of ecological diversification and physiological plasticity[4,5]. In addition to their ecological versatility, gobies exhibit remarkable genome plasticity, with extensive karyotype variation both across species and within *Gobius paganellus* (most common diploid numbers from 2n = 44–48)[6–8], suggesting rapid chromosomal evolution and structural genomic diversity.

The rock goby (*Gobius paganellus*) is a small benthic marine fish widely distributed along the northeastern Atlantic coasts of Europe and North Africa, as well as throughout the Mediterranean Sea (Fig. 1). A characteristic inhabitant of rocky intertidal and shallow subtidal ecosystems, *G. paganellus* plays a key ecological role in nearshore food webs and exhibits behavioural and physiological adaptations to highly dynamic environments, including tidal fluctuations, temperature variability, and hypoxia[9–11]. A genome assembly of *Gobius paganellus* (NCBI accession: GCA_965278345.1)[12] is currently available in GenBank, derived from a specimen collected in the Atlantic Ocean, with a karyotype of n = 24 chromosomes. However, given the species' broad distribution and pronounced phenotypic plasticity, focusing on multiple populations is essential to better capture its genomic variability. Providing a high-quality reference genome for *G. paganellus* from the Mediterranean Sea fills an important gap in marine genomics and strengthens the genomic framework of the family Gobiidae. Moreover,

[1]Department of Ecological and Biological Sciences, University of Tuscia, Viale dell'Università s.n.c., Viterbo, Italy. [2]Science for Life Laboratory, National Bioinformatics Infrastructure Sweden (NBIS), Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden. ✉e-mail: canestrelli@unitus.it

Fig. 1 *Gobius paganellus* specimen from the Tyrrhenian Sea (photo by Armando Macali).

it provides a valuable resource for investigating the genetic basis of local adaptations across contrasting environments and for assessing the potential impacts of climate change on this ecologically versatile species.

The Mediterranean Sea is a hotspot of marine biodiversity and endemism, shaped by a complex geological history and pronounced environmental gradients[13,14]. Species inhabiting this semi-enclosed sea are frequently subjected to strong local selective pressures, such as seasonal shifts in salinity, temperature extremes, and anthropogenic pollution[15]. These conditions make Mediterranean fish ideal candidates for exploring the genomic basis of local adaptation, particularly in the context of climate change and environmental stress. *G. paganellus*, with its widespread presence in intertidal habitats and tolerance to variable conditions, might serve as a powerful model for such studies.

In this study, we present a chromosome-level reference genome of *Gobius paganellus*, assembled using a combination of high-fidelity (HiFi) PacBio long reads and chromosome conformation capture (Hi-C) derived short reads. We further annotated repetitive elements and protein-coding genes using a newly generated multi-tissue RNA-Seq dataset alongside extensive protein reference databases. This high-quality genome provides a valuable resource for future studies on the ecological and evolutionary dynamics of *G. paganellus*. Additionally, it will support comparative genomic analyses across the Gobiidae family and more broadly among fishes.

Methods

Sample collection and data generation. The reference genome of *Gobius paganellus* was assembled from a single individual collected along the coast of Monte Argentario, Italy (42.43363° N, 11.15819° E) in 2024. The specimen measured 11.5 cm in total length and weighed 9 g. Sex could not be determined as the gonads were not macroscopically distinguishable. Following capture by line fishing, the fish was euthanized by immersion in an overdose of tricaine methanesulfonate (MS-222) to ensure rapid loss of consciousness and death and immediately dissected. Liver tissue was snap-frozen in liquid nitrogen, transported to the laboratory for long-term storage at −80 °C, and subsequently shipped on dry ice to the sequencing provider (BGI Genomics, Hong Kong).

In addition, eight tissues (brain, eyes, fins, gills, gonads, liver, muscle, and skin) were dissected from the same specimen and preserved in RNAlater for transcriptomic analysis. To isolate RNA, a FastPrep-24 homogenizer (MP Biomedicals) was first used to process each sample (30 s at 4.0 M), and then the total RNA was isolated using a Qiagen RNeasy Mini Kit using the manufacturer's recommendations. RNA quantity was assessed using a Qubit Flex fluorometer (Life Technologies, Darmstadt, Germany). RNA samples were shipped on dry ice to BGI, which performed quality assessment (Supplementary Fig. S1), RNA-Seq library preparation and PE150 DNBSEQ-G400 sequencing.

