

RESEARCH NOTE

Open Access



Improvement of the ayu (*Plecoglossus altivelis*) draft genome using Hi-C sequencing

Masatoshi Nakamoto¹ and Takashi Sakamoto^{1*}

Abstract

Objective The ayu or sweetfish *Plecoglossus altivelis* is ray-finned fish that is widely distributed in East Asia. The genome size of ayu was estimated at approximately 420 Mb. Previously, we reported on ayu draft genome assembly by whole-genome shotgun using Illumina short reads and PacBio long reads; however, the assembly was not to chromosome level. Therefore, to improve the draft genome sequence of ayu to chromosome level, we performed *in situ* Hi-C sequencing as a source of linkage information.

Results The ayu genome assembly yielded 28 large scaffolds that corresponded to the karyotype of ayu ($n = 28$). The resulting ayu genome assembly has a N50 scaffold length of 17.0 Mb, improved from 4.3 Mb. The high-quality reference genome will be helpful for phylogenetic research on bony fishes and for breeding programs in ayu aquaculture.

Keywords Ayu, Sweetfish, *Plecoglossus altivelis*, Genome quality, Hi-C

Introduction

The ayu or sweetfish *Plecoglossus altivelis* is a ray-finned fish with a wide distribution in East Asia [1–3]. The ayu is a typical amphidromous fish and an economically important aquaculture species in Japan. The species belongs to the teleost group Stomiati and the order Osmeriformes; Stomiati is phylogenetically classified as a sister group of the Neoteleostei. The divergence of Protacanthopterygii (which includes salmon and pike) and the common ancestor of Stomiati and Neoteleostei is estimated to have occurred approximately 190 million years ago [4]. Thus, the ayu holds an important position in teleost fish evolution.

Previously, we reported the ayu draft genome by whole-genome shotgun assembly using Illumina short reads and PacBio long reads [5]. The ayu genome assembly yielded 4,035 scaffolds longer than 1,000 bp. The longest scaffold was 16.8 Mb, with an N50 scaffold length of 4.3 Mb. Scaffolds of the ayu genome assembly were anchored to genetic linkage maps using ALLMAPS [6]; 90.7% of the scaffolds were anchored to linkage maps, and 72.4% were oriented. Thus, the draft genome of ayu was incomplete and the continuity of the assemblies relied on the sequencing and assembly methods.

Hi-C analysis captures the spatial conformation of chromatin [7, 8]. To characterize the three-dimensional architecture of whole genomes, Hi-C detects the physical contacts between chromatin regions through digestion of cross-linked DNA molecules with restriction enzymes and proximity ligation between closely contacted genomic regions. To improve the draft genome sequence of ayu, we performed *in situ* Hi-C sequencing as a source of linkage information and constructed chromosome-level scaffolds.

*Correspondence:

Takashi Sakamoto
takashis@kaiyodai.ac.jp

¹Department of Marine Biosciences, Tokyo University of Marine Science and Technology, Tokyo 108-8477, Japan



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

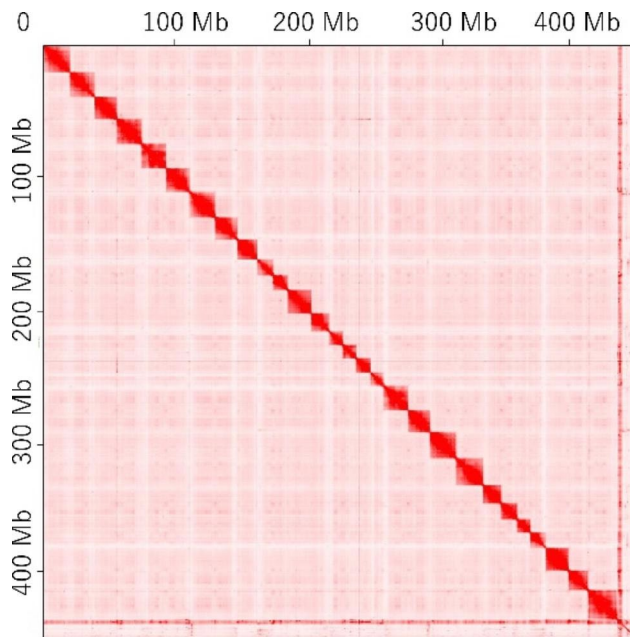


Fig. 1 Contact map for the ayu Hi-C scaffolds; the ayu genome assembly scaffolded 28 chromosomes based on Hi-C data. The interaction probability between pairs of genomic regions is indicated using the Juicebox Assembly Tools standard color scheme.

