scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Highly contiguous genome assembly and gene annotation of the short-finned eel (*Anguilla bicolor pacifica*)

Hyeongwoo Choi[1]✉, Jiwon Nam[2], Siyoung Yang[3] & Seong-il Eyun[1]✉

In East Asia, anguillid eels are commercially important. However, unlike other species, they have not been successfully cultivated throughout their lifecycle. Facing population decline due to overharvesting and environmental pressures, the industry is turning to alternatives, such as *Anguilla bicolor pacifica* (short-finned eel). However, genomic data for short-finned eels are unavailable. Here, we present in-depth whole-genome sequencing results for short-finned eel obtained using two sequencing platforms (PacBio Revio, and Illumina). In this study, we achieved a highly contiguous genome assembly of the short-finned eel, comprising 19 pseudochromosomes encompassing 99.76% of the 1.087 Gb genome sequence with an N50 of 16.88 and 61.07 Mb from contig and scaffold, respectively. Transcripts from four different tissues led to the annotation of 23,095 protein-coding genes in the eel genome, 98.66% of which were functionally annotated. This high-quality genome assembly, along with the annotation data, provides a foundation for future functional genomic studies of short-finned eels.

Background & Summary

Anguillid eels are commercially important fish in East Asia, with approximately 270,000 metric tons of eels being cultivated worldwide[1–3]. Despite considerable and persistent efforts over many years, achieving a successful full-life cycle culture from egg to adult remains elusive in eels, distinguishing them from numerous other fish species where such comprehensive cultivation has become feasible[4]. Consequently, the eel farming industry relies heavily on the collection of wild glass eels that migrate towards estuaries or inland freshwater habitats[5]. Nonetheless, several factors, including impediments to migration, pollution, climate change, habitat loss, and overexploitation of juvenile glass eels, have significantly reduced eel population[6,7]. Notably, the Japanese eel is categorized as "Endangered," while the European eel holds a "Critically Endangered" classification by the International Union for Conservation of Nature Red List (https://www.iucnredlist.org/search?query=anguilla&searchType=species). Consequently, alternative anguillid eel species, such as *Anguilla marmorata* and *Anguilla bicolor*, have garnered considerable interest[8,9]. Moreover, a recent study has shown that *A. bicolor pacifica* exhibits a faster growth rate than *A. marmorata*, indicating its potential suitability for cage culture[10]. Due to its comparable taste and texture, *A. bicolor pacifica* is recognized as the second-preferred choice, following *A. japonica*, indicating its significant economic importance concerning market demand[11].

Anguilla bicolor (short-finned eel) is globally distributed throughout the Indo-Pacific region, ranging from East Africa to Papua New Guinea, including the Philippines and Indonesia[12]. However, because of its allopatric distribution and slight morphological variations, *A. bicolor* has been divided into two subspecies: *A. bicolor bicolor* (inhabiting the Indian Ocean) and *A. bicolor pacifica* (inhabiting the western Pacific Ocean)[13].

Anguillid eels, known for their catadromous life patterns, adapt to varying environments throughout their life cycle, starting in marine environments as larvae, transitioning to brackish or inland shore waters as juveniles, and settling in fresh waters as adults[14,15]. Their notable resilience to an extensive range of salt concentrations provides opportunities to investigate how they coordinate osmotic pressure during migration. Understanding the osmoregulatory mechanisms will provide valuable insights into the ability of anguillid eels to achieve

[1]Department of Life Science, Chung-Ang University, Seoul, 06974, Korea. [2]Department of Fisheries Science, Chonnam National University, Yeosu, 59626, Korea. [3]Department of Biological Science, Sungkyunkwan University, Suwon, 16419, Korea. ✉e-mail: creo9447@cau.ac.kr; eyun@cau.ac.kr