The PacBio Revio sequencing generated 2.12 million (M) HiFi long-reads (mean length: 15,980 bases), totalling 33.9 Gb. Considering the assembled genomes (see below), the depth of coverage of the long-reads was equal to 42x. DNBSeq Hi-C sequencing produced 402.1 M quality-filtered short-reads (60.5 Gb), equalling 74x coverage. Finally, we obtained 269.3 M clean RNA-Seq short-reads from eight tissues (40.4 Gb) (the statistics of the three sequencing types were detailed in Supplementary Table S1).

Genome size estimation. The PacBio long-reads were used to estimate the genome size of *Gobius paganellus* based on k-mer frequency (k-mer value: 21). First, Jellyfish v2.3.1[16] was used to count the occurrence of each k-mer, values that were processed by Genomescope v2.0[17] to estimate genome heterozygosity, repeat content and size using a k-mer-based statistical approach. To further investigate genome characteristics, particularly ploidy and the presence of distinct haplotypes, the k-mer profiles were also analyzed using Smudgeplot v0.4.0[17]. Smudgeplot leverages two-dimensional k-mer coverage histograms to visualize and estimate ploidy, infer chromosome numbers, and detect whole-genome duplications or significant aneuploidies. The haploid genome size was estimated at 688 Mb using GenomeScope2 (Supplementary Fig. S2), while Smudgeplot analysis revealed a clear diploid genome signature (Supplementary Fig. S3).