Main text

Materials and methods

The Hi-C sequencing library was constructed using the Proximo Hi-C Kit for Animal Samples (v3.0; Phase Genomics, Seattle, WA) according to the manufacturer's instructions. As the input sample, 0.3 g of fin tissue obtained from one male ayu individual was rapidly chilled in liquid nitrogen and then ground to a powder. The libraries were quantified using a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA), and the size profile was analyzed on the TapeStation system with the D1000 ScreenTape assay (Agilent, Santa Clara, CA). A library fragment size was approximately 1000 bp. Sequencing of the Hi-C library was carried out on an Illumina HiSeq 4000 system, by Eurofins Genomics K.K. (Tokyo, Japan), with 100-bp paired-end sequencing. The Hi-C data have been deposited in the DDBJ Sequence Read Archive (DRA) under accession number DRA013867. Low-quality bases were removed using the tool Trimmomatic (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:15 TRAILING:15 SLIDING-WINDOW:4:15 MINLEN:50) [9]. The Hi-C reads were aligned to the ayu draft genome (DDBJ accession numbers BNHK01000001–BNHK01004035) using Juicer [10]. 36.64% read pairs were mapped that span at least 10 kbp. Candidate chromosomal scaffolds were constructed with the 3D *de novo* assembly (3D-DNA) pipeline with the parameter `-r 5 -i 10,000` [11]; the candidate scaffolds were then manually reviewed using Juicebox Assembly

Tools [12]. The Hi-C assembly, the linkage map anchored assembly, medaka (*Oryzias latipes*) genome and northern pike (*Esox lucius*) genome were compared using Synima and visualized using Circos [13, 14]. The genome sequences of medaka were obtained from the Ensembl [15].

The genome sequences of northern pike were obtained from the NCBI (accession numbers GCF_004634155).

Results and discussion

After Hi-C sequencing and quality trimming, 296.1 million paired-end reads (read length 100 bp) were obtained. As a result of scaffolding by Hi-C, the N50 of the ayu genome assembly was improved from 4.3 Mb to 17.0 Mb, and the longest scaffold was improved from 16.8 Mb to 22.4 Mb. Contact maps for the ayu Hi-C data indicated that the ayu genome assembly was constructed of 28 large scaffolds that contained 97.5% (435.7 Mb) of the nucleotides of the draft genome (Fig. 1). The number of large scaffolds corresponded to the karyotype of ayu ($n=28$) [16]. The improved genome assembly has been deposited in the DDBJ database under the accession number BROE01000001–BROE01004119. Finally, we compared the Hi-C assembly and the previous linkage map anchored assembly and the genome of other teleost fishes. (Fig. 2). These assemblies corresponded almost one-to-one. The major difference was that linkage group 24 and 26 corresponded to Hi-C chromosome 24. Hi-C chromosome 26 was not mapped to the linkage map. Hi-C chromosome 24 corresponded to medaka chromosome 20 and northern pike chromosome 21. Hi-C chromosome 26 corresponded to medaka chromosome 4 and northern pike chromosome 8. Ayu sex determining gene *Amhr2bY* was located in Hi-C chromosome 26, indicating that Hi-C chromosome 26 is sex chromosome of ayu. Further study will be needed to resolve the differences between the Hi-C data and the linkage maps.

Our results show that the ayu genome sequence was scaffolded to chromosome level with Hi-C data. A high-quality reference genome is useful to detect structural variants. We anticipate that our ayu genome assembly will contribute to research on teleost fish evolution as well as to the aquaculture of ayu as a basic food resource.