Platform	Tissue	Insert size (bp)	Raw data (bp)	Clean data (bp)	Average read length (bp)	Coverage (X)	SRA accession
Illumina	Muscle	550	45,294,013,152	40,770,397,562	151	50.33	SRR27869078
PacBio Revio	Muscle	—	83,238,023,138	—	15,968	92.49	SRR27869073
RNA	Muscle	155	5,243,858,188	4,790,527,610	101	5.83	SRR27869074
	Eye	155	6,323,479,912	5,774,190,954	101	7.03	SRR27869075
	Liver	155	5,458,459,150	4,920,790,300	101	6.06	SRR27869076
	Heart	155	6,912,335,970	6,340,204,750	101	7.68	SRR27869077

Table 1. Statistical results for the sequencing data of the *Anguilla bicolor pacifica*.

	Contig assembly	Scaffold assembly	Chromosome-level scaffold assembly
Contig/Scaffold number	405	30	19
N50 (bp)	16,884,858	61,071,980	61,071,980
N90 (bp)	1,598,979	39,552,249	39,552,249
L50	17	8	8
L90	99	16	16
Total length (bp)	1,087,597,977	1,087,610,635	1,085,035,931
Maximum length (bp)	57,332,747	94,400,178	94,400,178
Average length (bp)	2,685,430	36,253,688	57,107,200
GC (%)	43.5	43.5	43.5
BUSCO analysis			
Complete (%)	3,488 (95.8)	3,512 (96.4)	3,510 (96.4)
Single-copy (%)	3,170 (87.1)	3,194 (87.7)	3,192 (87.7)
Duplicated (%)	318 (8.7)	318 (8.7)	318 (8.7)
Fragmented (%)	57 (1.6)	36 (1.0)	36 (1.0)
Missing (%)	95 (2.6)	92 (2.6)	94 (2.6)

Table 2. Statistical results of *Anguilla bicolor pacifica* genome assembly.

homeostasis. Currently, the genomic sequences of six eel species are available. Chromosome-level genome assemblies have been established for three eels: *A. japonica*[16], *A. anguilla*[17], and *A. rostrata* (GCA_018555375.3). The genomes of *A. obscura*, *A. marmorata*, and *A. megastoma* were assembled at the scaffold level[18]. However, genomic information on *A. bicolor pacifica* is lacking.

In summary, the genomic resources presented in this study are valuable for studying the molecular mechanisms that drive evolutionary adaptations in migratory euryhaline fish. Additionally, the chromosome-scale genome of short-finned eels will facilitate comparative genomic studies, which will shed light on the adaptive strategies employed by catadromous fish that enable them to survive and thrive across a diverse range of saline environments.

Methods

Ethics statement. The experimental protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Chonnam National University (CNU IACUC-YS-2023-9).

DNA sample collection, library construction, and DNA sequencing. Short-finned eels (*Anguilla bicolor pacifica*) collected from an aquafarm in Jeonnam, South Korea, were transported to the laboratory and kept in a 250 L aerated tank at a water temperature of 24 °C (Supplementary Table S1). For DNA extraction, the muscle tissue was sampled from a male eel with a standard body length of 20 cm (Supplementary Figure S1). A short-fragment library was generated using the TruSeq DNA Nano 550 bp kit with an insert size of 550 bp. The paired-end library was sequenced using an Illumina NovaSeq 6000 platform. The DNA, ranging from 2 to 5 μg were placed in a single lane of a BluePippin 0.75% gel. Electrophoresis was used to collect libraries of 9–13 kb and $>$ 15 kb. The library was sequenced by circular consensus sequencing (CCS) on a PacBio Revio platform. Genomic DNA was extracted from the muscle tissue of an individual eel. Two platforms were used for the DNA sequencing. In total, 45.3 Gb (50 $\times$ coverage) of Illumina reads and 83.24 Gb (92 $\times$ coverage) of HiFi reads were generated (Table 1).

