Supplementary Methods

Details of manual curation

In every strain, we call as v3.0.0 the initial assembly automatically generated with Hifiasm and RagTag, and the final assembly as v3.1. The Hd-rR v3.1 assembly reached the gapless, telomere-to-telomere assembly, and for the other two, HNI and HSOK, we achieved nearly complete assemblies. Specifically, we identified potentially misassembled regions (PMRs) in the initial scaffolds in light of abnormality in read depths and the conflicts in nucleotide sequences between scaffolds and mapped HiFi/ONT reads from which the scaffolds are assembled.

The curation process consisted of 1) sequence patching from another genome assembler, Verkko, 2) consensus sequence refinement in each PMR using ONT-UL reads spanning the PMR, and 3) telomere extension using ONT-UL reads. During the curation process of the Hd-rR genome assembly from v3.0.0 to v3.1, in total 164, 116, and 44 loci were modified for each category, and all 19 gaps in the initial chromosomal scaffolds were filled. Below we describe the details of each assembly version update for each strain.

The T2T Hd-rR assembly

From v3.0.0 (initial assembly with Hifiasm) to v3.0.1:

- Identified Y-specific sequences and its corresponding X-specific region in chromosome 1 based on the sequencing depths calculated by mapping each of the female and male HiFi reads. Suppose the global diploid sequencing coverage is given as *C*. Then, X-specific contigs should exhibit a sequencing depth of *C* for the female read dataset and *C*/2 for the male read dataset, whereas Y-specific contigs should exhibit a sequencing depth of 0 for the female read dataset and *C*/2 for the male read dataset. In addition, X-/Y-specific contigs should exist as branching paths (i.e. primary contigs and alternative contigs) in a connected component representing chr1 in the assembly graph. Using the information of sequencing depths and graph connectivity, we found that the entire X-specific region was assembled and embedded in a large primary contig of chr1 and the Y-specific region was divided and contained in two alternative contigs (atg000001l and atg000005l) corresponding to the X-specific region. We concatenated these two Y-specific sequences into a single ~12-Mbp sequence and polished the gap between the two sequences using ONT-UL reads spanning the gap. We named the Y-specific sequence as chr1_Y.

- Determined a representative unit sequence for each of the 45S and 5S ribosomal DNA arrays. We did this by mapping human ribosomal DNA array sequences to the Hd-rR scaffolds, finding the tandemly repeated sequences in them, and extracting the unit sequence of the tandem repeat array.
- Removed short contigs placed outside telomeres by RagTag and re-assigned them as unlocalized contigs. This is likely to be caused because the reference genome used as the guide for RagTag was the prior v2.2.4 assembly published in 2017, which did not have telomere sequences at most of the chromosomes.

From v3.0.1 to v3.0.2:

- Replaced sequences at 162 PMRs with those of Verkko contigs.
- Modified 69 PMRs using spanning ONT reads.

From v3.0.2 to v3.0.3:

- Modified 47 PMRs using spanning ONT reads.
- The 24 chromosomal scaffolds became gapless at this stage, although there existed several potentially misassembled regions including rDNA arrays.

From v3.0.3 to v3.0.4:

- Extended telomere sequences at both ends of the chromosomes, except the 3'-end of chromosome 9, the 5'-end of chromosome 17, the 5'-end of chromosome 19, and the 3'-end of chromosome 23. Among these four exceptions, the 5'-end of chromosome 23 was not extended at this stage because of a highly complicated 45S ribosomal DNA array at the location, although the other three were not because those ends did already have sufficiently long telomere sequences.
- Removed all unlocalized sequences from the assembly. This resulted in excessive read mappings at six locations on chromosomal sequences such as regions near 5S and 45S rDNA arrays and subtelomeric regions, which were resolved in the subsequent curation steps below.
- Analyzed sequence variations in 45S rDNA arrays at the subtelomeric regions of chromosome 18 and chromosome 20 where only several copies of rDNA unit sequences are found. We discovered multiple ONT-UL reads containing minor sequence patterns that were different from the sequence of the chromosomal scaffolds, and concluded that these minor sequences we observed were variations among multiple individuals or somatic variations.
- Similarly to the shorter 45S rDNA regions on chromosome 18 and chromosome 20, from the largest 45S rDNA region on chromosome 23 we discovered at least three distinct sequence patterns. Moreover, there existed a huge (~1.5 Mbp)