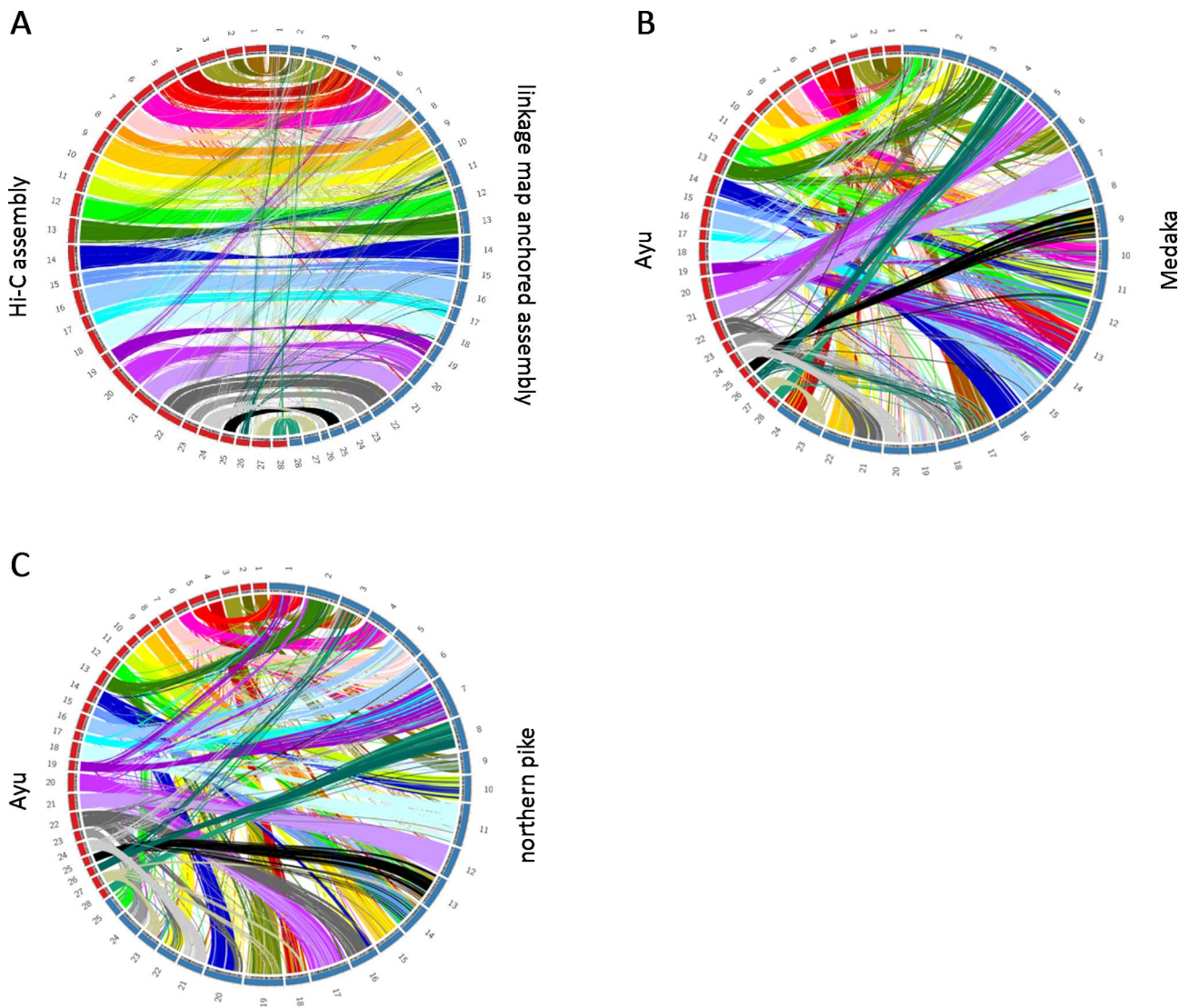


Fig. 2 Synteny analysis for ayu Hi-C chromosome. (A) Comparison of the Hi-C assembly with the linkage map anchored assembly of the ayu *Plecoglossus altivelis* genome, visualized with Circos software. Circos plots between Hi-C assembly of ayu and genome of medaka (B) and northern pike (C).

Abbreviations

Mb Mega base pair
DRA DDBJ Sequence Read Archive

Acknowledgements

The authors thank Cynthia Kulongowski with Edanz (<https://jp.edanz.com/ac>) for editing the language of a draft of this manuscript.

Authors' contribution

Conceived and designed the experiments: TS and MN. Performed the experiments: MN. Analyzed the data: MN. Wrote the paper: MN and TS.

Funding

This work was supported by funds from the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant numbers 20H00431 to TS, and 18K05816 to MN) and from the Japan International Cooperation Agency (JICA) for Science and Technology Research Partnership for Sustainable Development (SATREPS) to TS.

Data Availability

The improved genome assembly has been deposited in the DDBJ database under the accession number BROE01000001-BROE01004119.

