

A Chromosome-level Assembly of the Japanese Eel Genome, Insights into Gene Duplication and Chromosomal Reorganization

Hongbo WANG³, Hin Ting WAN², Bin WU⁴, Jianbo JIAN⁴, Alice HM NG², Claire Yik-Lok CHUNG⁵, Eugene Yui-Ching CHOW⁵, Jizhou ZHANG⁵, Anderson OL WONG^{1,6}, Keng Po LAI^{1,7}, Ting Fung CHAN^{1,5}, Eric Lu Zhang^{3*}, Chris Kong-Chu WONG^{1,2*}

¹Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), ²Croucher Institute for Environmental Sciences, Department of Biology, ³Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR; ⁴BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China; ⁵School of Life Sciences, State Key Laboratory of Agrobiotechnology, Hong Kong Bioinformatics Centre, the Chinese University of Hong Kong, Hong Kong SAR; ⁶School of Biological Sciences, the University of Hong Kong, Hong Kong SAR; ⁷Key Laboratory of Environmental Pollution and Integrative Omics, Guilin Medical University, Guilin, PR China;

*Corresponding author,

Dr Eric Lu ZHANG

³Department of Computer Science,

Hong Kong Baptist University,

Hong Kong SAR

Email address: ericluzhang@comp.hkbu.edu.hk

Dr Chris KC Wong

¹Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou),

²Croucher Institute for Environmental Sciences,

Department of Biology,

Hong Kong Baptist University,

Hong Kong SAR

Email address: ckcwong@hkbu.edu.hk.

ABSTRACT

Japanese eels (*Anguilla japonica*) are commercially important species that have been harvested extensively for foods. Currently, this and related species (American and European eels) are difficult to breed on a commercial basis. Wild stock is used for aquaculture. Due to pollution, overfishing, and international trafficking, eel populations are declining. The International Union for Conservation of Nature lists Japanese eels as critically endangered and on its red list. Here we presented a high-quality genome assembly for Japanese eels and demonstrated that large chromosome reorganizations occurred in the events of third-round whole-genome duplications (3R-WRD). Following multiple chromosomal fusion and fission rearrangement, the *Anguilla* lineage has reduced the haploid chromosomal number of 19 from the ancestral proto-chromosomal number of 25. Phylogenetic analysis of expanded gene families showed the gene families of olfactory receptors and voltage-gated Ca^{2+} -channel expanded significantly. The expansion of olfactory receptors (group δ and ζ genes) and voltage-gated Ca^{2+} -channel gene families are important for olfaction and neurophysiological functions. Following 3R-WGD, additional tandem (TD) and proximal (PD) duplications occurred to acquire immune-related genes for adaptation. The Japanese eel assembly presented here can be used to study other *Anguilla* species that are related to evolution and conservation.

INTRODUCTION

Fishes are highly diverse species living in many ecological habitats, including freshwater, estuarine, and the ocean⁴¹. Over 99% of fish species are known to be stenohaline, inhabiting either freshwater or marine environments. While euryhaline fishes are known to be diadromous, migrating between freshwater and marine environments in their life cycles³². Catadromous fishes like eels spawn in the sea and migrate to inland freshwater to grow and mature. Eels are ecologically and economically important, serving as indicators of the healthiness of coastal environments and resources in aquaculture. The fish are not bred in captivity⁴⁷. In current practices, glass eels (juvenile life stage) are captured from the wild and raised on farms. Over 90 % of freshwater eels consumed worldwide are farm-raised. Since the 1960s, catches of Anguillid eels, like European and Japanese eels, have declined by over 50-80 %. In a 2014 report from the International Union for Conservation of Nature (IUCN), the American, European, and Japanese eels have been listed as being at the high risk of extinction. The decline in eel populations is abetted by soaring demand from global markets. In addition, overfishing, habitat loss, dams⁷⁵, water pollution²⁶, parasites³⁵, eel larvae predation by mesopelagic fishes⁴⁸, climate change, and altered ocean currents¹⁴ are known to cause population decline.

From the evolutionary perspective, eels are among the most basal extant groups of teleost fishes and close to the non-teleost ray-finned fishes, including holostei (bowfin, gar), chondrostei (sturgeon, paddlefish, starlet), and cladistia (bichir, ropefish), those undergoing the first 2-round of vertebrate genome duplication, occurred before the divergence of ray finned and lobe-finned fishes 450 million years ago⁴². Eels are one of the first groups separated from the majority of the teleost fishes after the teleost-specific of whole-genome duplication (3R-WGD)⁶⁷. The comparison of eels with other teleosts would shed light on fish evolution. In 2012, the first draft sequences of the genomes of Japanese (genome size 1.15 Gb, consisting of 323,776 scaffolds) and European eels (0.923 Gb, N50 of 78Kbp) were published^{36,37}. The Japanese eel's draft genome's annotation was then enhanced using transcriptome data⁶⁰. Moreover, the genome sequence assembly of the European eel (0.86 Gb genome size) was improved using

nanopore sequencing⁴⁶. A draft genome of the American eel (with a total size of 1.41 GB) was published in 2017, and 26,564 genes were annotated⁷³. In 2019, the assembly of a Japanese genome of 1.13 Gb¹⁷ was improved with 256,649 contigs, 41,687 scaffolds, and a scaffold N50 of 1.03M. Currently, only the draft genome is available for Japanese eels. The purpose of this study was to provide high-quality genome assemblies and to understand the evolution of karyotypes in early ray-finned fishes. The genome-scale data can provide ecological and conservation information by identifying adaptive and disease-resistant alleles.

RESULTS

Genome Assembly and Annotation. In this study, MitoZ software was used to assemble and annotate the mitochondrial genome (16.686Kb) of our sample to confirm the species' identity (**Methods**). The data matched with the Japanese eel mitochondrial genome (GenBank ID AB038556.2) of the NR database from NCBI (**Supplementary Fig 1&2**). We hierarchically integrated the sequencing data from different platforms to characterize their strength in *de novo* assembly and annotation (**Supplementary Fig. 3**). The draft genome was generated using ONT contigs followed by error correction and scaffolding based on the genomic spans of different sequencing technologies²⁸ (**Methods**). A high-quality Japanese female eel's reference genome was then generated, through the integration of ONT long reads (234x, 239.64Gb), PacBio CLR (261x, 267Gb), 10x Chromium linked-reads (313x, 319.7Gb), Hi-C data (48x, 48.99 Gb), Illumina short-reads (148x, 151.89Gb) and mate-pair reads (127x, 130.5Gb). The contigs from ONT long reads resulted in a significantly improved N50 (25.82Mb) without losing many complete genes (54.6%) (**Supplementary Table 2**). With reduced assembly errors, the percentage of complete genes increased from 54.6% to 90.1%, indicating a higher base quality (**Supplementary Table 3**). For scaffolding, 10x linked-reads, Bionano, and Hi-C data were used sequentially according to fragment length to increase assembly continuity and assign scaffolds to 19 chromosomes (**Fig. 1, Supplementary Fig 3, and Supplementary Tables 4**). As a result, the genome size is 1.028Gb, the contig N50 is 21.48Mb, and the scaffold N50 is 58.7Mb. The chromosome lengths range from 19.93Mb to 94.28Mb. According to *actinopterygii_oddb10* in the BUSCO database, 94% of the single-copy direct homologs in the Ray-finned Fishes were assembled in Japanese eels (**Supplementary Table 5**). The repeat elements were predicted to account for 30.49% of the whole genome (**Supplementary Tables 6**). The TEs were excluded from gene annotation (**Supplementary Tables 7**). Japanese eels have a higher percentage (30.49%) of repetitive sequences, which may explain their larger genome, compared to European eels (*Anguilla anguilla* 0.979 Gb)⁴⁶. Even so, the Japanese and European eels have a 1:1

correspondence pattern of chromosomes and 19,325 homologous genes, demonstrating their matching structure (**Supplementary Fig 4**).

By combining gene annotations from homology, *de novo*, and transcriptome annotations (**Methods**), we identified 29,982 coding genes (**Table 1**). We functionally annotated 97.44 % (29,219) of these genes (**Supplementary Table 8**) using the publicly available databases (**Methods**). Additionally, 21,606 genes were annotated by all five major protein databases (**Supplementary Fig 5**), with signal transduction pathways most abundant in KEGG (**Supplementary Fig 6**) and KOG (**Supplementary Fig 7**). BUSCO analysis showed that 94.7 % of the single-copy orthologs could be found in the ray-finned fish single-copy direct homology gene database actinopterygii_odb10 (**Supplementary Table 9**). The protein-coding genes in Japanese eels have an average length of 10.2kbp and contain approximately 9 exons (**Table 1**), which has an average length of 1.6kbp (**Supplementary Table 10**). The gene structure of Japanese eels is similar to those of four closely related species (**Supplementary Fig 8**). The genome assembly has a greater number of predicted genes (29,982 genes) than the Atlantic species, European (25,903 genes), and America (26,565 genes) eels. Additionally, 17,095 noncoding RNAs were predicted, including 1,042 transfer RNAs (tRNAs), 1,771 ribosomal RNAs (rRNAs), and 3,974 microRNAs in Japanese eels.

Phylogenomics and demographic history. The orthology analysis of 12 species' coding genes identified 21,653 gene family clusters. *Anguilla japonica*'s genome contains 29,982 coding genes, including 3347 single-copy orthologs, 8204 multiple-copy orthologs, 233 unique paralogs, 12,662 other orthologs, and 5536 unclustered genes. A phylogenetic tree was reconstructed by identifying the fourfold synonymous third-codon transversion (4dTv) loci in the 1,131 single-copy orthologs from the 12 fish species (**Fig 2**). American and European eels diverged from their ancestors about 27.0 million years ago (MYA). With a divergence time of approximately 44.1 MYA, Japanese eel was distant from the Atlantic eel species. The 4dTv-orthology density plot showed that Japanese eels are phylogenically closer to American than European eels. In comparison with the three freshwater eels (Anguilliformes) and tarpons (Elopiformes), the members of the Order

Elopomorpha, their common ancestor, diverged 196.1 MYA. Elopomorpha and Osteoglossomorpha (i.e., arowana) are the closest evolutionary relatives at the basal branch of teleosts, separating 240.9 MYA. *Gadiformes* (e.g., Atlantic cod) and *Cypriniformes* (e.g., medaka, zebrafish) diverged from *Elopomorpha* and *Osteoglossomorpha* at 262.5 MYA. Above are groups of fishes, which had undergone three rounds of whole-genome duplication (3R-WGD). Compared to the outgroups, spotted gars, reed fish, coelacanths, and Australian ghost sharks underwent only 2R-WGD.

