Initial data release and announcement of the Fish10K: Fish 10,000 Genomes Project

Guanngyi Fan[1,5,*], Yue Song[1,*], Xiaoyun Huang[1,*], Liandong Yang[2,*], Suyu Zhang[1], Mengqi Zhang[1], Xianwei Yang[1], Yue Chang[1], He Zhang[1,5], Yongxin Li[3], Shanshan Liu[1], Lili Yu[1], Inge Seim[8,9], Chenguang Feng[3], Wen Wang[3], Kun Wang[3], Jing Wang[4,6,7], Xun Xu[5], Huanming Yang[1,5], Nansheng Chen[4,6,7,†], Xin Liu[1,5,†] & Shunping He[2,†].

[1]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China

[2]Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

[3]Center for Ecological and Environmental Sciences, Northwestern Polytechnical University, Xi'an, China.

[4]CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, Shandong 266071, China

[5]BGI-Shenzhen, Shenzhen 518083, China

[6]Marine Ecology and Environmental Science Laboratory, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao, Shandong 266237, PR China

[7]Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, Shandong 266400, PR China

[8]Integrative Biology Laboratory, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China;

[9]Comparative and Endocrine Biology Laboratory, Translational Research Institute-Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology, Brisbane 4102, Queensland,

Australia;

*These authors contributed equally

†e-mail: Shunping He (clad@ihb.ac.cn), Xin Liu (liuxin@genomics.cn) and Nansheng Chen (chenn@qdio.ac.cn).

Abstract

With more than 30,000 species, fish are the largest and most ancient vertebrate group. Despite their critical roles in many ecosystems and human society, fish genomics lags behind work on birds and mammals. This severely limits our understanding of evolution and hinders progress on the conservation and sustainable utilization of fish. Here, we announce the Fish10K project, an international collaborative project or initiative? aiming to sequence 10,000 representative fish genomes under a systematic context within ten years, and officially welcome collaborators to join this effort. As a step towards this goal, we herein describe a feasible workflow for the procurement and storage of biospecimens, and sequencing and assembly strategies. To illustrate, we present the genomes of ten fish species from a cohort of 93 species chosen for technology development.

Keywords: Fish10K, fish, genome sequencing, phylogenetics, evolution, stLFR

Background

Fish genomes sequenced to date

As of writing, genome assemblies are publicly available for less than 1% of fish species (216 species of 56 orders) (**Supplementary Table 1**). Their assembly length ranges from 302.36 Mb (*Diretmus argenteus*) to 3.81Gb (*Zeus faber*), with an average length of 813.74 Mb. The average scaffold N50 and contig N50 values are 8.53 Mb and 860.51 Kb, while the median scaffold N50 and contig N50 are 406.7Kb and 20.3Kb, respectively. There are 97 species with a scaffold N50 of more than 1Mb, of which 31 have a contig N50 above 1Mb (**Figure 1**). These genomes has fueled a number of studies on the phylogeny and evolution of fish (e.g., the African coelacanth genome and tetrapod evolution), evolutionary processes of specific fish subgroups (e.g., elephant shark genome illustrating the phylogenetic relationship of Chondrichthyes as a sister group to bony vertebrates) [1], genetic mechanisms of adaptation to different environments (e.g. the deep-sea Mariana Trench snailfish and cave-dwelling fish) [2], and specific biological processes (for example, the tonguefish *Cynoglossus semilaevis* genome for understanding ZW sex chromosome evolution) [3]. Nevertheless, the current fish genome sequencing results are only a drop in the ocean, and numerous critical research questions remain to be resolved. A non-exhaustive list includes gaining a comprehensive and clear understanding of fish phylogeny, genome size diversity and chromosome evolution, diverse environmental adaptations, morphology evolution, respiratory system, immune system, the evolution and function of ultra-conservative (UCE) and conserved nonexonic elements

3

(CNEEs).

The era of genome consortiums

With the rapid development of DNA sequencing technology, this is the time for large-scale, collaborative genomic studies. The first such project was the Genome 10K (G10K) Project established in 2009, which aimed to sequence and assemble genomes of about 10,000 vertebrate species [4]. Further advances in sequencing have extended this vision. The Vertebrate Genomes Project (VGP) was launched in 2017 to generate chromosome-level, haplotype-phased genome assemblies of vertebrate species [5]. The Bird 10,000 Genomes Project (B10K) was initiated [6] after the successful phylogenomic study on 45 avian genomes in 2014 [7]. The B10K projects aims to sequence and assemble all known bird species in three phases. Similar efforts have been made for bats [8], plants [9], and other species [10, 11]. Despite current challenges in funding, sampling, sequencing, assembly, and data analysis, these projects have already made substantial progress. For fish, which makes up more than half of all vertebrate species, no projects at the similar scale has been initiated. The only large-scale genomic study to our knowledge was Fish-T1K, which aimed to sequence the transcriptomes (RNA-seq) of ray-finned fishes [12]. However, the insights gained from transcriptome data alone is relatively limited. Accelerating fish genomics by large-scale genome sequencing efforts would undoubtedly boost research into fish biodiversity, speciation, adaptation, as well as aiding the conservation and sustainable utilization of fish.