Genome size estimation, genome assembly, and quality assessment. We removed any Illumina reads shorter than 120 bp that contained adaptor sequences, low quality (Phred score $<$ 20), or unknown bases (*Ns*) using `Trim_galore` (ver. 0.6.7)[19]. The trimmed reads were then used to count the 21-mer reads using `jellyfish` (ver. 2.3.0)[20]. Subsequently, based on the 21-mer histogram, the genome size of *A. bicolor pacifica* was estimated to be 899.9 Mb, with a heterozygosity rate of 1.25% by `GenomeScope` (ver. 2.0)[21] (Supplementary Figure S2).

HiFi reads were used for draft assembly to produce highly contiguous draft contigs using `Hifiasm` (ver. 0.16.1)[22]. This process resulted in the generation of 405 contigs with a total length of 1.09 Gb and contig N50 of 16.9 Mb (Table 2).

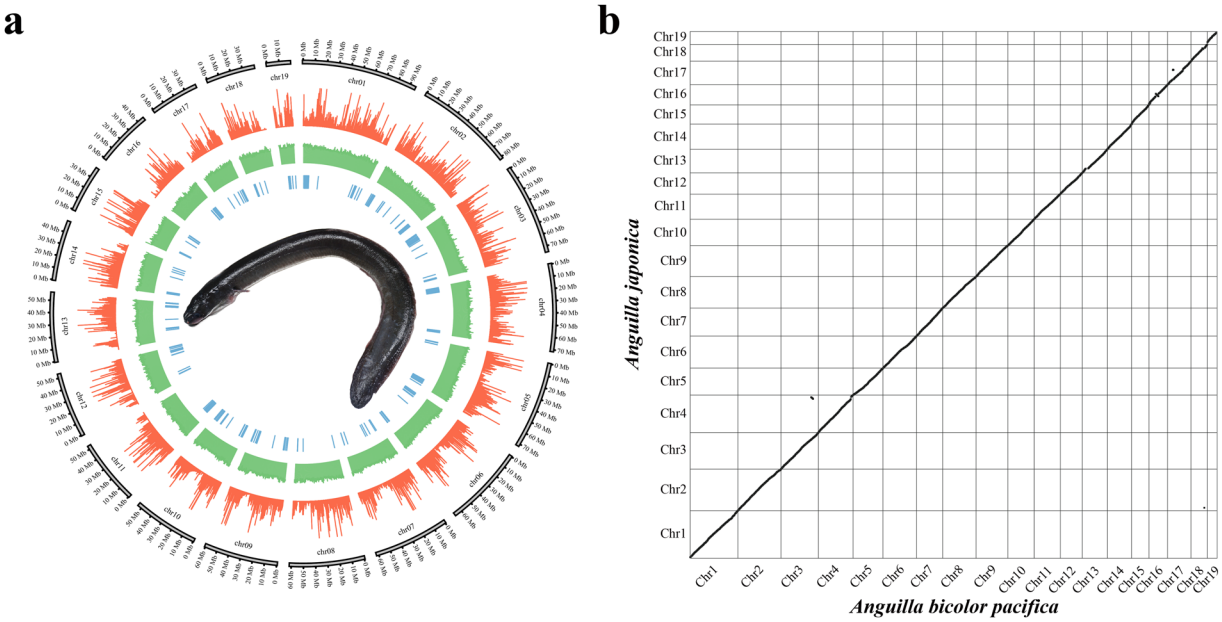

Fig. 1 (**a**) Genomic landscape of the assembled pseudochromosomes of *Anguilla bicolor pacifica*, featuring thick marks every 10 Mb including the gene density (red), and GC content (green) calculated using a 100 Kb sliding window approach. Gaps between contigs were marked with blue lines. (**b**) Syntenic dot plots of the whole genome of *A. bicolor pacifica* against *A. japonica*. The x- and y-axes represent the 19 pseudochromosomes of the *A. bicolor pacifica* and chromosomes of *A. japonica* genomes, respectively.