Expanded Gene Families and Gene duplication. The expansion and contraction of gene families reflect the evolution of organisms' adaptations to their environments. Ortholog analysis of genes from the 12 species (**Methods**) identified 21,652 gene family clusters. By removing gene families with too many (≥ 200) or too few (≤ 2) genes, we achieved 129,862 genes to evaluate the expansion and contraction of gene families (**Fig 2**). Compared to the nine other species (**Methods**), the three freshwater eels had expanded 771 and contracted 467 gene families, resulting in an increase of 919 and loss of 531 genes, respectively (**Supplementary Table 11**). Among those, the three freshwater eel species exhibited a significant expansion in the olfactory receptor (OR) gene family, which is crucial for detecting odor molecules under varying environmental conditions. A retrospective analysis of the OR receptors across 10 species' genomes was performed and seven types of OR receptors were identified [alpha (α), beta (β), gamma (γ), delta (δ), epsilon (ϵ), zeta (ζ) and eta (η)] based on a previous study⁶⁹. Compared to other fish species, the Japanese eels had a significantly higher number of OR genes (394) (**Fig 3**), located on the four chromosomes - Chr4 (2 genes), Chr9 (153 genes), Chr11 (1 gene), and Chr12 (238 genes). Similarly, the European eel contains 392 OR genes. The δ and ζ genes are the major OR genes in the eels.

Comparing Japanese eel to the other 11 species, 433 gene families increased, with a total increase of 551 genes. A total of 943 genes were lost from 782 gene families (**Supplementary Table 12**). It is interesting to note that Ca^{2+} and K^{+} channel families were identified. Calcium and potassium play significant roles in neuronal excitability,

muscle contraction, fertilization and energy metabolism. Interestingly, the other expanded gene families include (i) the assembly of myosin thick filament in skeletal muscle, (ii) lipoprotein receptor-related protein (metabolic and morphogenetic pathways) and (iii) isocitrate and isopropylmalate dehydrogenases family (carbohydrate and amino acid metabolism).

It was reported that freshwater eels (European and Japanese) had a large number of paralogous pairs, after splitting from the Osteoglossomorpha lineage⁷⁹. The observation suggested 4R-WGD or lineage-specific rediploidization in some duplicated genomic regions. Our data showed that the genome size (1.028G) of Japanese eels is less than half that of the Atlantic salmon (2.97G)⁵⁹, which underwent salmonid-specific 4R-WGD. In addition, we studied the distribution of 4dTv and Ks values of genome-wide direct homologous gene pairs in Japanese eels, European eels, and tarpons. There were 4dTv values of 0.402, 0.386, and 0.317 for *A. japonica*, *A. Anguilla*, and *M. cyprinoides*, respectively (**Fig 4A**, and **Supplementary Fig 9**). Additional WGD events were not detected. We also compared the syntenic blocks at Hox A-D loci with those in spotted gar that underwent 2R-WGD (**Fig 4B**). Japanese eel's genome has eight clusters of Hox loci on chromosomes 1, 2, 3, 8, 11, 13, 15, and 17 while spotted gar has four clusters on chromosomes 4, 11, 12, and 13. Collectively, the data do not support the presence of 4R-WGD in Japanese and European eels.

There are a total of 21,249 duplicated genes identified among the 29,982 coding genes in the Japanese eel genome. Based on their duplication patterns, DupGen_finder (**Methods**) classified the duplicated genes into five categories, (i) 9,890 whole-genome duplicates (WGD, 46.54%), (ii) 1,420 tandem duplicates (TD, 6.68%), (iii) 768 proximal duplicates (PD, 3.61%), (iv) 3,975 transposed duplicates (TRD, 18.71%), and (v) 5,196 dispersed duplicates (DSD, 24.45%). We then calculated the Ks and Ka/Ks values for these five gene categories. Ks distribution indicates that TD and PD revealed additional duplication, post-3R-WGD (**Fig 4C**). In addition, both TD and PD duplicates exhibited high Ka/Ks ratios, an indication of high selection pressure, which was probably related to environmental adaptation. TD and PD duplicated genes are mainly involved in immune

responses (e.g. the production of interleukin-8, virus and biotic stress, somatic hypermutation of immunoglobulin genes, diversification and production of immunoglobulins and immunoreceptors) (**Fig 4D**). Nonetheless, WGD was associated with 32.98 % of the total number of coding genes (29,982) in Japanese eels. Gene duplications in other fish species were also analyzed using the DupGen_finder pipeline (Qiao et al. 2019) and compared. Japanese eels were found to share the same level of WGD duplication of coding genes as arowana (37.60%), as both are extant members of the basal teleost group. However, it differs from the majority of teleosts, such as medaka (6.09%) zebrafish (9.51%) and Atlantic cod (4.68%). In Japanese eel, these duplicated gene functions were associated with neuronal (dendrites, synapses, neuron projections, obsolete synapses) and cell-cell junctions (cellular periphery, cell junctions, integral components of plasma membranes, obsolete plasma membranes, and cell projections). TRD shows a similar profile of changes. In DSD, duplication genes function in microtubules, reproduction (oocyte fate determination, fertilization), and ATP metabolism.

Evolution of chromosome number in Japanese eels. When comparing chromosome numbers of the fishes all undergone 3R-WGD, the haploid chromosome number (n) is 25 for tarpons, arowana, zebrafish; 24 for medaka; 23 for Atlantic cod. Japanese eels have a lower haploid chromosome number ($n = 19$). To assess the extent of inter-chromosomal rearrangements in Japanese eels, we reconstructed the karyotype of the common ancestral teleosts karyotype (ATK) and ancestral eel-tarpon karyotype (AETK) (**Fig 5A**). According to our results, the ATK and AETK had 24 and 25 haploid chromosome numbers, respectively. The 14 AETK's chromosomes (Chr1, 3, 4, 6, 7, 9, 16-20, 22, 23, 24) undergone 10-fusion and 10-fission to form the 14 tarpon's chromosomes (Chr1, 4-8, 10, 11, 18, 19, 21-23, 25) (**Supplementary Table 13**). The remaining 11 AETK's chromosomes (Chr2, 5, 8, 11, 25, 21, 15, 10, 12, 13, 14) correspond to those in tarpons (Chr2, 3, 9, 12-17, 20, 24). This chromosome rearrangement resulted in the same haploid chromosome number ($n = 25$) in tarpons. Comparatively, the 19 AETK's chromosomes (Chr1, 2, 4, 7, 9-14, 16-21, 23-25) underwent 24-fusion and 18-fission to form the 13

chromosomes (Chr1-8, 11, 13, 15-17) in Japanese eels (**Supplementary Table 14**). The remaining 6 AETK's chromosomes (Chr3, 5, 6, 8, 15, 22) correspond to the 6 chromosomes (Chr9, 10, 12, 14, 18 & 19) in Japanese eels. This chromosome rearrangement resulted in the reduction of the chromosome number ($n = 19$) in Japanese eels. Of which, Chr1, Chr3-7 rearrangements are unique to Japanese eels, and might play a role in speciation. Japanese eel's chromosomes (Chr2, 11, 15) were essentially derived from AETK's (Chr1, 3, 21) with slight rearrangements. The patterns of chromosome rearrangements in Chr8, Chr13, and Chr16-17 of Japanese eels were comparable with Chr21, Chr6, Chr8, and Chr19 in tarpons. Without rearrangement, Japanese eel's chromosomes 10, 14 & 19 were equivalent to AETK's chromosomes 3, 6, 22. In addition, there were 3 chromosomes in the Japanese eel (Chr9, Chr12, & Chr18) derived directly from AETK's chromosomes (Chr5, Chr8, & Chr15), those also corresponding to tarpon's chromosomes (Chr3, Chr9, & Chr15), respectively. **Figure 5B** and **Supplementary Fig 10** show the alignment of Japanese eel's chromosomes to tarpon's and arowana's chromosomes and highlighted distinct conservation of orthologous segments.

DISCUSSION

In the past 10 years, the high-resolution whole-genome sequences of the teleosts, flatfish, zebrafish⁴⁰, flatfish¹⁶, killifish⁹⁰, salmon⁵⁹, and the non-teleost ray-finned fishes, including spotted gar¹¹, starlet sturgeon²², the early ray-finned fishes (i.e., bichir, paddlefish, bowfin & alligator gar)⁷ were published. However, as the basal extant group of teleosts, a high-resolution genome assemble of Atlantic or Pacific *Anguilla* species was not achieved. Here, we report the high-quality chromosomal-level Japanese eel's genome for understanding the evolution of this basal extant group and providing the genome database for identifying adaptive and disease-resistant alleles.

The phylogenetic analysis of olfactory receptor (OR) genes identified from the genome sequences of medaka, Atlantic cod, zebrafish, gar, coelacanth and Australian ghost shark indicated that the delta (δ) and zeta (ζ) group genes in the freshwater eels expanded enormously, comprising about 86% of the entire gene family. Delta (δ) and ζ belong to the type I genes⁷, which are specialized for detecting water-soluble odorants and are uniquely expressed in the water-filled lateral diverticulum of the nasal cavity^{25,30}. The mammalian type I (alpha group, α) and (gamma group, γ) genes detect airborne odor molecules. In teleost fishes, the group α genes are absent⁷. Interestingly, the group γ genes were found to have 26 in European eels and 7 in Japanese eels. Since eels can briefly live on land, they may have retained the group γ genes. The group β genes, which detect airborne and water-soluble odor molecules were low in the freshwater eels, but high in numbers in Arowana (35) and spotted gar (20). The group eta (η) genes (type 2) is the third major OR gene group in the freshwater eels. The group η genes are mainly expressed in fishes and is absent in mammals⁷⁰.

The voltage-gated Ca^{2+} channels were found to be the major expanded gene families in Japanese eels. Genome studies suggest that the cellular functions of voltage-gated ion channels emerged early in Metazoan evolution^{64,82}, in determining physiology and behavior at the time of early divergence. It is probably associated with the physiological challenge of Japanese eels to maintain a narrow range of intrinsic Ca^{2+} , during migration

between waters with great variations of calcium contents. A gene expression study in marbled eel (*Anguilla marmorata*) showed the high expression of voltage-gated Ca^{2+} -channels in brain, skin and osmoregulatory tissues (i.e., gills, intestine, and kidneys), and its response to change in water calcium levels¹². Besides controlling Ca^{2+} homeostasis, Ca^{2+} -signaling coordinates a wide range of physiological processes, including skeletal muscle contractions, nervous system activity, cardiac and reproductive functions. The expanded gene families of myosin thick filament in skeletal muscle imply enhanced coordination of muscle contraction and performance⁸¹, especially for this distinct clade of elongated bodies inhabiting a diverse range of habitats⁷⁴. Additionally, the expanded gene families in lipoprotein receptor-related protein, and the isocitrate and isopropylmalate dehydrogenases unravel the importance of this fundamental metabolic and morphogenetic functions in this lineage. Interestingly, lipoprotein receptor-related proteins first appeared during an evolutionary burst associated with the first multicellular organisms, and are multifunctional receptors in nervous system to modulate signals in brains^{21,38}. Isocitrate dehydrogenase is an important enzyme of carbohydrate metabolism, while isopropylmalate dehydrogenase is involved in leucine biosynthesis. Although Japanese eels underwent 3R-WGD, an additional TD and PD duplication was detected. These duplication events, genetic raw materials were provided to facilitate new adaptations to the changing environment⁶⁵. The duplicated gene functions of eels are immune-related to respond against different pathogens, which has been linked to the major factor contributing to the decline of eel populations^{4,19,53}. Presumably, physiological fitness for adaptation might have been weakened by changes in the ecological environment, causing these evolutionary novelties⁵. Notably, the positive selection of immune-related genes indicates the adaptive advantages from the additional TD and PD duplication. Intriguingly, duplicated immune genes was also observed in salmon⁵⁵ and sturgeon²².