The Fish10K Genome Project

We here announce the Fish10K Genome Project, aiming to sample, sequence, assemble, and analyze genomes of 10,000 fish species. We are proposing an effective and integrated workflow, in which major genomics challenges are addressed, to construct high-quality reference genomes. Through developing and applying effective analysis methods, we will be able to address critical evolutionary and biological research questions related to fish. In order to prove the efficiency of our workflow and the feasibility of this large-scale genome project, we are releasing ten high-quality genomes as part of a pilot project. We hope the released genomes, along with the other genomes generated by Fish10K, will be valuable resources for fish researchers as well as to fishery industry.

Main text

Feasibility test and the release of ten fish genomes

In order to establish cost-effective strategies and assess the feasibility of a large-scale genome project, we initiated a pilot study in June 2017. Over the last two years, we went on four expeditions across lakes, rivers, and coastal waters of China, collecting 324 fish species. After careful documentation of sample information and species identification, the tissues of 93 species were selected for DNA extraction and sequencing. We used single tube long fragment reads technology (stLFR) [13] and the DNBSEQ platform to sequence the species, generating long read (Nanopore or PacBio) and Hi-C data for a subset. In this way, we were able to test the feasibility of three different sequencing and assembly strategies (**Figure 2**): stLFR data alone (synthetic long reads generated using second-generation sequencing platform) (Strategy I); stLFR data combined with low-depth, long reads (~10× raw Nanopore data to fill in the gaps) (Strategy II); and high-depth long reads (~80× raw Nanopore data) combined with second generation short reads (either short insert size libraries or stLFR) (Strategy III). We have sequenced all the 93 species using stLFR (**Supplementary Table 2**), 26 species using Nanopore at low depth, and 13 using Nanopore at high depth. To date, we have assembled the genomes of 48 species, which are currently undergoing curation, and are finishing assembly for the remaining 57. For the 48 assembled fish genomes, the average contig N50 and the average scaffold N50 are 1.53 Mb and 5.20 Mb, respectively. The average Benchmarking Universal Single-Copy Orthologs (BUSCO) [14] completeness estimate is 93.48%

(**Supplementary Table 3**). A comparison of assembly statistics revealed that assemblies generated with Strategy II and Strategy III were more continuous, achieving an average contig N50 of 3.99 Mb and an average scaffold N50 of 9.51 Mb. They are notably superior to Strategy I (average contig N50 of 79.04 Kb and average scaffold N50 of 4.73Mb). To illustrate our effort, we are releasing the genomes of ten representative bony fish genomes covering the three assembly strategies (

Table 1). The contig N50s of seven of these genomes are more than 1 Mb and a minimum 93% of BUSCO genes were found, indicating the genome assemblies are of high quality. Three genomes were assembled at chromosome-level, with more than 92% scaffold-anchored ratio using Hi-C data.

The Fish10K Genome Project: from 100 to 10,000

With the experience gained in the Fish10K pilot study and our published results, we believe that the project can scale-up. Thus, we are proposing a roadmap (**Figure 3**) in which we will construct high-quality reference genomes for representative species in all orders (Phase I) and families (Phase III) in concert with the generation of draft genome sequences for additional related species (Phase III).

An interrogation of FishBase [15] and "Fishes of the world" (5th ed.) [16] revealed information on 34,115 fish species from ~5,000 genera, ~529 families, and ~80 orders (**Supplementary Table 2**). The species were divided into six lineages (Elasmobranchii, Holocephali, Actinopterygii, Sarcopterygii, Cephalaspidomorphi, and Myxini), in which Elasmobranchii and Holocephali belong to Chondrichthyes (cartilaginous fishes); and Actinopterygii and Sarcopterygii belong to Osteichthyes (bony fishes). As mentioned above, there are reference genomes available for at least one species of 56 orders, while for the rest of the orders reference genomes are required. Also, there are fish orders with a large number of species (e.g., Perciformes has 62 families; Siluriformes has 40 families; and Scorpaeniformes has 39 families), suggesting that additional high-quality reference genomes are required to represent the diverse biological characteristics. Thus, in Phase I we aim to sequence 450 bony

fish and 50 cartilages fish species, covering all 80 orders (**Supplementary Table 2**).