Chromosome	Length (bp)	% of assembly
Chr1	94,400,178	8.68
Chr2	80,975,899	7.45
Chr3	74,441,218	6.84
Chr4	71,026,713	6.53
Chr5	71,492,640	6.57
Chr6	65,665,778	6.04
Chr7	61,399,891	5.65
Chr8	61,071,980	5.62
Chr9	60,052,471	5.52
Chr10	53,064,840	4.88
Chr11	52,293,486	4.81
Chr12	52,944,484	4.87
Chr13	56,178,325	5.17
Chr14	48,846,382	4.49
Chr15	38,257,336	3.52
Chr16	45,095,081	4.15
Chr17	39,308,995	3.61
Chr18	39,552,249	3.64
Chr19	18,967,985	1.74
Total	1,085,035,931	99.76%

Table 3. Statistical results of the 19 pseudochromosomes of *Anguilla bicolor pacifica*.

To improve contig contiguity, final scaffolds were generated using `RagTag` (ver. 2.1.0)[23]. In scaffolding process, chromosomes of *A. japonica* genome were used as a reference[16]. During this process, gaps, indicated by "N" characters, were inserted between adjacent query sequences. These gaps represent regions within the sequences that remained unidentified. This step included reordering, orienting, and connecting the sequences using these gaps. Consequently, 405 contigs were integrated into 30 linear scaffolds. This final assembly comprised 19 pseudochromosome-level scaffolds (99.76%) and 11 unplaced scaffolds (0.24%) with an N50 value of 61.07 Mb and a total length of 1.09 Gb (Fig. 1a, Tables 2, 3).

Species	No. of Scaffolds	No. of Chromosomes	Longest Size (Mb)	GC Content(%)	Total Size(Gb)	N50 (Mb)/L50	BUSCO analysis (C, [S, D], F, M) (%)	References
Anguilla bicolor pacifica	30	19	94.4	43.5	1.09	61.07/8	96.4, [87.7, 8.7], 1.0, 2.6	The present study
A. japonica	86	19	94.29	43.64	1.03	58.7/8	94.2, [86, 8.2], 2, 3.8	Wang *et al.*[16]
A. anguilla	1,466	19	90.99	43.51	1.03	56/8	98.4, [89.1, 9.2], 0.4, 1.2	Elise *et al.*[17]
A. rostrata	967	19	85.92	43.73	0.98	56.7/8	98.2, [89.1, 9.1], 0.6, 1.2	GCA_018555375.3
A. obscura	43,284	—	0.81	43.45	0.88	0.54/3,803	78.9, [72.9, 6], 6.6, 14.5	GCA_901111295.1
A. marmorata	38,756	—	0.96	43.33	0.89	0.65/3,157	80.9, [74.9, 6], 6.2, 12.9	GCA_901111315.1
A. megastoma	40,052	—	1.28	43.33	0.88	0.62/3,291	79.9, [73.8, 6.1], 6.5, 13.6	GCA_901111305.1

Table 4. Comparative analysis of the genomes of six anguillids and *Anguilla bicolor pacifica*.

Repeat type	Number	Length (bp)	% of genome
DNA Transposons	338,995	125,214,892	11.51
SINEs	66,166	9,567,464	0.88
LINEs	125,525	32,128,877	2.95
LTR Elements	38,838	13,276,593	1.22
Unclassified	577,431	131,389,954	12.08
Small RNAs	3,346	819,667	0.08
Simple Repeats	558,481	30,100,792	2.77
Total Interspersed		311,577,780	28.65

Table 5. Statistics of repetitive sequences in the genome of *Anguilla bicolor pacifica*.

To evaluate the completeness of the assembled short-finned eel genome, benchmark universal single-copy orthologs (`BUSCO`) (ver. 5.4.3)[24] was used to compare the 3,640 orthologous genes present in Actinopterygii_odb10. The GC contents and genome sizes of the seven anguillid species were found to be comparable (Table 4). Additionally, `Bowtie2` (ver. 2.4.5)[25] was used to align Illumina short reads generated from DNA using the following parameters:`--no-unal --very-sensitive-local`. This resulted in an alignment ratio of 99.41%. Finally, genome quality was examined using `QUAST` (ver. 5.2.0)[26].