The acquisition of evolutionary novelty by WGD duplication and the subsequent fate change of duplicated genes is necessary for phenotype alteration, environmental adaptation, and speciation⁶⁵. The large scale of genomic reshaping after the third rounds

of WGD affect evolutionary complexity and novelty in teleost fishes^{29,43}. It has been widely established that chromosomal numbers are the most fundamental genomic characteristic of an organism or a lineage⁶². On the basis of the hypothesis that genome duplication resulted in chromosomal rearrangements⁴⁵, an understanding of chromosome rearrangement in the eel genome may provide insight into the evolution of karyotype numbers at the base of the teleost evolutionary tree. The majority of fishes today have between 40 and 60 chromosomes (diploid number), while some commonly ancestral fishes are thought to have 48 chromosomes. Chromosome rearrangement and duplication have been the principal mechanisms involved in fish evolution, including the generation of new species and development of sex chromosomes. It is noted that freshwater fishes generally have higher number of chromosomes (the modal diploid number = 54) than marine fishes (the modal diploid number = 48). It has been suggested that the higher number of chromosomes in freshwater fishes is related to a less stable freshwater environment with greater topographical barriers⁷¹. A large capacity for dispersal in the marine environment, on the other hand, would contribute to homogenization of populations, reducing karyotype diversity². Retrospectively, freshwater species seem to speciate more frequently than marine ones⁹. In a study of reconstructing the vertebrate ancestral genome to reveal dynamic genome reorganization, the 3R-WGD in the teleosts ancestor resulted in the number of chromosomes reaching haploid number (n) 26⁶⁶. Evolutionarily, chromosome numbers peak at $n=24$ or 25 in extant teleost species. In this study, we were able to reconstruct the ancestral proto-chromosomes AETK ($n = 25$) to describe the cross-species chromosome collinearity and underpin the lineage-specific genome reorganization. The *Anguilla* species ($n=19$) diverged from *Megalops cyprinoides* ($n=25$) at 196.1 MYA, and their common ancestor from *Scleropages formosus* ($n=25$) at 240.9 MYA. The Anguilliformes are made up of 15 families with remarkable karyotypic diversity⁹¹. The haploid number ranges from 18 to 25, with a prevalence of $n = 19$ and 21. The *Anguilla* lineage underwent a significant structural rearrangement upon their divergence from tarpons (*Megalops cyprinoides*). The fusion and fission of their chromosome structure was the major drivers to reduce the haploid chromosome number to 19.

METHODS

Genome Sequencing. A market-purchased female Japanese eel, *Anguilla japonica* was kept in a freshwater tank for a week with aeration. Blood and muscle samples were taken from the fish, snapped frozen in liquid nitrogen, and then stored at -80°C. Genomic DNA was extracted from the blood samples. DNA sequencing data were generated by different platforms, including Oxford Nanopore (ONT) long reads, PacBio continuous long reads (CLR), Illumina short reads, Illumina mate-pair reads, 10X Chromium linked-reads, DNase Hi-C (Omni-C), and Bionano optical mapping.

For ONT long reads sequencing, the library was prepared by Ligation Sequencing Kit and sequenced using Nanopore PromethION P48 sequencer. For PacBio CLR sequencing, the SMRTbell templates were prepared using Sequel Binding Kit 1.0 and sequenced on the PacBio Sequel System. For Illumina short reads and mate-pair sequencing, the libraries were prepared using TruSeq DNA PCRFree Kit and Nextera Mate Pair Library Preparation Kit (gel plus), respectively. They were sequenced with 2×150bp reads on an Illumina HiSeq X Ten instrument. The library for linked-reads was prepared by 10X Genomics Chromium system with Chromium Genome library v2, and sequenced with 2×150bp reads on an Illumina NovaSeq 6000 instrument. Dovetail Omni-C Kit was used for Hi-C library preparation, which used NEBNext Ultra enzyme and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads prior to PCR enrichment. The library was sequenced with 2×150bp reads on an Illumina HiSeqX platform. The Bionano optical mapping was generated by three enzymes, two from Irys (Nt.BspQI and Nb.BssSI) and one from Saphyr (DLE1). We stretched and captured the images of fluorescently labeled DNA molecules in Irys and Saphyr G1.2 chips. The labeling distances were extracted from the images and recorded into the raw molecule files. Molecules over 150 kbp were assembled into consensus maps using Bionano Solve for further analysis (**Supplementary Table 1**).

Genome Assembly on ONT Long Reads. MitoZ software⁶³ was used to assemble and annotate the mitochondrial genome of Japanese eel. We assembled ONT long reads using

Canu⁵⁷, Wtdbg2⁸⁰, and Flye⁵⁶ separately and merged their contigs using Quickmerge¹³ to achieve a balance between contig N50 and percentage of complete genes (PCGs) in vertebrate species. We used Racon⁹² for two rounds and Medaka (<https://github.com/nanoporetech/medaka>) for one round to self-correct assembly errors using ONT reads, respectively. The PacBio CLR were then incorporated for error correction using Racon for two rounds. As the last step, we further improved the assembly by integrating Illumina short-reads and mate-pair libraries using Pilon⁹³ for two rounds.

Scaffolding on 10x linked-reads, Bionano and Hi-C. We applied Tigrint⁴⁴ and ARKS¹⁸ to correct misassembled contigs and linking contigs into scaffolds according to the shared barcodes from 10x linked-reads. We used OMGS⁷² to integrate three enzymes used in Bionano optical mapping for scaffolding. We further extended the scaffolds using 3D-DNA²³ based on the Hi-C data from Dovetail Omni-C library and refined the scaffolds manually by JuiceBox²⁴ to extend the scaffolds to the corresponding chromosome scale.

Tandem Repeats and Transposable Elements Annotation. Tandem Repeats Finder v4.09⁶ was applied to annotate tandem repetitive sequences. We utilized homolog-based and *de novo* approaches to annotate transposable elements (TEs) in Japanese eel genome. For homolog-based approach, RepeatMasker v4.0.7⁸⁸ and RepeatProteinMask v4.0.7 (<http://www.repeatmasker.org/cgi-bin/RepeatProteinMaskRequest>) were used to identify the repeats by aligning the known TE sequences from RepBase v21.12 database⁵⁰ to the genome. LTR_FINDER v1.07⁹⁶ was used to infer long terminal repeat retrotransposons. For *de novo* approach, RepeatModeler (<http://www.repeatmasker.org/RepeatModeler>) was used to detect the TE families and repeat boundaries by integrating three complementary *de novo* repeat finding programs. RepeatMasker collected the union of these tools' results and annotated the genome accordingly.

Genes and their Functional Annotation. Three types of methods were used to annotate the protein-coding genes in the genome, including *de novo*, homology-based, and transcriptome-based annotations. Maker (v2.31.8)³⁹ was adopted for homology

annotation using the protein sequences from the five closely related species, including European eel (*Anguilla Anguilla*), zebrafish (*Danio rerio*), Indo-Pacific tarpons (*Megalops cyprinoides*), Asian arowana (*Scleropages formosus*), and spotted gar (*Lepisosteus oculatus*), based on the phylogeny of teleost fishes⁸.

De novo annotation was performed using Augustus⁸⁵ and SNAP⁴⁹ by training a model using 3,000 complete genes obtained from homology prediction. Transcriptome annotation was performed by aligning RNA-seq data to the genome with HISAT2 (version 2.1.0)⁵⁴, and assembling transcript sequences with Trinity³⁴. Pasa_lite (https://github.com/PASApipeline/PASA_Lite) was used to correct assembly errors to obtain the final transcripts. Maker (v2.31.8) was further used to integrate the three annotations, followed by the second round of homology annotation to refine the final gene set.

Gene functional annotation was performed by aligning the predicted gene sequences to protein sequences using BLAST v2.2.31¹ in the six databases, including NCBI Non-Redundant Protein Sequence (NR), Kyoto Encyclopedia of Genes and Genomes⁵¹, SwissProt¹⁰, KOG⁸⁹, Gene Ontology³, and TrEMBL (Uniprot version 2020-06). We further searched the secondary structure domain database for gene function prediction using InterProscan⁹⁸.

Evaluation of Genome Assembly and Gene annotation. BUSCO⁸³ was used to evaluate genome assembly and gene annotation by calculating the completeness of single-copy orthologs. We selected the Ray-finned Fish single-copy orthologs direct homologous gene database actinopterygii_odb10 (which contains 3640 core single-copy direct homologous gene proteins), the closest relative to Japanese eel in the OrthoDB database (<https://www.orthodb.org/>) to compare.

Annotation of Conserved Noncoding Elements. tRNAscan-SE 1.3.1⁶¹ was used to identify tRNA sequences in the genome families. We annotated the rRNA sequences by aligning the conserved rRNA sequences from the five closely related fish species (European eel, zebrafish, tarpons, arowana, and spotted gar) to the genome using

BLASTN⁹⁹. The microRNAs and snRNAs were annotated by aligning the corresponding sequences from Rfam³¹ to the genome.

Phylogenetic Analysis, Gene Expansion, and Gene Contraction. OrthoMCL 2.0⁵⁸ (<http://orthomcl.org/orthomcl/>) was used to identify gene families by grouping orthologous proteins. We applied maximum likelihood method³³ and RAXML⁸⁴ (<http://sco.h-its.org/exelixis/web/software/raxml/index.html>) to reconstruct the phylogenetic tree using four-fold degenerate sites (4DTv) in single-copy orthologs from the 12 fish species, including *Anguilla rostrata* (American eel, GenBank assembly: GCA_001606085.1), *Anguilla anguilla* (European eel, GCA_013347855.1), *Anguilla japonica* (Japanese eel), *Megalops cyprinoides* (tarpons, GCA_013368585.1), *Scleropages formosus* (arowana, GCA_900964775.1), *Gadus morhua* (Atlantic cod, GCA_902167405.1), *Oryzias latipes* (medaka, GCA_002234675.1), *Danio rerio* (zebrafish, GCA_000002035.4), *Lepisosteus oculatus* (spotted gar, GCA_000242695.1), *Erpetoichthys calabaricus* (reed fish, GCA_900747795.2), *Latimeria chalumnae* (coelacanth, GCA_000225785.1) and *Callorhynchus milii* (Australian ghost shark, GCA_000165045.2). We estimated the divergence times for single-copy orthologs using mcmctree (<http://abacus.gene.ucl.ac.uk/software/paml.html>) in PAML package⁹⁷ based on the predefined times from TimeTree website (<http://www.timetree.org/>) [*Danio rerio* with *Oryzias latipes* (214.9 - 253.9 Mya), *Megalops cyprinoides* with *Anguilla anguilla* (179.3 - 219.3 Mya), *Callorhynchus milii* with *Danio rerio* (452.6 - 496.5 Mya) and *Erpetoichthys calabaricus* with *Danio rerio* (367.9 - 405.5 Mya)]. To estimate gene family expansion and contraction, we used CAFÉ²⁰ (<http://sourceforge.net/projects/cafehahnlab/>) to model gene expansions and contractions, as well as the divergence times.