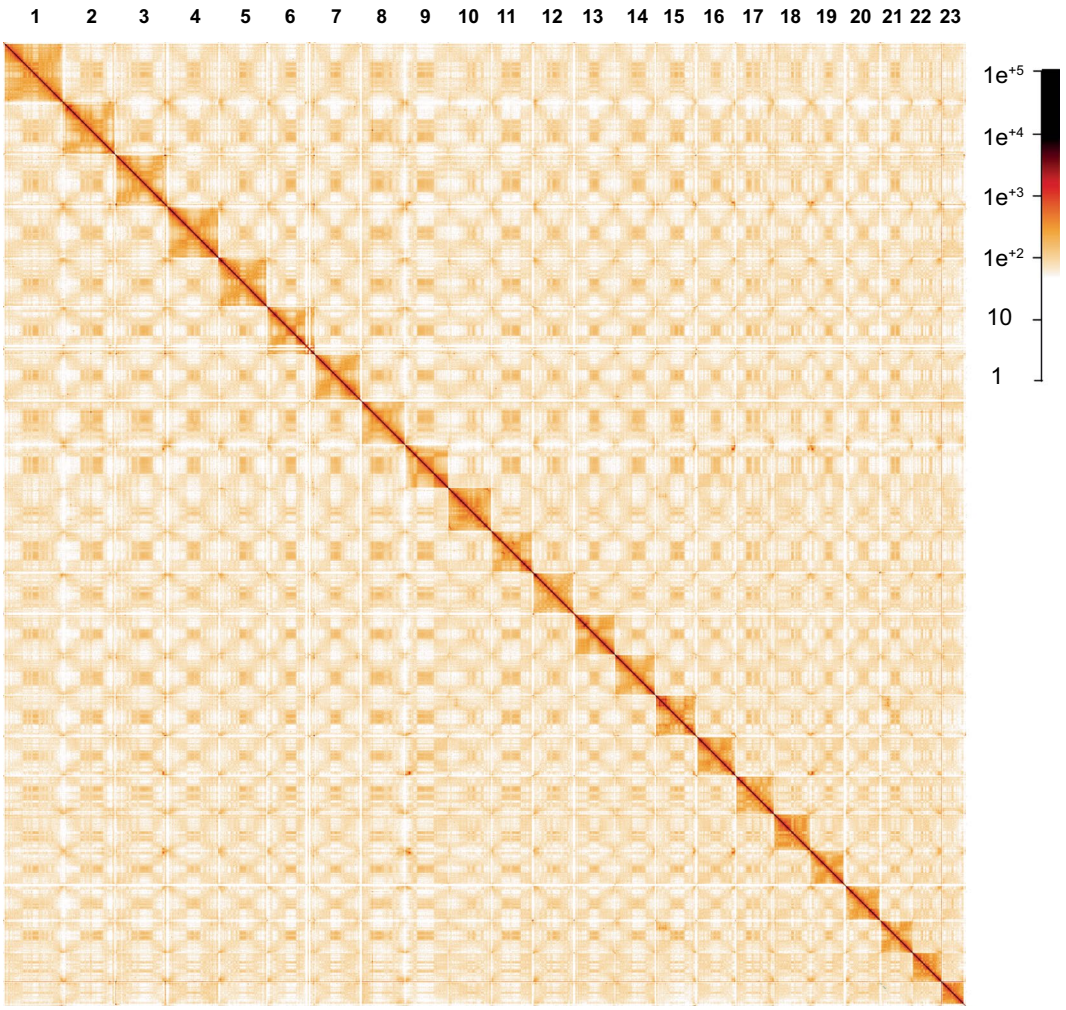

Fig. 2 Hi-C contact map.

Genome assembly. The genome assembly of *Gobius paganellus* was carried out with Hifiasm v0.25.0-r726[18] with the '–primary' option enabled. Remaining haplotypic duplications were removed with purge_dups v1.2.5[19] with default settings. The potential presence of contamination in the assembled contigs was evaluated with Blobtools v4.4.5[20]. HiC reads were aligned to the contigs following Arima mapping pipeline v03, and the resulting BAM file was processed by Yahs v1.2.2[21] to scaffold the contigs with parameters '-v 1–no-contig-ec -q 1–telo-motif AACCCT'. The manual curation of the scaffolds was performed with PretextView v1.0.3 (https://github.com/wtsi-hpag/PretextView) and the GRIT Rapid Curation pipeline (GRIT_Rapid: https://gitlab.com/wtsi-grit/rapid-curation). Two rounds of manual curation were done. The assembly and gene annotation of the mitochondrial genome were carried out using MitoHifi 3.2.1[22]. Genome completeness was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.8.2[23] using the Actinopterygii gene set (odb_10 release). Additionally, the consensus quality value (QV) and k-mer completeness of the *G. paganellus* assembly was evaluated by Merqury v1.3[24]. The assembly pipeline allowed us to obtain a high-quality chromosome-level primary genome of 813 Mb, including 23 pseudochromosome molecules and only four short unplaced contigs, with approximately 99.98% of the assembled sequences anchored to chromosomes (Fig. 2). The massive PacBio and Hi-C coverage was instrumental to assemble a highly contiguous contig-level (86 contigs with N50 of 20.3 Mb) and scaffold/chromosome-level genome (N50: 36.4 Mb), respectively. Blobtools identified no contamination in the assembled contigs (assembly statistics were reported in Supplementary Table S2 and Supplementary Fig. S4).

Repeat elements annotation. Before protein-coding gene prediction, we generated a soft-masked version of the genome assembly, in which repetitive elements were converted to lowercase letters. Repeat annotation was carried out using RepeatModeler v2.0.3 and RepeatMasker v4.1.2-p1, included in the Dfam TEtools Docker/Singularity container v1.3[25]. *De novo* repeat families were identified using RepeatModeler with the 'LTRStruct' option enabled. The resulting *de novo* library was then merged with curated transposable elements (TEs) from the "Actinopterygii" taxonomic group retrieved from the Dfam TE database using famdb.py v2.0.1[25]. RepeatMasker was run in high-sensitivity mode ('-s'), with low-complexity sequence masking disabled ('-nolow'), using this combined repeat library.