Transcriptome sequencing and assembly. Four different tissue types, namely, the eye, heart, liver, and muscle, were collected from an individual eel. All collected samples were immediately preserved in RNAlater and stored at −80 °C until RNA extraction. Total RNA was extracted from the four samples using a TruSeq Stranded mRNA sample preparation kit following the manufacturer's protocol. Complementary DNA libraries were constructed and sequenced using the Illumina NovaSeq 6000 platform to generate 5.24–6.91 Gb of paired-end reads. A total of 21.83 Gb of clean reads was obtained using `Trim_galore` and assembled using `Trinity` (ver. 2.15.1)[27] through the default option (see Table 1).

Genome structure annotation. To identify and screen repetitive sequences within the genomes of short-finned eels, we integrated homology- and *de novo*-based prediction approaches using `RepeatModeler` (ver. 2.0.1) and `RepeatMasker` (ver. 4.1.2)[28,29]. *A. bicolor pacifica* repeat library was annotated by `RepeatModeler` with the National Center for Biotechnology Information (NCBI) searching engine `RMBlast` (ver. 2.9.0). This custom repeat library and two repeat libraries from *Actinopterygii* and *Anguilla* in Dfam[30] were used by `RepeatMasker`. A total of 28.39% of the repetitive sequences were present in the genome of *A. bicolor pacifica* (Table 5).

`BRAKER` Pipeline (ver. 3.0.6)[31] was used to predict the gene models in the genome of *A. bicolor pacifica*. This process began with soft-masking repeats in the genomes generated using `RepeatModeler` and `RepeatMasker`. `GeneMark-ETP` (ver. 1)[32] was used to generate hints from the RNA-Seq and protein data. For the protein data, we combined the Metazoa ortholog data from OrthoDB 11[33] and six actinopterygian species (*A. anguilla*, GCF_013347855.1; *Danio rerio*, GCF_000002035.6; *Pleuronectes platessa*, GCF_947347685.1; *Poecilia reticulata*, GCF_000633615.1; *Scleropages formosus*, GCF_900964775.1; and *Takifugu rubripes*, GCF_901000725.2). To train the gene sets, *ab initio* gene prediction was performed using `Augustus` software (ver. 3.4.0)[34], incorporating hints provided by `GeneMark-ETP`. Finally, the results were integrated with those of `TSEBRA`[35]. Gene models were annotated by combining evidence from homology, *de novo*, and transcriptome data, yielding 23,095 non-redundant protein-coding genes. The BUSCO analysis identified 3,448 (94.7%) actinopterygian orthologous genes (Table 6).

Genome annotation. The functions of the integrated gene models were annotated using the SWISS-PROT protein database[36] and the NCBI non-redundant database (https://www.ncbi.nlm.nih.gov/protein). The `diamond` (ver. 2.1.9.163)[37] blastx was used with the following parameters: `--dbsize 530000000 --max-targetseqs. 1 --outfmt 6 --evalue 1e-5`. Furthermore, we used `eggNOG-mapper`

Gene structure statistics		
No. of protein-coding gene	23,095	
No. of complete BUSCO orthologues (%)	3,448 (94.7)	
Functional annotation	**Number**	**Percent (%)**
SwissProt	21,803	94.41
Nr	22,784	98.65
GO	15,334	66.4
KEGG	16,319	70.66
KOG	22,090	95.65
Pfam	21,347	92.43
Annotated	22,786	98.66
Unannotated	298	1.34

Table 6. Protein coding genes statistics and functional annotation results of the *Anguilla bicolor pacifica* genome.

(ver. 2.1.8)[38] to annotate their functions against the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), euKaryotic Orthologous Group (KOG), and protein families (Pfam) databases. The "`--evalue 1e-5 -m diamond`" parameters in eggnog-mapper were applied. In summary, 98.66% of gene models were functionally annotated using publicly accessible databases (Table 6).