Identification of olfactory receptor (OR) genes. We identified OR genes using the pipeline described in Github (https://github.com/MaximePolicarpo/Olfactory_receptor_genes)⁷⁶, while candidate genes were filtered via the NR database. The OR gene identified in a previous study⁶⁹ was used as a query sequence. TBLASTN²⁷ was used to identify genomic regions containing OR genes in the 10 fish species (European eel,

Japanese eel, tarpons, arowana, medaka, Atlantic cod, zebrafish, spotted gar, coelacanth, and Australian ghost shark). Only the non-overlapping BLAST hits regions were extracted, and the 1kb upstream and downstream flanking regions were used as the input to EMBOSS⁷⁸. EMBOSS was used to generate the Open Reading Frames (ORFs), translated the ORFs into protein sequences, and then ran BlastP to weed out sequences that did not match genes already known in SwissProt and NR. InterProscan was used to determine the secondary structures of the predicted OR genes. Some genes were filtered due to lacking the seven transmembrane domains. The maximum likelihood phylogenetic tree was reconstructed using IQ-TREE⁶⁸ based on the multiple sequencing alignments on the OR gene sequences with MAFFT⁵².

Genome evolution analysis. MCscanX⁹⁵ and macrosynteny visualization (jcvi) were used to screen for collinear blocks with at least 30 genes⁸⁷ in *Anguilla japonica*, *Anguilla anguilla*, *Anguilla rostrata*, *Megalops cyprinoides*, and *Lepisosteus oculatus*. The numbers of non-synonymous substitutions (Ka) and synonymous substitutions (Ks) were calculated using KaKs_calculator2.0⁹⁴. In addition, we calculated 4dTv values to estimate the WGD events in Japanese eel genome. We identified gene duplicates in the genomes of Japanese eel, zebrafish, arowana, medaka and Atlantic cod using the DupGen_finder pipeline⁷⁷, using spotted gars as an outgroup. It classified gene duplication patterns into five categories: whole genome duplications, tandem duplications, proximal duplications (non-tandem duplications that are separated by 10 genes on the same chromosome), transposable duplications, and scattered duplications (duplications other than the four categories mentioned above).

Ancestral Chromosome Reconfiguration. Ancestral eel/tarpon karyotype (AETK) was constructed using *Anguilla japonica* (Japanese eel), *Megalops cyprinoides* (tarpon) and *Scleropages formosus* (arowana, outgroup). The ancestral teleosts karyotype (ATK) was constructed using zebrafish, *Scleropages formosus* (arowana), and *Lepisosteus oculatus* (spotted gar, outgroup)¹¹. This was done by using BLASTP⁸⁶ to obtain homologous gene pairs between species. The default parameters of MCScanX were then used to obtain the

collinear blocks of chromosomes between species. Finally, the karyotype of the ancestor was constructed using ANGeSv1.01¹⁵.

Acknowledgement

This work was supported by the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC02) to CKCW, AOLW, TFC & KPL, and General Research Fund (Research Grant Council, HKBU12162016) to CKCW.

Data availability

The *Anguilla japonica* whole genome sequencing and assembly data has been deposited at National Library of Medicine, BioProject under the accession number PRJNA852364.

FIGURE LEGENDS

Figure 1. The genome landscape of Japanese eel, *Anguilla japonica*. From outer to inner circle: (A) length of 19 chromosomes (Mb); (B) Read-depth of ONT long-reads; (C) Read-depth of Illumina short-reads; (D) Distribution of transposon sequences; (E) Distribution of protein-coding gene; (F) GC content; (G) Collinear blocks of at least 10 genes in genome. The window size is 1MB.

Figure 2. Phylogenetic relationship, divergence times and gene families of *Anguilla* species, relevant bony and cartilaginous fishes. Expansions (numbers in green) and contractions (numbers in purple) of gene families are shown at individual lineages. Each node shows the estimated divergence times (blue numbers, millions of years ago, Mya) and the 95% confidence intervals for these dates. Red dots indicate times taken from the TimeTree website (<http://www.timetree.org/>). The orange star shows the 3R-WGD event. Geological periods from left to right: S= Silurian, D= Devonian, C= Carboniferous, P= Permian, T= Triassic, J= Jurassic, K= Cretaceous, Pa=: Paleogene, N= Neogene. A comparison of gene families associated with orthologs and paralogs in Japanese eel and the 11 fish species.

Figure 3. Number and classification of olfactory receptor (OR) genes for 10 fish species. On the left is the phylogenetic tree of the 10 species. The number of OR genes is shown on the right. The size of circle indicate the number of OR genes.

Figure 4. (A) Four-fold synonymous third-codon transversion rate (4dTv) distributions of homologous gene pairs for intra-species (paralogs density) and inter-species (orthologs density) comparisons. (B) The collinear relationships of syntenic blocks between *Anguilla japonica* and *Lepisosteus oculatus*. The numbers indicate the corresponding chromosomes for each species. In *Lepisosteus oculatus*, the 29th chromosome is 293.7 Kb long, which has no collinearity with that of *Anguilla japonica*. Based on homologous blocks of at least 10 genes, gene links between these two species were identified. The four collinear blocks that contain Hox genes are shown in green, yellow, red, and blue. (C) Ks distributions of syntenic gene pairs from different gene duplications (wgd: whole genome duplication, trd: transposable duplication, td: tandem duplication, pd: proximal duplication, dsd: dispersed duplication). The y-axis shows the distribution of Ks values. (D) Enrichment analysis of five duplicated expansion gene families, with the color of the circles representing the statistical significance of the GO. The circle size represents the number of genes.

Figure 5. Reconstruction of proto-chromosomes for the common ancestor of teleosts (ATK) and eel/tarpons (AETK). (A) A model for the distribution of chromosomal segments in the genomes of ATK, arowana, AETK, Japanese eels and tarpons. AETK is the common ancestor of tarpons and eels. The Circos plots indicate conservation of synteny between (B) Japanese eel and tarpon, as well as (C) arowana and Japanese eel.

References

- 1 S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, "Basic local alignment search tool," *J Mol. Biol* **215**(3), 403 (1990).
- 2 R. F. Artoni, J. P. Castro, U. P. Jacobina, P. A. Lima-Filho, G. W. da Costa, W. F. Molina, "Inferring Diversity and Evolution in Fish by Means of Integrative Molecular Cytogenetics," *ScientificWorldJournal*. **2015**, 365787 (2015).
- 3 M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.* **25**(1), 25 (2000).
- 4 I. Bandin, S. Souto, J. M. Cutrin, C. Lopez-Vazquez, J. G. Oliveira, C. Esteve, E. Alcaide, C. P. Dopazo, "Presence of viruses in wild eels *Anguilla anguilla* L, from the Albufera Lake (Spain)," *J Fish Dis.* **37**(7), 597 (2014).
- 5 A. Belyayev, "Bursts of transposable elements as an evolutionary driving force," *J Evol. Biol* **27**(12), 2573 (2014).
- 6 G. Benson, "Tandem repeats finder: a program to analyze DNA sequences," *Nucleic Acids Res.* **27**(2), 573 (1999).
- 7 X. Bi, K. Wang, L. Yang, H. Pan, H. Jiang, Q. Wei, M. Fang, H. Yu, C. Zhu, Y. Cai, Y. He, X. Gan, H. Zeng, D. Yu, Y. Zhu, H. Jiang, Q. Qiu, H. Yang, Y. E. Zhang, W. Wang, M. Zhu, S. He, G. Zhang, "Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes," *Cell* **184**(5), 1377 (2021).
- 8 C. Bian, Y. Hu, V. Ravi, I. S. Kuznetsova, X. Shen, X. Mu, Y. Sun, X. You, J. Li, X. Li, Y. Qiu, B. H. Tay, N. M. Thevasagayam, A. S. Komissarov, V. Trifonov, M. Kabilov, A. Tupikin, J. Luo, Y. Liu, H. Song, C. Liu, X. Wang, D. Gu, Y. Yang, W. Li, G. Polgar, G. Fan, P. Zeng, H. Zhang, Z. Xiong, Z. Tang, C. Peng, Z. Ruan, H. Yu, J. Chen, M. Fan, Y. Huang, M. Wang, X. Zhao, G. Hu, H. Yang, J. Wang, J. Wang, X. Xu, L. Song, G. Xu, P. Xu, J. Xu, S. J. O'Brien, L. Orban, B. Venkatesh, Q. Shi, "The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts," *Sci. Rep* **6**, 24501 (2016).
- 9 D. D. Bloom, J. T. Weir, K. R. Piller, N. R. Lovejoy, "Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of new world silversides (atherinopsidae)," *Evolution* **67**(7), 2040 (2013).