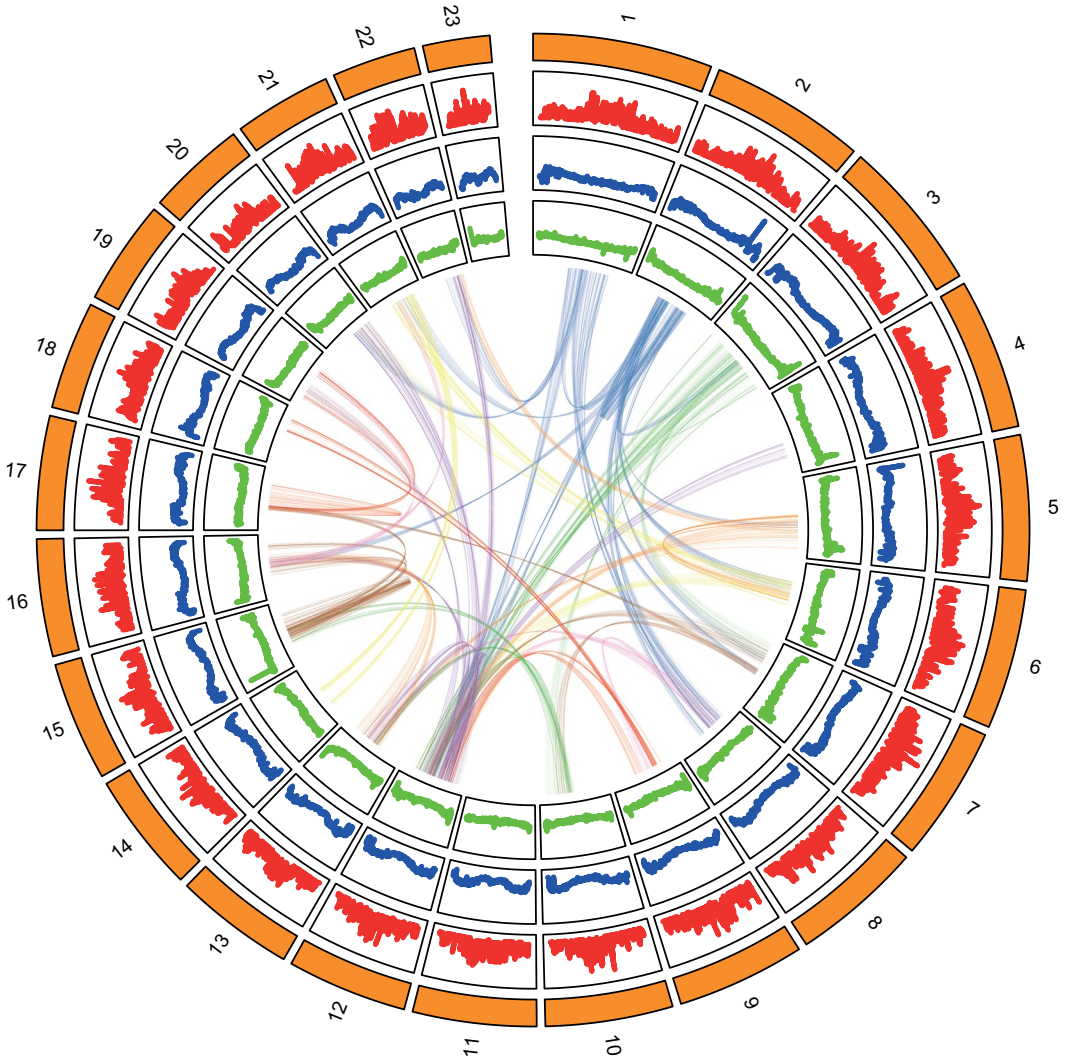

Fig. 3 Circular plot illustrating genomic features of the 23 assembled chromosomes. From outer to inner: orange blocks represent chromosomes scaled by length; gene density; transposable element (TE) density; GC content; synteny plot displaying colinear homologous gene blocks.

A further RepeatMasker run was carried out using the same combined repeat library, but disabling the '-nolow' parameter, and with the '-a' option to produce the alignment file. The repeat content identified using the RepeatModeler/RepeatMasker pipeline amounted to 351.9 Mb (43.3% of the genome). The majority of repeats consisted of DNA transposons (8.00%), followed by LINEs (6.27%), LTRs (3.69%) and SINEs (2.74%). A substantial portion of the repeats remained unclassified (21.30%) (Supplementary Table S3).