In Phase II, we aim to sequence approximately 3,000 species, covering almost all ~500 fish families. In Phase III, we will sequence ~6,500 fish genomes, covering ~5,000 genera.

Sampling, sequencing, assembly, and annotation

Sampling is a critical challenge in any large-scale genome consortium. We propose a centralised sampling mode (i.e., mirroring our 93-species pilot phase), with several sampling centres set up to collect samples. In addition to these sampling centres, we would like to obtain further samples from around the world. To make sure we have enough information for further analysis and to maximise the value of the genome data, we propose a sampling standard. The associated meta-data was designed to include as much information as possiblestressing the importance of collecting images of each specimen and adequate storage conditions (frozen or voucher specimen).

For sequencing, we propose to use both second- and third-generation sequencing technologies to generate high-quality genome assemblies. Based on our pilot study, and considering the feasibility of obtaining the required amount of high molecular DNA, we have chosen a 'stLFR data + low-depth Nanopore data + Hi-C data' strategy (Strategy II in **Figure 2**) for the majority of the species. For more complex genomes, we will generate high-depth Nanopore sequence data to ensure that good assemblies can be achieved (Strategy III, 'stLFR data + high-depth Nanopore data + Hi-C data'; **Figure 2**). For key species (to be determined by the working groups; see below), we will employ a Pacific Biosciences circular consensus sequencing (CCS) long

9

high-fidelity (HiFi) approach, allowing the generation of highly-accurate long reads [17]. For the large-scale sequencing of 6,000 species in Phase III, we propose to employ stLFR alone (Strategy I in **Figure 2**). For a diploid species with a genome size of less than 10Gb (generated using our preferred Strategy II), we will require the contig N50 and scaffold N50 to be longer than 1Mb and 10Mb, respectively, and (if applicable) to anchor more than 90% of the assembled sequences to chromosomes. The same criteria will apply to assemblies generated using a high-depth long read strategy (Strategy III). For assemblies generated using stLFR sequencing alone (Phase III) assemblies must have a contig N50 and scaffold N50 longer than 100 Kb and 1Mb, respectively. All assemblies must have a BUSCO completeness estimate higher than 90%. Finally, genome feature annotations (e.g., repeat and gene annotations) will be performed using well-established in-house pipelines.

Fish10k data sharing

Per the Fort Lauderdale [18] and Toronto International Data Release Workshop guidelines [19], all sequencing data (including raw data, assemblies, and annotations) will be deposited in the appropriate INSDC databases, alongside GigaDB and China National Gene Bank (CNGB). The website of Fish10K (http://icg-ocean.genomics.cn/index.php/fish10kintroduction) will provide detailed information on the project status, as well as continuously updated information on the sequenced species. It also provides a portal for data download (in particular for assembled genomes).

Organisation of Fish10K consortium

Fish10K has been initiated by a core group of researchers, forming the steering committee of Fish10K (**Figure 4**). The steering committee oversees the project and is responsible for fundraising, expanding the steering committee, organising the scientific groups and species groups, and coordinating sampling, sequencing, assembly, and analysis strategies. The steering committee is also responsible for the generation of genomic data. Various scientific groups will focus on technical and scientific questions related to this project. The scientific groups, which will have advance access to all generated data, will include a sampling group, a sequencing and assembly group, and a series of groups focusing on different fish-related scientific questions. We wish to receive proposals from researchers who would like to take part in scientific groups. We also invite researchers who are studying fish species which are rare or extinct to join Fish10K as members in the species group (with or without associated funding for sequencing). In addition to obtaining the genome sequences of their area of interest, joining the consortium provides immediate access to all genomes currently being assembled by Fish10K.

Conclusions

Fish10K will generate an unprecedented, comprehensive data set of fish, the largest and most diverse vertebrate group. Our effort will allow us to complete the genomic tree for fish and, in concert with other projects such as VGP and B10K, vertebrates in general.

Abbreviations

B10K: Bird 10,000 Genomes Project

VGP: Vertebrate Genomes Project

EBP: Earth BioGenome Project

Fish-T1K: Transcriptomes of 1,000 Fishes Project

Fish10K: Fish 10,000 Genomes Project

CNGB: China National Gene Bank

stLFR: single tube Long Fragment Read

NCBI: National Center for Biotechnology Information

Data availability

The ten fish genome assemblies in the pilot have been deposited in the CNSA (https://db.cngb.org/cnsa/) of CNGBdb with accession codes CNP0000597 and CNP0000691.