- 10 B. Boeckmann, A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003," *Nucleic Acids Res.* **31**(1), 365 (2003).
- 11 I. Braasch, A. R. Gehrke, J. J. Smith, K. Kawasaki, T. Manousaki, J. Pasquier, A. Amores, T. Desvignes, P. Batzel, J. Catchen, A. M. Berlin, M. S. Campbell, D. Barrell, K. J. Martin, J. F. Mulley, V. Ravi, A. P. Lee, T. Nakamura, D. Chalopin, S. Fan, D. Weisel, C. Canestro, J. Sydes, F. E. Beaudry, Y. Sun, J. Hertel, M. J. Beam, M. Fasold, M. Ishiyama, J. Johnson, S. Kehr, M. Lara, J. H. Letaw, G. W. Litman, R. T. Litman, M. Mikami, T. Ota, N. R. Saha, L. Williams, P. F. Stadler, H. Wang, J. S. Taylor, Q. Fontenot, A. Ferrara, S. M. Searle, B. Aken, M. Yandell, I. Schneider, J. A. Yoder, J. N. Volff, A. Meyer, C. T. Amemiya, B. Venkatesh, P. W. Holland, Y. Guiguen, J. Bobe, N. H. Shubin, Palma F. Di, J. Alföldi, K. Lindblad-Toh, J. H. Postlethwait, "The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons," *Nat. Genet.* **48**(4), 427 (2016).
- 12 Q. Cao, P. Chu, J. Gu, H. Zhang, R. Feng, X. Wen, D. Wang, W. Xiong, T. Wang, S. Yin, "The influence of Ca(2+) concentration on voltage-dependent L-type calcium channels' expression in the marbled eel (*Anguilla marmorata*)," *Gene* **722**, 144101 (2020).
- 13 M. Chakraborty, J. G. Baldwin-Brown, A. D. Long, J. J. Emerson, "Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage," *Nucleic Acids Res.* **44**(19), e147 (2016).
- 14 Y. K. Chang, Y. Miyazawa, M. J. Miller, K. Tsukamoto, "Potential impact of ocean circulation on the declining Japanese eel catches," *Sci. Rep.* **8**(1), 5496 (2018).
- 15 C. Chauve, E. Tannier, "A methodological framework for the reconstruction of contiguous regions of ancestral genomes and its application to mammalian genomes," *PLoS. Comput. Biol.* **4**(11), e1000234 (2008).
- 16 S. Chen, G. Zhang, C. Shao, Q. Huang, G. Liu, P. Zhang, W. Song, N. An, D. Chalopin, J. N. Volff, Y. Hong, Q. Li, Z. Sha, H. Zhou, M. Xie, Q. Yu, Y. Liu, H. Xiang, N. Wang, K. Wu, C. Yang, Q. Zhou, X. Liao, L. Yang, Q. Hu, J. Zhang, L. Meng, L. Jin, Y. Tian, J. Lian, J. Yang, G. Miao, S. Liu, Z. Liang, F. Yan, Y. Li, B. Sun, H. Zhang, J. Zhang, Y. Zhu, M. Du, Y. Zhao, M. Scharl, Q. Tang, J. Wang, "Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle," *Nat. Genet.* **46**(3), 253 (2014).
- 17 W. Chen, C. Bian, X. You, J. Li, L. Ye, Z. Wen, Y. Lv, X. Zhang, J. Xu, S. Yang, R. Gu, X. Lin, Q. Shi, "Genome Sequencing of the Japanese Eel (*Anguilla japonica*) for Comparative Genomic Studies on *tbx4* and a *tbx4* Gene Cluster in Teleost Fishes," *Mar Drugs* **17**(7) (2019).

- 18 L. Coombe, J. Zhang, B. P. Vandervalk, J. Chu, S. D. Jackman, I. Birol, R. L. Warren, "ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers," *BMC. Bioinformatics*. **19**(1), 234 (2018).
- 19 L. Danne, L. Horn, A. Feldhaus, D. Fey, S. Emde, H. Schutze, M. Adamek, J. Hellmann, "Virus infections of the European Eel in North Rhine Westphalian rivers," *J Fish Dis*. **45**(1), 69 (2022).
- 20 Bie T. De, N. Cristianini, J. P. Demuth, M. W. Hahn, "CAFE: a computational tool for the study of gene family evolution," *Bioinformatics*. **22**(10), 1269 (2006).
- 21 M. Dieckmann, M. F. Dietrich, J. Herz, "Lipoprotein receptors--an evolutionarily ancient multifunctional receptor family," *Biol Chem*. **391**(11), 1341 (2010).
- 22 K. Du, M. Stock, S. Kneitz, C. Klopp, J. M. Woltering, M. C. Adolphi, R. Feron, D. Prokopov, A. Makunin, I. Kichigin, C. Schmidt, P. Fischer, H. Kuhl, S. Wuertz, J. Gessner, W. Kloas, C. Cabau, C. Iampietro, H. Parrinello, C. Tomlinson, L. Journot, J. H. Postlethwait, I. Braasch, V. Trifonov, W. C. Warren, A. Meyer, Y. Guiguen, M. Scharl, "The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization," *Nat. Ecol Evol*. **4**(6), 841 (2020).
- 23 O. Dudchenko, S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, A. P. Aiden, E. L. Aiden, "De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds," *Science* **356**(6333), 92 (2017).
- 24 N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, E. L. Aiden, "Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom," *Cell Syst*. **3**(1), 99 (2016).
- 25 J. Freitag, J. Krieger, J. Strotmann, H. Breer, "Two classes of olfactory receptors in *Xenopus laevis*," *Neuron* **15**(6), 1383 (1995).
- 26 C. Geeraerts, C. Belpaire, "The effects of contaminants in European eel: a review," *Ecotoxicology*. **19**(2), 239 (2010).
- 27 E. M. Gertz, Y. K. Yu, R. Agarwala, A. A. Schaffer, S. F. Altschul, "Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST," *BMC. Biol* **4**, 41 (2006).
- 28 J. Ghurye, M. Pop, "Modern technologies and algorithms for scaffolding assembled genomes," *PLoS. Comput. Biol* **15**(6), e1006994 (2019).
- 29 S. M. Glasauer, S. C. Neuhauss, "Whole-genome duplication in teleost fishes and its evolutionary consequences," *Mol. Genet. Genomics* **289**(6), 1045 (2014).

- 30 G. Glusman, A. Bahar, D. Sharon, Y. Pilpel, J. White, D. Lancet, "The olfactory receptor gene superfamily: data mining, classification, and nomenclature," *Mamm. Genome* **11**(11), 1016 (2000).
- 31 S. Griffiths-Jones, S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, A. Bateman, "Rfam: annotating non-coding RNAs in complete genomes," *Nucleic Acids Res.* **33**(Database issue), D121-D124 (2005).
- 32 M. R. Gross, R. M. Coleman, R. M. McDowall, "Aquatic productivity and the evolution of diadromous fish migration," *Science* **239**(4845), 1291 (1988).
- 33 S. Guindon, O. Gascuel, "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood," *Syst. Biol* **52**(5), 696 (2003).
- 34 B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, A. Regev, "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis," *Nat. Protoc.* **8**(8), 1494 (2013).
- 35 J. L. Hein, S. A. Arnott, W. A. Roumillat, D. M. Allen, Buron de, I, "Invasive swimbladder parasite *Anguillicoloides crassus*: infection status 15 years after discovery in wild populations of American eel *Anguilla rostrata*," *Dis. Aquat. Organ* **107**(3), 199 (2014).
- 36 C. V. Henkel, E. Burgerhout, D. L. de Wijze, R. P. Dirks, Y. Minegishi, H. J. Jansen, H. P. Spaik, S. Dufour, F. A. Weltzien, K. Tsukamoto, G. E. van den Thillart, "Primitive duplicate Hox clusters in the European eel's genome," *PLoS. ONE.* **7**(2), e32231 (2012).
- 37 C. V. Henkel, R. P. Dirks, D. L. de Wijze, Y. Minegishi, J. Aoyama, H. J. Jansen, B. Turner, B. Knudsen, M. Bundgaard, K. L. Hvam, M. Boetzer, W. Pirovano, F. A. Weltzien, S. Dufour, K. Tsukamoto, H. P. Spaik, G. E. van den Thillart, "First draft genome sequence of the Japanese eel, *Anguilla japonica*," *Gene* **511**(2), 195 (2012).
- 38 J. Herz, H. H. Bock, "Lipoprotein receptors in the nervous system," *Annu Rev Biochem.* **71**, 405 (2002).
- 39 C. Holt, M. Yandell, "MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects," *BMC. Bioinformatics.* **12**, 491 (2011).
- 40 K. Howe, M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M.

Caccamo, C. Churcher, C. Scott, J. C. Barrett, R. Koch, G. J. Rauch, S. White, W. Chow, B. Kilian, L. T. Quintais, J. A. Guerra-Assuncao, Y. Zhou, Y. Gu, J. Yen, J. H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd, E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Elliot, G. Threadgold, G. Harden, D. Ware, S. Begum, B. Mortimore, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Lloyd, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthavadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C. M. Dooley, Z. Ersan-Urun, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kim, J. Konantz, M. Konantz, M. Oberlander, S. Rudolph-Geiger, M. Teucke, C. Lanz, G. Raddatz, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, S. C. Schuster, N. P. Carter, J. Harrow, Z. Ning, J. Herrero, S. M. Searle, A. Enright, R. Geisler, R. H. Plasterk, C. Lee, M. Westerfield, P. J. de Jong, L. I. Zon, J. H. Postlethwait, C. Nusslein-Volhard, T. J. Hubbard, Crollius H. Roest, J. Rogers, D. L. Stemple, "The zebrafish reference genome sequence and its relationship to the human genome," *Nature* **496**(7446), 498 (2013).

- 41 L. C. Hughes, G. Orti, Y. Huang, Y. Sun, C. C. Baldwin, A. W. Thompson, D. Arcila, R. Betancur, C. Li, L. Becker, N. Bellora, X. Zhao, X. Li, M. Wang, C. Fang, B. Xie, Z. Zhou, H. Huang, S. Chen, B. Venkatesh, Q. Shi, "Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data," *Proc. Natl. Acad. Sci. U. S. A* **115**(24), 6249 (2018).
- 42 I. A. Hurley, R. L. Mueller, K. A. Dunn, E. J. Schmidt, M. Friedman, R. K. Ho, V. E. Prince, Z. Yang, M. G. Thomas, M. I. Coates, "A new time-scale for ray-finned fish evolution," *Proc. Biol. Sci.* **274**(1609), 489 (2007).
- 43 J. Inoue, Y. Sato, R. Sinclair, K. Tsukamoto, M. Nishida, "Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling," *Proc. Natl. Acad. Sci. U. S. A* **112**(48), 14918 (2015).
- 44 S. D. Jackman, L. Coombe, J. Chu, R. L. Warren, B. P. Vandervalk, S. Yeo, Z. Xue, H. Mohamadi, J. Bohlmann, S. J. M. Jones, I. Birol, "Tigmint: correcting assembly errors using linked reads from large molecules," *BMC. Bioinformatics.* **19**(1), 393 (2018).