Protein-coding gene annotation. To predict protein-coding genes in the newly generated *G. paganellus* genome assembly, we used the BRAKER3 v3.0.8 pipeline[26], which integrates RNA-Seq and protein evidence using GeneMark-ETP[27], AUGUSTUS[28], and TSEBRA[29] for fully automated genome annotation. BRAKER3 was run with three input files: (1) the soft-masked version of the *G. paganellus* genome assembly in fasta format; (2) the whole RNA dataset generated in this study in fastq format (see below); and (3) a comprehensive protein reference dataset compiled by concatenating protein sequences in fasta format from multiple sources. This latter dataset included a total of 10,398,228 proteins comprising 9,805,801 entries from the *Vertebrata* OrthoDB database v12.1[30], 98,394 curated proteins from UniProtKB/Swiss-Prot release 2025_02[31], and 494,033 proteins from 10 fish species from Ensembl release 114[32], selected based on either the close phylogenetic relation with *G. paganellus* or because they are well annotated model species in genomics (*Astyanax mexicanus*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Neogobius melanostomus*, *Oreochromis niloticus*, *Oryzias latipes*, *Periophthalmus magnuspinnatus*, *Takifugu rubripes*, *Xiphophorus maculatus*). RNA-Seq alignments for gene prediction were processed with StringTie2[33], which assembles and quantifies transcripts from RNA-Seq alignments. GFF files were managed using GffRead and GffCompare[34] for conversion and comparison of annotation formats. BEDTools[35] was used for genome feature manipulation during preprocessing. Using the BRAKER3 pipeline with comprehensive RNA-Seq and protein datasets, we predicted 23,493 protein-coding genes (Fig. 3). Following gene prediction, functional annotation of the resulting gene models was performed using the *agat_sp_manage_functional_annotation.pl*

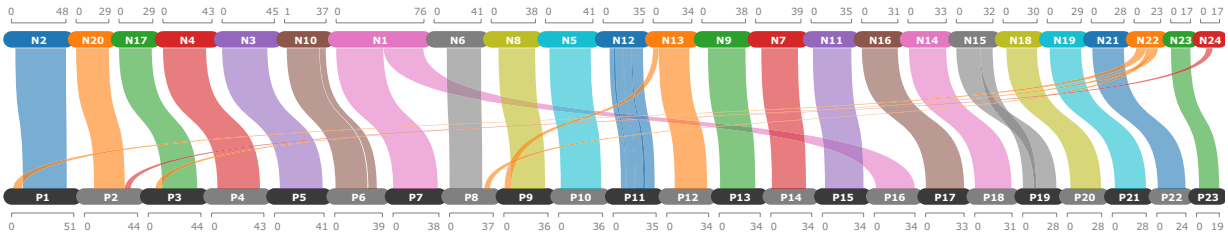

Fig. 4 SynVisio plot illustrating chromosome-scale synteny between *Gobius niger* (top) and *Gobius paganellus* (bottom) genomes. Each colored ribbon represents a colinear block of homologous genes identified from pairwise alignments. Conserved one-to-one correspondences appear as continuous ribbons, while crossing or split ribbons indicate structural rearrangements such as inversions, translocations, fissions or fusions. Chromosomes are labelled and scaled according to their physical length in Megabases (Mb).

script from the AGAT toolkit v1.4.2[36]. Gene names and descriptions were integrated by processing the BLASTp output generated from aligning the newly predicted proteins against the UniProtKB/Swiss-Prot database.

Synteny analysis. To identify potential chromosomal rearrangements and evaluate chromosome conservation within the genus, we conducted a comparative synteny analysis between the newly generated *Gobius paganellus* genome and the *Gobius niger* genome (NCBI accession: GCA_951799975.1)[37] using MCScanX v1.0.0[38]. Pairwise comparisons of the longest protein isoforms for each gene were performed using BLASTp with the parameters '-outfmt 6 -evalue 1e-5'. The resulting alignments were then processed with MCScanX with default settings. Syntenic relationships were visualised using a parallel synteny plot generated with the SynVisio web platform (https://synvisio.github.io/#/) and a dot plot produced with the JCVI synteny visualisation pipeline[39].

Ethic statement. This study was conducted in full compliance with the European Directive 2010/63/EU and the Italian Legislative Decree No. 26/2014 (Art. 1 and 2, paragraphs 1e–f) concerning the protection of animals used for scientific purposes. The specimen, belonging to a non-protected species, was euthanized immediately after capture, and only post-mortem analyses were performed. As no experimental procedures were conducted on live animals, no specific ethical authorization was required under current Italian and European regulations.