tandem repeat array consisting of 144-bp unit sequences at the upstream region of the 45S rDNA array. The largest 5S rDNA array on chromosome 16 also has two tandem repeat arrays of ~1 Mbp in total consisting of 408-bp and 597-bp unit sequences, respectively, at its upstream.

From v3.0.4 to v3.0.5:

- Modified the sequence around the largest 45S rDNA array on chromosome 23. Specifically, we manually concatenated the sequence of chromosome 23 of the v3.0.4 assembly, a part of a Verkko contig, and an unlocalized contig of the v3.0.4 assembly in this order, by finding overlaps between these sequences. The unlocalized contig contained telomere sequences, and thus the chromosome 23 became a T2T sequence. There still existed some fluctuation in sequencing depths calculated with ONT-UL reads mapped to the assembled sequences, but given the minor variations on 45S rDNA arrays, we accepted this fluctuation.
- Modified the sequence of the huge tandem repeat array at the upstream of the 5S rDNA array on chromosome 16. Similarly to the modification for the 45S rDNA array above, we glued a Verkko contig and ONT-UL reads with the sequence of chromosome 16 of the v3.0.4 assembly. We first placed a raw ONT-UL read and then polished the sequence with other ONT-UL and HiFi reads mapped to the region.
- Replaced a minor 45S rDNA array sequence at the 5'-end of chromosome 18 with the consensus sequence manually generated.
- Identified two additional short Y-specific sequences, chr1_alt_atg000009l_r and chr1_alt_atg000026l_r, other than the main ~12 Mbp sequence named chr1_Y. We named them as chr1_Y_2 and chr1_Y_3, respectively.

From v3.0.5 to v3.0.6:

- Polished the T2T chromosomal sequences with the HiFi and ONT-UL reads. Specifically, we first mapped these reads to the T2T chromosomal sequences and called small variants on them using DeepVariant. We extracted homozygous variants and incorporated these variants into the T2T chromosomal sequences using "bcftools consensus" so that reads and assembled sequences became consistent. After this polish, the number of homozygous variants detected with HiFi and ONT-UL reads was reduced from 11,922 and 19,762 to 1,167 and 7,801, respectively. The remaining variants are mostly homopolymer variants, which are thought to be due to sequencing errors in reads, not assemblies.

From v3.0.6 to v3.1:

- Took reverse complement of the sequences for each of the twelve chromosomes 1, 2, 7, 8, 9, 11, 12, 13, 18, 19, 22, 24 so that each chromosome begins with its short p-arm, as we first revealed the accurate locations of centromeres and thus p-arms and q-arms. We also flipped the Y-specific sequences of chr1_Y, chr1_Y_2, and chr1_Y_3. To make the prior assembly compatible with the flipped assembly, we also flipped the sequences of the same chromosomes of the published v2.2.4 assembly and called the flipped assembly as the v2.3 assembly.

The HNI assembly

From v3.0.0 (initial assembly with hifiasm) to v3.0.1:

- Removed short contigs placed outside telomeres by RagTag and re-assigned them as unlocalized contigs.
- Cut a false large duplication on chr1 and re-assigned it as an unlocalized contig, which was supposed to be a misplacement of a Y-specific sequence.

From v3.0.1 to v3.0.2:

- Replaced sequences at 9 PMRs with those of Verkko contigs.

From v3.0.2 to v3.0.3:

- Identified X-/Y-specific sequences from unlocalized contigs. Because the sequence of chr1 on its sex-specific region was a mosaic of X- and Y-specific sequences, we swapped unlocalized X-specific sequences and their corresponding Y-specific sequences contained in chr1 and then performed scaffolding of Y-specific sequences using both the X-specific sequence and Hd-rR's X-/Y-specific sequences to make scaffold sequences of the X-/Y-specific regions.