- 45 O. Jaillon, J. M. Aury, F. Brunet, J. L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot, S. Nicaud, D. Jaffe, S. Fisher, G. Lutfalla, C. Dossat, B. Segurens, C. Dasilva, M. Salanoubat, M. Levy, N. Boudet, S. Castellano, V. Anthouard, C. Jubin, V. Castelli, M. Katinka, B. Vacherie, C. Biemont, Z. Skalli, L. Cattolico, J. Poulain, Berardinis De, V, C. Cruaud, S. Duprat, P. Brottier, J. P. Coutanceau, J. Gouzy, G. Parra, G. Lardier, C. Chapple, K. J. McKernan, P. McEwan, S. Bosak, M. Kellis, J. N. Volff, R. Guigo, M. C. Zody, J. Mesirov, K. Lindblad-Toh, B. Birren, C. Nusbaum, D. Kahn, M. Robinson-Rechavi, V. Laudet, V. Schachter, F. Quetier, W. Saurin, C. Scarpelli, P. Wincker, E. S. Lander, J. Weissenbach, Crollius H. Roest, "Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype," *Nature* **431**(7011), 946 (2004).
- 46 H. J. Jansen, M. Liem, S. A. Jong-Raadsen, S. Dufour, F. A. Weltzien, W. Swinkels, A. Koelewijn, A. P. Palstra, B. Pelster, H. P. Spaink, G. E. V. D. Thillart, R. P. Dirks, C. V. Henkel, "Rapid de novo assembly of the European eel genome from nanopore sequencing reads," *Sci. Rep* **7**(1), 7213 (2017).
- 47 P. Jehannet, A. P. Palstra, L. T. N. Heinsbroek, L. Kruijt, R. P. Dirks, W. Swinkels, H. Komen, "What Goes Wrong during Early Development of Artificially Reproduced European Eel *Anguilla anguilla*? Clues from the Larval Transcriptome and Gene Expression Patterns," *Animals*. (Basel) **11**(6) (2021).
- 48 M. R. Jensen, S. W. Knudsen, P. Munk, P. F. Thomsen, P. R. Moller, "Tracing European eel in the diet of mesopelagic fishes from the Sargasso Sea using DNA from fish stomachs," **165**, 130 (2018).
- 49 A. D. Johnson, R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'Donnell, P. I. de Bakker, "SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap," *Bioinformatics*. **24**(24), 2938 (2008).
- 50 J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, "Repbase Update, a database of eukaryotic repetitive elements," *Cytogenet. Genome Res.* **110**(1-4), 462 (2005).
- 51 M. Kanehisa, S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.* **28**(1), 27 (2000).
- 52 K. Katoh, D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," *Mol. Biol. Evol.* **30**(4), 772 (2013).
- 53 C. R. Kennedy, "The pathogenic helminth parasites of eels," *J Fish Dis.* **30**(6), 319 (2007).
- 54 D. Kim, B. Langmead, S. L. Salzberg, "HISAT: a fast spliced aligner with low memory requirements," *Nat. Methods* **12**(4), 357 (2015).

- 55 E. Kjaerner-Semb, F. Ayllon, T. Furmanek, V. Wennevik, G. Dahle, E. Niemela, M. Ozerov, J. P. Vaha, K. A. Glover, C. J. Rubin, A. Wargelius, R. B. Edvardsen, "Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection," *BMC. Genomics* **17**(1), 610 (2016).
- 56 M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, "Assembly of long, error-prone reads using repeat graphs," *Nat. Biotechnol.* **37**(5), 540 (2019).
- 57 S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, "Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation," *Genome Res.* **27**(5), 722 (2017).
- 58 L. Li, C. J. Stoeckert, Jr., D. S. Roos, "OrthoMCL: identification of ortholog groups for eukaryotic genomes," *Genome Res.* **13**(9), 2178 (2003).
- 59 S. Lien, B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent, T. Nome, T. R. Hvidsten, J. S. Leong, D. R. Minkley, A. Zimin, F. Grammes, H. Grove, A. Gjuvsland, B. Walenz, R. A. Hermansen, Schalburg K. von, E. B. Rondeau, Genova A. Di, J. K. Samy, Vik J. Olav, M. D. Vigeland, L. Caler, U. Grimholt, S. Jentoft, D. I. Vage, Jong P. de, T. Moen, M. Baranski, Y. Palti, D. R. Smith, J. A. Yorke, A. J. Nederbragt, A. Tooming-Klunderud, K. S. Jakobsen, X. Jiang, D. Fan, Y. Hu, D. A. Liberles, R. Vidal, P. Iturra, S. J. Jones, I. Jonassen, A. Maass, S. W. Omholt, W. S. Davidson, "The Atlantic salmon genome provides insights into rediploidization," *Nature* **533**(7602), 200 (2016).
- 60 Y. C. Liu, S. D. Hsu, C. H. Chou, W. Y. Huang, Y. H. Chen, C. Y. Liu, G. J. Lyu, S. Z. Huang, S. Aganezov, M. A. Alekseyev, C. D. Hsiao, H. D. Huang, "Transcriptome sequencing based annotation and homologous evidence based scaffolding of *Anguilla japonica* draft genome," *BMC. Genomics* **17 Suppl 1**, 13 (2016).
- 61 T. M. Lowe, S. R. Eddy, "tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence," *Nucleic Acids Res.* **25**(5), 955 (1997).
- 62 I. Mayrose, M. A. Lysak, "The Evolution of Chromosome Numbers: Mechanistic Models and Experimental Approaches," *Genome Biol Evol.* **13**(2) (2021).
- 63 G. Meng, Y. Li, C. Yang, S. Liu, "MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization," *Nucleic Acids Res.* **47**(11), e63 (2019).
- 64 Y. Moran, H. H. Zakon, "The evolution of the four subunits of voltage-gated calcium channels: ancient roots, increasing complexity, and multiple losses," *Genome Biol Evol.* **6**(9), 2210 (2014).

- 65 Y. Moriyama, K. Koshiba-Takeuchi, "Significance of whole-genome duplications on the emergence of evolutionary novelties," *Brief. Funct. Genomics* **17**(5), 329 (2018).
- 66 Y. Nakatani, H. Takeda, Y. Kohara, S. Morishita, "Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates," *Genome Res.* **17**(9), 1254 (2007).
- 67 T. J. Near, R. I. Eytan, A. Dornburg, K. L. Kuhn, J. A. Moore, M. P. Davis, P. C. Wainwright, M. Friedman, W. L. Smith, "Resolution of ray-finned fish phylogeny and timing of diversification," *Proc. Natl. Acad. Sci. U. S. A* **109**(34), 13698 (2012).
- 68 L. T. Nguyen, H. A. Schmidt, Haeseler A. von, B. Q. Minh, "IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies," *Mol. Biol. Evol.* **32**(1), 268 (2015).
- 69 Y. Niimura, "On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species," *Genome Biol. Evol.* **1**, 34 (2009).
- 70 Y. Niimura, M. Nei, "Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods," *Proc. Natl. Acad. Sci. U. S. A* **102**(17), 6039 (2005).
- 71 G. Nikolsky, "THE INTERRELATION BETWEEN VARIABILITY OF CHARACTERS, EFFECTIVENESS OF ENERGY UTILISATION, AND KARYOTYPE STRUCTURE IN FISHES," *Evolution* **30**(1), 180 (1976).
- 72 W. Pan, T. Jiang, S. Lonardi, "OMGS: Optical Map-Based Genome Scaffolding," *J. Comput. Biol.* **27**(4), 519 (2020).
- 73 S. A. Pavey, M. Laporte, E. Normandeau, J. Gaudin, L. Letourneau, S. Boisvert, J. Corbeil, C. Audet, L. Bernatchez, "Draft genome of the American Eel (*Anguilla rostrata*)," *Mol. Ecol. Resour.* **17**(4), 806 (2017).
- 74 C. Pfaff, R. Zorzin, J. Kriwet, "Evolution of the locomotory system in eels (Teleostei: Elopomorpha)," *BMC. Evol. Biol.* **16**(1), 159 (2016).
- 75 A. T. Piper, C. Manes, F. Siniscalchi, A. Marion, R. M. Wright, P. S. Kemp, "Response of seaward-migrating European eel (*Anguilla anguilla*) to manipulated flow fields," *Proc. Biol. Sci.* **282**(1811) (2015).
- 76 M. Policarpo, K. E. Bemis, J. C. Tyler, C. J. Metcalfe, P. Laurenti, J. C. Sandoz, S. Retaux, D. Casane, "Evolutionary Dynamics of the OR Gene Repertoire in Teleost Fishes: Evidence of an Association with Changes in Olfactory Epithelium Shape," *Mol. Biol. Evol.* **38**(9), 3742 (2021).

- 77 X. Qiao, Q. Li, H. Yin, K. Qi, L. Li, R. Wang, S. Zhang, A. H. Paterson, "Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants," *Genome Biol* **20**(1), 38 (2019).
- 78 P. Rice, I. Longden, A. Bleasby, "EMBOSS: the European Molecular Biology Open Software Suite," *Trends Genet.* **16**(6), 276 (2000).
- 79 C. Rozenfeld, J. Blanca, V. Gallego, V. Garcia-Carpintero, J. G. Herranz-Jusado, L. Perez, J. F. Asturiano, J. Canizares, D. S. Penaranda, "De novo European eel transcriptome provides insights into the evolutionary history of duplicated genes in teleost lineages," *PLoS. ONE.* **14**(6), e0218085 (2019).
- 80 J. Ruan, H. Li, "Fast and accurate long-read assembly with wtdbg2," *Nat. Methods* **17**(2), 155 (2020).
- 81 M. F. Schneider, W. K. Chandler, "Voltage dependent charge movement of skeletal muscle: a possible step in excitation-contraction coupling," *Nature* **242**(5395), 244 (1973).
- 82 A. Senatore, H. Raiss, P. Le, "Physiology and Evolution of Voltage-Gated Calcium Channels in Early Diverging Animal Phyla: Cnidaria, Placozoa, Porifera and Ctenophora," *Front Physiol* **7**, 481 (2016).
- 83 F. A. Simao, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs," *Bioinformatics.* **31**(19), 3210 (2015).
- 84 A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics.* **30**(9), 1312 (2014).
- 85 M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, "AUGUSTUS: ab initio prediction of alternative transcripts," *Nucleic Acids Res.* **34**(Web Server issue), W435-W439 (2006).
- 86 D. J. States, W. Gish, "Combined use of sequence similarity and codon bias for coding region identification," *J Comput. Biol* **1**(1), 39 (1994).
- 87 H. Tang, J. E. Bowers, X. Wang, R. Ming, M. Alam, A. H. Paterson, "Synteny and collinearity in plant genomes," *Science* **320**(5875), 486 (2008).
- 88 M. Tarailo-Graovac, N. Chen, "Using RepeatMasker to identify repetitive elements in genomic sequences," *Curr. Protoc. Bioinformatics.* **Chapter 4**, Unit (2009).
- 89 R. L. Tatusov, N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, D. A. Natale, "The

- COG database: an updated version includes eukaryotes," *BMC. Bioinformatics.* **4**, 41 (2003).
- 90 D. R. Valenzano, B. A. Benayoun, P. P. Singh, E. Zhang, P. D. Etter, C. K. Hu, M. Clement-Ziza, D. Willemsen, R. Cui, I. Harel, B. E. Machado, M. C. Yee, S. C. Sharp, C. D. Bustamante, A. Beyer, E. A. Johnson, A. Brunet, "The African Turquoise Killifish Genome Provides Insights into Evolution and Genetic Architecture of Lifespan," *Cell* **163**(6), 1539 (2015).
 - 91 A. J. Vasconcelos, W. F. Molina, "Cytogenetical studies in five Atlantic Anguilliformes fishes," *Genet. Mol. Biol* **32**(1), 83 (2009).
 - 92 R. Vaser, I. Sovic, N. Nagarajan, M. Sikic, "Fast and accurate de novo genome assembly from long uncorrected reads," *Genome Res.* **27**(5), 737 (2017).
 - 93 B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, "Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement," *PLoS. ONE.* **9**(11), e112963 (2014).
 - 94 D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, "KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies," *Genomics Proteomics. Bioinformatics.* **8**(1), 77 (2010).
 - 95 Y. Wang, H. Tang, J. D. Debarry, X. Tan, J. Li, X. Wang, T. H. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger, A. H. Paterson, "MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity," *Nucleic Acids Res.* **40**(7), e49 (2012).
 - 96 Z. Xu, H. Wang, "LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons," *Nucleic Acids Res.* **35**(Web Server issue), W265-W268 (2007).
 - 97 Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Mol. Biol Evol.* **24**(8), 1586 (2007).
 - 98 E. M. Zdobnov, R. Apweiler, "InterProScan--an integration platform for the signature-recognition methods in InterPro," *Bioinformatics.* **17**(9), 847 (2001).
 - 99 Z. Zhang, S. Schwartz, L. Wagner, W. Miller, "A greedy algorithm for aligning DNA sequences," *J Comput. Biol* **7**(1-2), 203 (2000).