Data Records

The raw sequencing data and genome assembly of *Gobius paganellus* have been deposited at the National Center for Biotechnology Information (NCBI) under project number PRJNA1298813. The PacBio HiFi, Hi-C, and RNA-Seq datasets are available in the NCBI Sequence Read Archive (SRA)[40], while the assembled genome is accessible through the NCBI Genome database[12]. Additionally, the protein-coding gene structure annotation has been deposited in the Figshare repository (https://doi.org/10.6084/m9.figshare.29820362)[41]. A detailed list of all raw data analyzed in this study is provided in Supplementary Table S1.

Technical Validation

Three complementary approaches were used to assess the quality of the *Gobius paganellus* genome assembly. First, the BUSCO pipeline detected 3,621 (99.5%) conserved Actinopterygii genes, of which 98.8% were complete and 0.7% were fragmented (Supplementary Table S4). Second, Merqury estimated a consensus quality value (QV) of 67 and a k-mer completeness of 92.16%. Third, alignment of PacBio, Hi-C, and RNA-Seq reads from eight tissues to the assembly yielded average mapping rates of 99.7%, 98.2%, and 92.3%, respectively. Together, these results indicate a high level of genome completeness.

The completeness of the predicted protein-coding genes was further supported by a BUSCO analysis, which detected 96.9% Actinopterygii genes (96.5% complete and 0.4% fragmented) (Supplementary Table S5).

Additionally, a synteny analysis with the congeneric species *G. niger* was performed as an independent assessment of assembly quality. A total of 2,586 collinear blocks comprising 22,419 genes were identified between *G. paganellus* and *G. niger*. As shown in the parallel plot (Fig. 4), most chromosomes displayed a clear one-to-one correspondence, reflecting a largely conserved karyotype organization within the genus *Gobius*. Nevertheless, the analysis revealed a large chromosomal fusion or fission event involving chromosome 1 of *G. niger* and chromosomes 7 and 16 of *G. paganellus*, as well as several minor rearrangements (Fig. 4 and Supplementary Fig. S5).

Data availability

All data supporting this study are publicly available. Raw PacBio HiFi, Hi-C, and RNA-Seq reads are deposited in the NCBI Sequence Read Archive (BioProject PRJNA1298813); the genome assembly is available in GenBank (WGS accession JBPXNH000000000; version JBPXNH010000000 analyzed here). The protein-coding gene annotation is archived on Figshare (https://doi.org/10.6084/m9.figshare.29820362). A complete inventory of files and formats is provided in Supplementary Table S1.

Code availability

No custom scripts or proprietary code were used in this study. All software tools employed are publicly available, and all parameters are explicitly detailed in the Methods section.

Received: 6 August 2025; Accepted: 19 January 2026;
Published online: 29 January 2026