Declarations

Ethics approval and consent to participate

All animal experiments and methods were performed in accordance with the guidelines and approval of the Institutional Animal Care and Use Committee of the Tokyo University of Marine Science and Technology.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 2 August 2022 / Accepted: 19 May 2023

Published online: 30 May 2023

References

- Iguchi K, Tanimura Y, Takeshima H, Nishida M. Genetic variation and geographic population structure of amphidromous ayu *Plecoglossus altivelis* as examined by mitochondrial DNA sequencing. *Fisheries Sci.* 1999;65:63–7. <https://doi.org/10.2331/fishsci.65.63>.
- Iguchi K, Nishida M. Genetic biogeography among insular populations of the amphidromous fish *Plecoglossus altivelis* assessed from mitochondrial DNA analysis. *Conserv Genet.* 2000;1:147–56. <https://doi.org/10.1023/A:1026582922248>.
- Kwan YS, Song HK, Lee HJ, Lee WO, Won YJ. Population genetic structure and evidence of demographic expansion of the ayu (*Plecoglossus altivelis*) in East Asia. *Anim Syst Evol Divers.* 2012;28:279–90. <https://doi.org/10.5635/ASED.2012.28.4.279>.
- Betancur-RR, Wiley EO, Arratia G, Bailly N, Miya M, Lecointre G, Ortí G. Phylogenetic classification of bony fishes. *BMC Evol Biol.* 2017;17:162. <https://doi.org/10.1186/s12862-017-0958-3>.
- Nakamoto M, Uchino T, Koshimizu E, Kuchiishi Y, Sekiguchi R, Wang L, Sudo R, Endo M, Guiguen Y, Scharl M, Postlethwait JH, Sakamoto T. A Y-linked anti-Müllerian hormone type-II receptor is the sex-determining gene in ayu, *Plecoglossus altivelis*. *PLoS Genet.* 2021;17(8):e1009705. <https://doi.org/10.1371/journal.pgen.1009705>.
- Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, Schnable PS, Eric Lyons E, Lu J. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol.* 2015;16:3. <https://doi.org/10.1186/s13059-014-0573-1>.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009;326(5950):289–93. <https://doi.org/10.1126/science.1181369>.
- van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp.* 2010;39:e1869. <https://doi.org/10.3791/1869>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–2020. <https://doi.org/10.1093/bioinformatics/btu170>.
- Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Lieberman Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 2016;3(1):95–8. <https://doi.org/10.1016/j.cels.2016.07.002>.
- Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. De-novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* 2017;356(6333):92–5. <https://doi.org/10.1126/science.aal3327>.
- Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 2016;3(1):99–101. <https://doi.org/10.1016/j.cels.2015.07.012>.
- Farrer RA. Synima: a synteny imaging tool for annotated genome assemblies. *BMC Bioinformatics.* 2017;18:507. <https://doi.org/10.1186/s12859-017-1939-7>.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19:1639–45. <https://doi.org/10.1101/gr.092759.109>.
- Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amodè MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, Brooks L, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, Donaldson S, El Houdaigui B, El Naboulsi T, Fatima R, Giron CG, Genez T, Martinez JG, Guijarro-Clarke C, Gymer A, Hardy M, Hollis Z, Hourlier T, Hunt T, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Marugán JC, Mohanan S, Mushtaq A, Naven M, Ogeh DN, Parker A, Parton A, Perry M, Piližota I, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Pérez-Silva JG, Stark W, Steed E, Sutinen K, Sukumar R, Sumathipala D, Suner MM, Szpak M, Thomann A, Tricomi FF, Urbina-Gómez D, Veidenberg A, Walsh TA, Walts B, Willhoft N, Winterbottom A, Wass E, Chakiachvili M, Flint B, Frankish A, Giorgetti S, Haggerty L, Hunt SE, Ilsley GR, Loveland JE, Martin FJ, Moore B, Mudge JM, Muffato M, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Dyer S, Harrison PW, Howe KL, Yates AD, Zerbino DR. Flicek PEnsembl 2022. *Nucleic Acids Res.* 2022;50(1):D988–95. <https://doi.org/10.1093/nar/gkab1049>.
- Ueno K, Ikenaga Y, Kariya H. Potentiality of application of triploidy to the culture of ayu, *Plecoglossus altivelis* Temminck et Schlegel. *Jpn J Genet.* 1986;61:71–7.

Publisher's Note

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