Figure 1

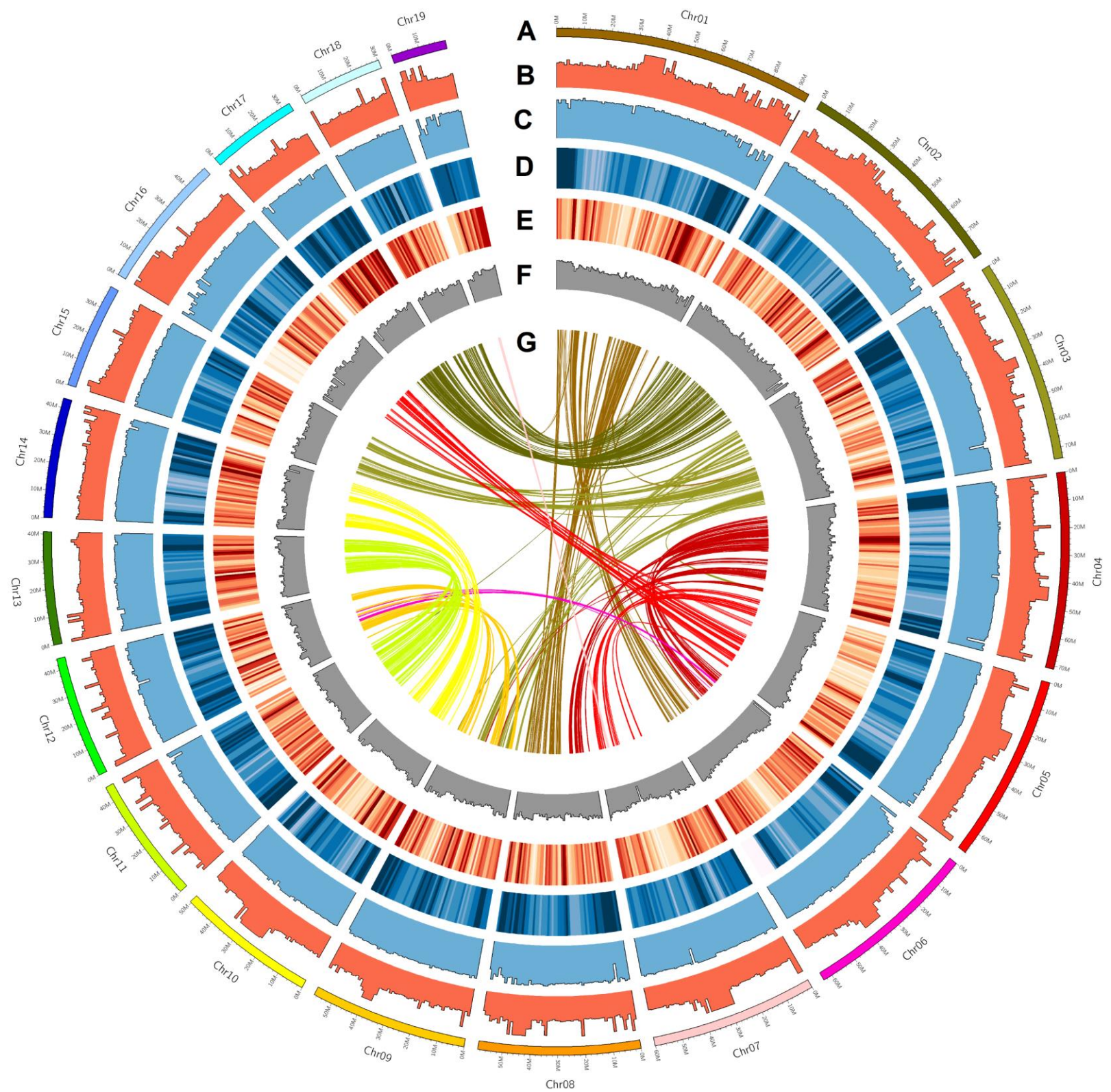


Figure 2

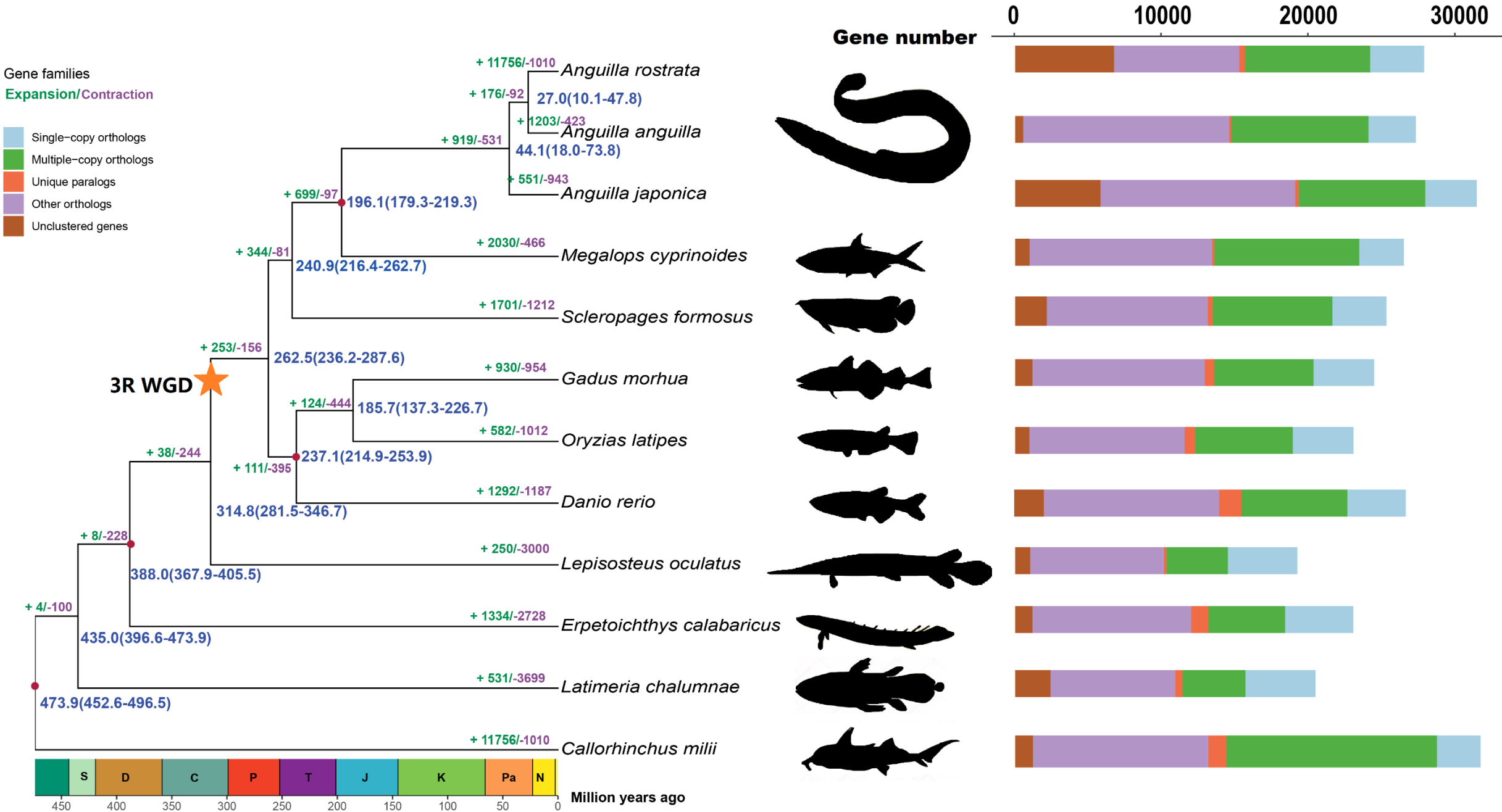


Figure 3

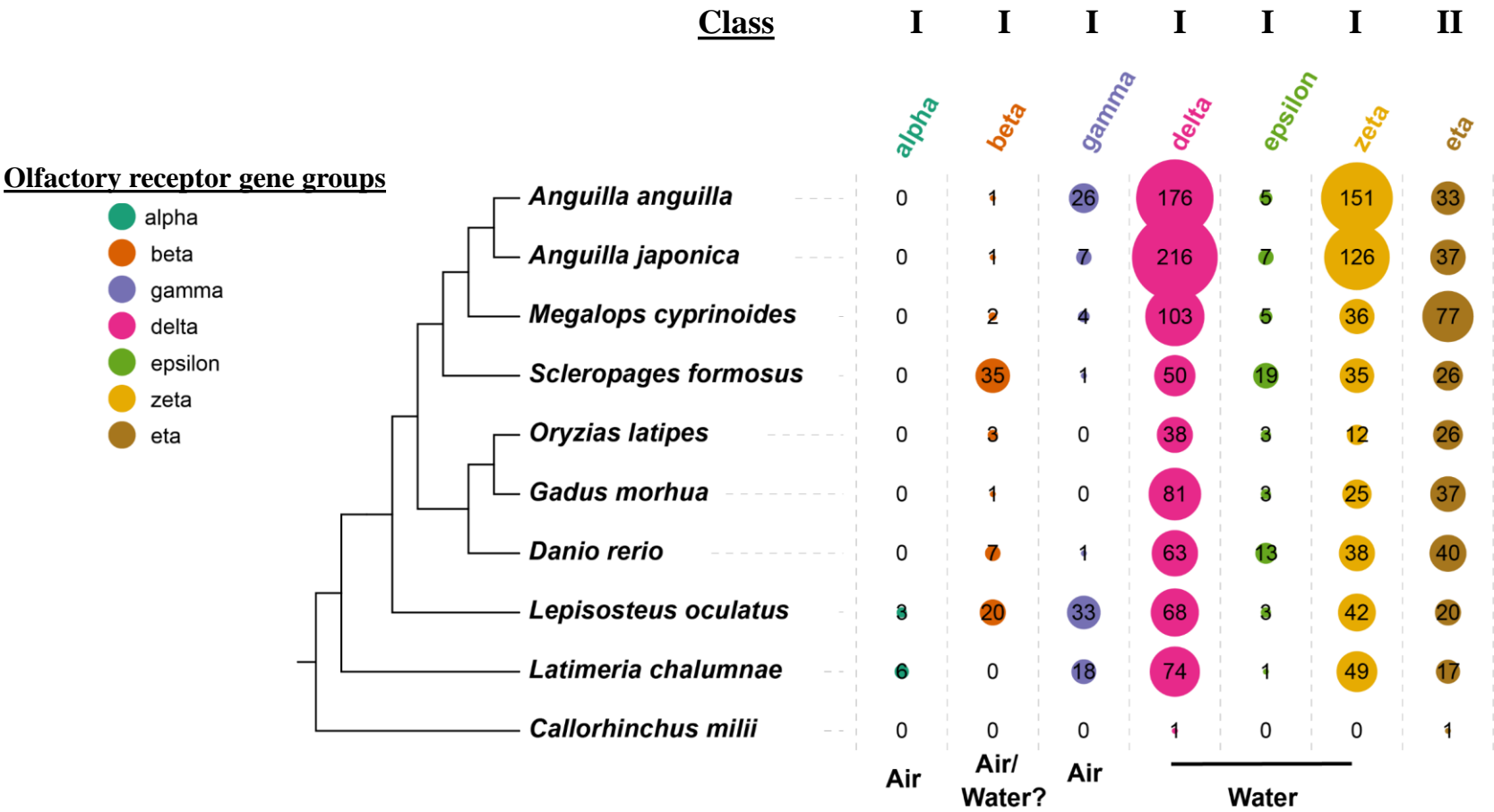
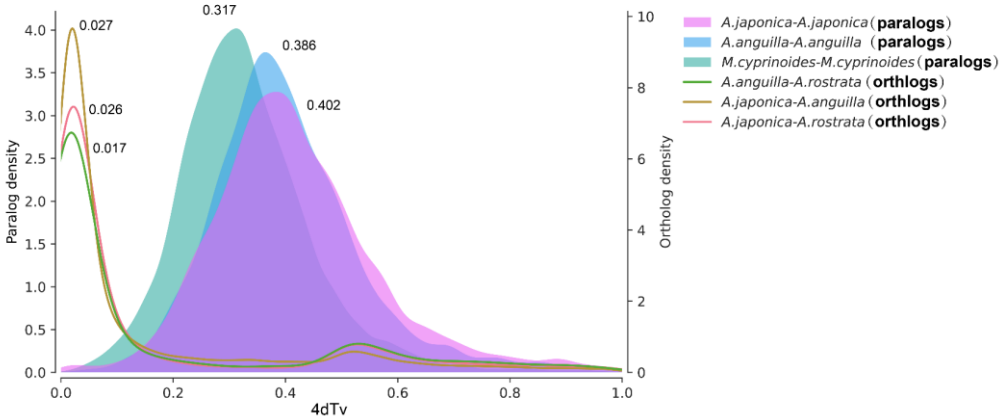