References

1. Nelson, J. S., Grande, T. C. & Wilson, M. V. H. *Fishes of the World*. (John Wiley & Sons, 2016).
2. Thacker, C. E. & Roje, D. M. Phylogeny of Gobiidae and identification of gobiid lineages. *Syst. Biodivers.* **9**, 329–347 (2011).
3. Shang, Y. *et al*. Adaptability and Evolution of Gobiidae: A Genetic Exploration. *Animals* **12**, 1741 (2022).
4. Salvanes, A. G. V., Utne-Palm, A. C., Currie, B. & Braithwaite, V. A. Behavioural and physiological adaptations of the bearded goby, a key fish species of the extreme environment of the northern Benguela upwelling. *Mar. Ecol. Prog. Ser.* **425**, 193–202 (2011).
5. Top, N., Karakuş, U., Tepeköy, E. G., Britton, J. R. & Tarkan, A. S. Plasticity in habitat use of two native Ponto-Caspian gobies, Proterorhinus semilunaris and Neogobius fluviatilis: implications for invasive populations. *Knowl. Manag. Aquat. Ecosyst.* **40** https://doi.org/10.1051/kmae/2019031 (2019).
6. Thode, G., Giles, V. & Alvarez, M. C. Multiple Chromosome Polymorphism in Gobius paganellus (Teleostei, Perciformes). *Heredity* **54**, 3–7 (1985).
7. Amores, A., Giles, V., Thode, G. & Alvarez, M. C. Adaptative character of a Robertsonian fusion in chromosomes of the fish Gobius paganellus (Pisces, Perciformes). *Heredity* **65**, 151–155 (1990).
8. da Silva, S. A. S. *et al*. High chromosomal evolutionary dynamics in sleeper gobies (Eleotridae) and notes on disruptive biological factors in Gobiiformes karyotypes (Osteichthyes, Teleostei). *Mar. Life Sci. Technol.* **3**, 293–302 (2021).
9. Madeira, D. *et al*. Physiological and biochemical thermal stress response of the intertidal rock goby *Gobius paganellus*. *Ecol. Indic.* **46**, 232–239 (2014).
10. Vinagre, C. *et al*. Effect of increasing temperature in the differential activity of oxidative stress biomarkers in various tissues of the Rock goby, *Gobius paganellus*. *Mar. Environ. Res.* **97**, 10–14 (2014).
11. Paul, N., Novais, S. C., Lemos, M. F. L. & Kunzmann, A. Chemical predator signals induce metabolic suppression in rock goby (Gobius paganellus). *PLOS ONE* **13**, e0209286 (2018).
12. Gobius paganellus genome assembly ASM5352551v1. *NCBI Genbank* https://identifiers.org/ncbi/insdc.gca:GCA_053525515.1 (2025).
13. Bianchi, C. N. & Morri, C. Marine Biodiversity of the Mediterranean Sea: Situation, Problems and Prospects for Future Research. *Mar. Pollut. Bull.* **40**, 367–376 (2000).
14. Coll, M. *et al*. The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats. *PLOS ONE* **5**, e11842 (2010).
15. Boulanger, E. *et al*. Climate differently influences the genomic patterns of two sympatric marine fish species. *J. Anim. Ecol.* **91**, 1180–1195 (2022).
16. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
17. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
18. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
19. Guan, D. *et al*. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
20. Challis, R. J., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. L. BlobToolKit – Interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics*, **10**, 1361–1374 (2020).
21. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
22. Uliano-Silva, M. *et al*. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics* **24**, 288 (2023).
23. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
24. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
25. Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob. DNA* **12**, 2 (2021).
26. Gabriel, L. *et al*. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769–777 (2024).
27. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**, 757–768 (2024).
28. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
29. Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M. & Stanke, M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* **22**, 566 (2021).
30. Kriventseva, E. V. *et al*. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).
31. The UniProt Consortium. *et al*. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
32. Yates, A. D. *et al*. Ensembl 2020. *Nucleic Acids Res.* gkz966 https://doi.org/10.1093/nar/gkz966 (2019).
33. Kovaka, S. *et al*. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
34. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **9**, ISCB Comm J-304 (2020).
35. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinforma.* **47**, 11.12.1–34 (2014).
36. Dainat, J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format (Version 1.4.2). *Zenodo* https://doi.org/10.5281/zenodo.3552717.
37. Gobius niger genome assembly fGobNig1.1. *NCBI* https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_951799975.1/.
38. Wang, Y. *et al*. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
39. Bandi, V. & Gutwin C. Interactive Exploration of Genomic Conservation. In Proceedings of the 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020 (GI'20). *Can. Hum.-Comput. Commun. Soc. Waterloo CAN*.
40. *SRP606519 - SRA - NCBI*. https://identifiers.org/ncbi/insdc.sra:SRP606519.
41. Gobius paganellus protein-coding gene annotation. *figshare* https://doi.org/10.6084/m9.figshare.29820362.v1 (2025).

Acknowledgements

This research project was funded by the European Union, NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment 1.4, Project title 'National Biodiversity Future Center (NBFC)' (project code CN_00000033), CUP J83C22000860007.

Author contributions

P.F. conceived the study, performed data analysis, and wrote the first draft of the manuscript. G.G. conducted field and laboratory work and performed data analysis. M.P. performed data analysis. A.M. conducted field work. D.C. conceived the study. All authors contributed to writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-026-06659-9.

Correspondence and requests for materials should be addressed to D.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.