Genome-wide collinearity analysis. Two anguillid genomes were used for comparative analyses. Macrosynteny pairs between short-finned and Japanese eels were obtained using `MCscan` with the default option (https://github.com/tanghaibao/jcvi/wiki/MCscan, Python version)[39]. Macrosynteny blocks were visualized using the Python scripts provided by `MCscan`. The 19 pseudochromosomes from *A. bicolor pacifica* showed a highly conserved collinear relation with the chromosomes of *A. japonica* (Fig. 1b).

Data Records

The raw sequencing data for this study are deposited in the NCBI under BioProject ID: PRJNA1073276. Illumina, transcriptome, and PacBio sequencing data are available under the Sequence Read Archive ID: SRR27869073–SRR27869078[40]. The assembled genome has been deposited in the GenBank database under the accession number JBDGNX020000000[41]. Additionally, assembled genome and annotations can be downloaded from Figshare[42] under https://doi.org/10.6084/m9.figshare.25139891. All data sets used in this study are available at: http://eyunlab.cau.ac.kr/shortfinned_eel.

Technical Validation

Evaluation of genome assembly and annotation. Five methods were applied to evaluate the completeness, accuracy, and contiguity of the *A. bicolor pacifica* genome assembly. These included statistics of N50, `BUSCO` analysis, mapping of short reads of DNA to the genome, and comparison of synteny blocks in the genomes of *A. bicolor pacifica* and *A. japonica*. Furthermore, the total size of the assembled genome is similar to that estimated by `jellyfish`. All assessments indicated that the genome assembly was contiguous, and of high quality.

Code availability

The software used in this study is publicly available. Parameters for all commands used in this study were described in Method. The default parameters were applied for any commands where specific parameters were not mentioned. No custom script or code was used in this study.

Received: 14 February 2024; Accepted: 21 August 2024;
Published online: 30 August 2024

References

1. Sugeha, H. Y. & Genisa, M. U. External and internal morphological characteristics of glass eels *Anguilla bicolor bicolor* from the Cibaliung River Estuary, Banten, Indonesia. *OLDI* **41**, 37–48 (2015).
2. Marini, M. *et al*. Genetic diversity, population structure and demographic history of the tropical eel *Anguilla bicolor pacifica* in Southeast Asia using mitochondrial DNA control region sequences. *GECCO* **26**, e01493 (2021).
3. Yuan, Y., Yuan, Y., Dai, Y., Gong, Y. & Yuan, Y. Development status and trends in the eel farming industry in Asia. *N. Am. J. Aquacult.* **84**, 3–17 (2022).
4. Tanaka, H. Progression in artificial seedling production of Japanese eel *Anguilla japonica*. *Fish. Sci.* **81**, 11–19 (2015).
5. Liao, I. C., Hsu, Y. K. & Lee, W. C. Technical innovations in eel culture systems. *Rev. Fish. Sci.* **10**, 433–450 (2002).
6. Guhl, B., Stürenberg, F. J. & Santora, G. Contaminant levels in the European eel (*Anguilla anguilla*) in North Rhine-Westphalian rivers. *Environ. Sci. Eur.* **26**, 26 (2014).
7. Belpaire, C. G. J. *et al*. Decreasing eel stocks: survival of the fattest? *Ecol. Freshwat. Fish* **18**, 197–214 (2009).
8. Muthmainnah, D., Honda, S., Suryati, N. K. & Prisantoso, B. I. Understanding the current status of anguillid eel fisheries in Southeast Asia. *Fish for the People* **14**, 19–25 (2016).
9. Cuvin-Aralar, M. L., Aya, F. A., Romana-Eguia, M. R. R. & Logronio, D. J. *Nursery culture of tropical anguillid eels in the Philippines* (Aquaculture Department, Southeast Asian Fisheries Development Center, 2019).