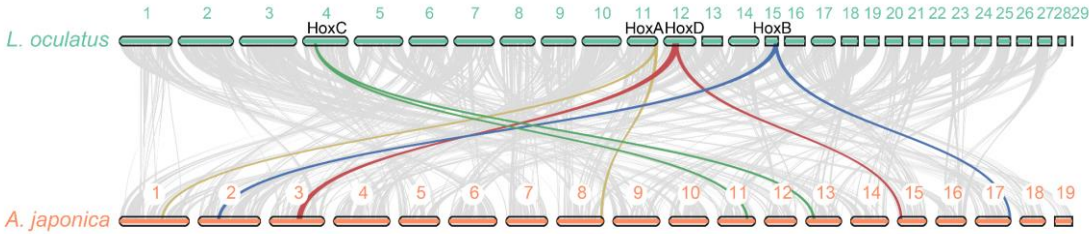


Figure 4

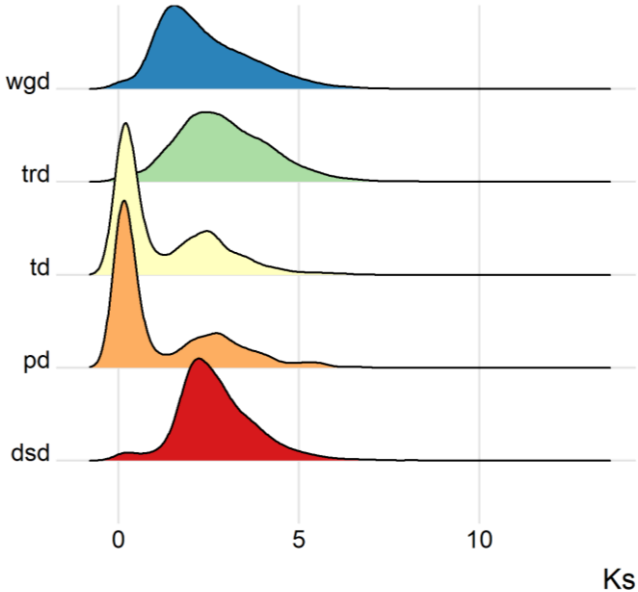
(A)



(B)



(C)



(D)

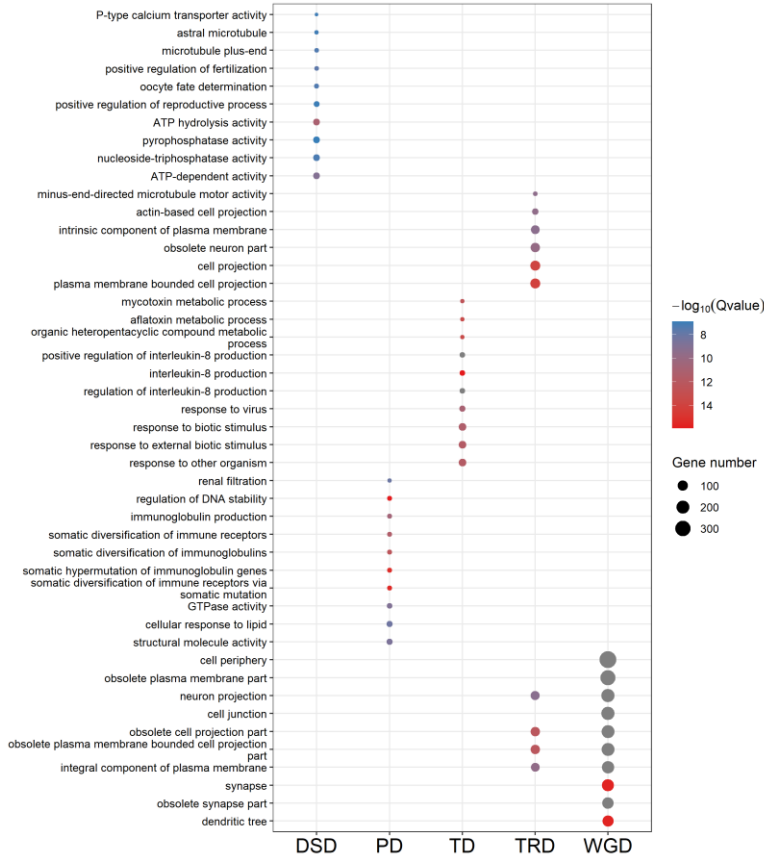
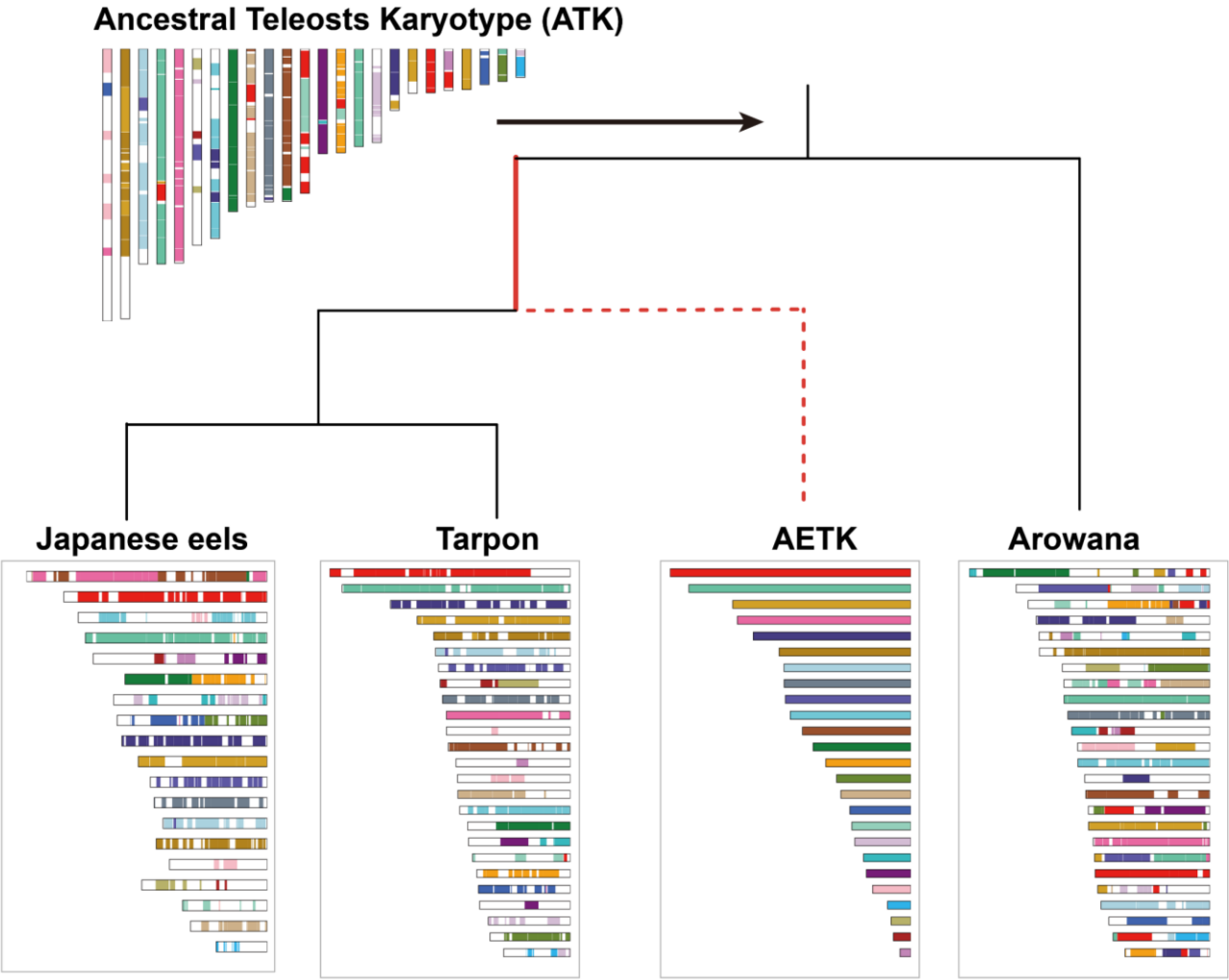


Figure 5

(A)



(B)