From v3.0.3 to v3.1:

- Took reverse complement of the sequences for each of the twelve chromosomes 1, 2, 7, 8, 9, 11, 12, 13, 18, 19, 22, 24 so that each chromosome begins with its short p-arm.

The HSOK assembly

From v3.0.0 (initial assembly with hifiasm) to v3.0.1:

- Removed short contigs placed outside telomeres by RagTag and re-assigned them as unlocalized contigs.
- Removed false duplications near centromeres.

From v3.0.1 to v3.0.2:

- Cut a false large duplication on chr1 and re-assigned it as an unlocalized contig, which was supposed to be a misplacement of a Y-specific sequence.
- Identified Y-specific sequences from unlocalized contigs.

From v3.0.2 to v3.0.3:

- Performed scaffolding of Y-specific sequences. For about a half of the Y-specific sequences, we reconstructed the Y-specific sequence containing the *DMY* gene based on the X-specific sequence in chr1. For the remaining half, the sequence content was too diverged from the X-specific region to scaffold, and thus we kept them as unlocalized.
- Took reverse complement of the sequences for each of the twelve chromosomes 1, 2, 7, 8, 9, 11, 12, 13, 18, 19, 22, 24 so that each chromosome begins with its short p-arm.

From v3.0.3 to v3.1:

- Corrected a misassembly where a portion of the sequence at the end of chr18 was mistakenly included in chr23, which was mistakenly connected probably due to a large tandem repeat. The sequence was moved to its proper location.

Details on updates from v2

Here we detailed some additional updates in the v3 assembly from the v2 assembly published in 2017[1].

Centromere locations

The centromere positions for chr4 and chr10 are different with those shown in Fig. 2 of the medaka v2 assembly paper. In the previous study, chr4 was classified as metacentric since the HSOK assembly had a large centromeric sequence at the metacentric position although a short centromeric sequence was found at the acrocentric position on chr4 in the Hd-rR assembly. In the v3 assemblies, all strains are now acrocentric. For chr10, in the previous study, the centromere could not be assembled in Hd-rR/HSOK, and only a small portion could be assembled at the metacentric position in HNI, so it was classified as metacentric. In the v3.1 assemblies, chr10 is now acrocentric in Hd-rR/HSOK and dicentric in HNI. It is likely that only the major centromere sequence at the middle of the dicentric chromosome of HNI was observed in the v2 assembly.

Distinct patterns of satellite array sequence and CpG methylation in acrocentric centromeres

Despite very limitedly, the distinctive centromere sequence architecture in acrocentric chromosomes was observed in the previous assembly in a few chromosomes (Hd-rR chr22 and HSOK chr4; see Fig. 3c and Supplementary Fig. 5 of ref. [1]). In this study we revealed the whole and prevalent picture of the distinct patterns of the acro-centromeres in all the strains (**Fig. 3a,b, Extended Data Fig. 10,11**).

Analysis on mitogenomes

Complete mitochondrial genomes reconstructed in this study were highly consistent with the previous medaka mitogenome study[2]. Specifically, compared to the previous ones, there existed only one and two substitutions in our complete mitogenomes in Hd-rR and HNI, respectively, excluding incompletely assembled regions of ~1 kbp in length in the previous ones. There existed three substitutions plus 42-bp deletion between our HSOK complete mitogenome and previously reported mitogenome of SOK, a slightly different strain from HSOK. These results indicate that the estimated divergence times of the three strains calculated based on mitogenomes remain valid.

Supplementary references

1. Ichikawa, K. *et al.* Centromere evolution and CpG methylation during vertebrate speciation. *Nat Commun* **8**, 1833 (2017).
2. Setiamarga, D. H. E. *et al.* Divergence time of the two regional medaka populations in Japan as a new time scale for comparative genomics of vertebrates. *Biol Lett* **5**, 812–816 (2009).