10. Aya, F. A. & Garcia, L. M. B. Cage culture of tropical eels, *Anguilla bicolor pacifica* and *A. marmorata* juveniles: Comparison of growth, feed utilization, biochemical composition and blood chemistry. *Aquacult. Res.* **53**, 6283–6291 (2022).
11. Arai, T. Do we protect freshwater eels or do we drive them to extinction? *Springerplus* **3**, 534 (2014).
12. Ege, V. A *revision of the genus Anguilla Shaw*. Vol. 16 8-256 (Brill, 1939).
13. Watanabe, S., Miller, M. J., Aoyama, J. & Tsukamoto, K. Evaluation of the population structure of *Anguilla bicolor* and *A. bengalensis* using total number of vertebrae and consideration of the subspecies concept for the genus *Anguilla*. *Ecol. Freshwat. Fish* **23**, 77–85 (2014).
14. Arai, T. Ecology and evolution of migration in the freshwater eels of the genus *Anguilla Schrank*, 1798. *Heliyon* **6**, e05176 (2020).
15. Wright, R. M. *et al*. First direct evidence of adult European eels migrating to their breeding place in the Sargasso Sea. *Sci. Rep.* **12**, 15362 (2022).
16. Wang, H. *et al*. A Chromosome-level assembly of the Japanese eel genome, insights into gene duplication and chromosomal reorganization. *GigaScience* **11**, giac120 (2022).
17. Parey, E. *et al*. Genome structures resolve the early diversification of teleost fishes. *Science* **379**, 572–575 (2023).
18. Barth, J. M. I. *et al*. Stable species boundaries despite ten million years of hybridization in tropical eels. *Nat. Commun.* **11**, 1433 (2020).
19. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10–12 (2011).
20. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinform.* **27**, 764–770 (2011).
21. 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).
22. 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).
23. Alonge, M. *et al*. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022).
24. Simão, F. A. *et al*. assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinform.* **31**, 3210–3212 (2015).
25. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
26. Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. & QUAST Quality assessment tool for genome assemblies. *Bioinform.* **29**, 1072–1075 (2013).
27. Haas, B. J. *et al*. *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
28. Smit, A. F. A. & Hubley, R. *RepeatModeler Open-1.0*, http://www.repeatmasker.org (2008–2015).
29. Smit, A. F. A., Hubley, R. & Green, P. *RepeatMasker Open-4.0* http://www.repeatmasker.org (2013–2015).
30. Hubley, R. *et al*. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81–89 (2016).
31. Gabriel, L. *et al*. BRAKER3: fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *bioRxiv* (2023).
32. Bruna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP: automatic gene finding in eukaryotic genomes in consistency with extrinsic data. *bioRxiv* (2023).
33. Kuznetsov, D. *et al*. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **51**, D445–D451 (2022).
34. Stanke, M. *et al*. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
35. Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M. & Stanke, M. TSEBRA: transcript selector for BRAKER. *BMC Bioinform* **22**, 566 (2021).
36. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
37. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
38. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional annotation, orthology assignments and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
39. Tang, H. *et al*. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
40. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRP488076 (2024).
41. Choi, H., Nam, J., Yang, S. & Eyun, S. *Anguilla bicolor pacifica*, whole genome sequencing project. *GenBank*. https://identifiers.org/ncbi/insdc:JBDGNX020000000 (2024).
42. Choi, H., Nam, J., Yang, S. & Eyun, S. Chromosome-level genome assembly and gene annotation of short-finned eel (*Anguilla bicolor pacifica*). *figshare*. https://doi.org/10.6084/m9.figshare.25139891.v5 (2024).

Acknowledgements

This work was supported by the National Research Foundation of Korea (2022R1A2C4002058) and the Korea Institute of Marine Science & Technology Promotion (RS-2022-KS221676) funded by the Ministry of Oceans and Fisheries.

Author contributions

S.E. supervised and conceived the project. H.C. and J.N. collected the sample. J.N. performed the experiments. H.C. performed bioinformatics analysis. H.C., S.Y. and S.E. wrote the article. All authors contributed to discussions and interpretations of the results and edited 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/s41597-024-03817-9.

Correspondence and requests for materials should be addressed to H.C. or S.-i.E.