Author contributions

S-P. H., Q. Q., N-S. C., X.X., X. L., W. W. and G-Y. F. conceived and designed the study. L-D. Y., M-Q. Z., S.-S. L., and D.S., C. performed sample collection and sequencing. Y. S., S.-Y. Z., and X.-Y. H. performed assembly and annotation. X. L., Y. S., and G. F. wrote the manuscript. Y.-X. L., N.-S. C., and all other authors revised and read the manuscript.

Acknowledgements

References

1.   Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, et al. Elephant shark genome provides unique insights into gnathostome evolution. 2014;505 7482:174.

2.   Wang K, Shen Y, Yang Y, Gan X, Liu G, Hu K, et al. Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation. 2019;3 5:823.

3.   Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, et al. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nature genetics. 2014;46 3:253.

4.   Koepfli K-P, Paten B, Scientists GKCo and O'Brien SJJARAB. The Genome 10K Project: a way forward. 2015;3 1:57-111.

5.   Vertebrate Genomes Project (VGP). https://genome10k.soe.ucsc.edu/vertebrate-genomes-project/. Accessed 10 September 2019.

6.   Zhang GJN. Genomics: Bird sequencing project takes off. 2015;522 7554:34.

7.   Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. 2014;346 6215:1311-20.

8.   BAT 1K. https://bat1k.ucd.ie/. Accessed 27 September 2019.

9.   Cheng S, Melkonian M, Smith SA, Brockington S, Archibald JM, Delaux P-M, et al. 10KP: A phylodiverse genome sequencing plan. 2018;7 3:giy013.

10.   DNA ZOO. https://www.dnazoo.org/. Accessed 27 September 2019.

11.   Ostrander EA, Wang G-D, Larson G, vonHoldt BM, Davis BW, Jagannathan V, et al. Dog10K: An international sequencing effort to advance studies of canine domestication, phenotypes, and health. 2019.

12.   Sun Y, Huang Y, Li X, Baldwin CC, Zhou Z, Yan Z, et al. Fish-T1K (Transcriptomes of 1,000 Fishes) Project: large-scale transcriptome data for fish evolution studies. 2016;5 1:18.

13.   Wang O, Chin R, Cheng X, Wu MKY, Mao Q, Tang J, et al. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. Genome research. 2019;29 5:798-808.

14.   Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2017;35 3 doi:10.1093/molbev/msx319.

15.   FishBase. https://www.fishbase.se/search.php. Accessed 10 September 2019.

16. Nelson JS, Grande TC and Wilson MV. Fishes of the World. John Wiley & Sons; 2016.

17. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Highly-accurate long-read sequencing improves variant detection and assembly of a human genome. 2019:519025.

18. The Fort Lauderdale Agreement. Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects. Available at: https://www.genome.gov/10506537/reaffirmation-and-extension-of-nhgri-rapid-data-release-policies. Accessed 10 September 2019.

19. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, et al. Prepublication data sharing. 2009;461 7261:168-70.

Figures and Tables
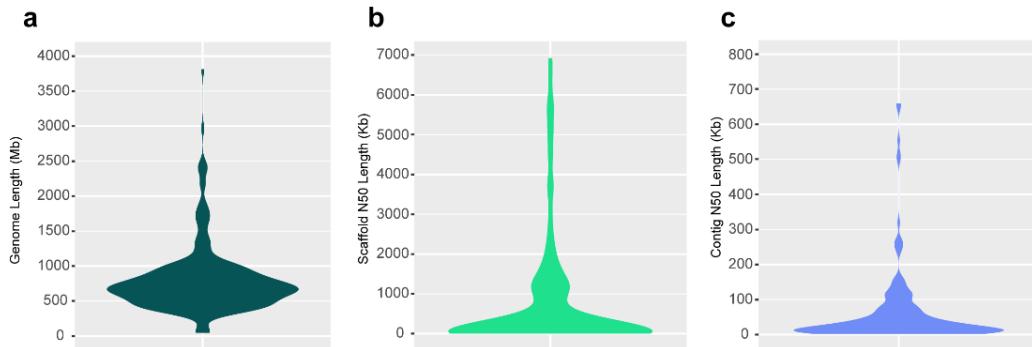
Figure 1. Assembly statistics of fish genomes in public databases. a) Summary of genome size. b, c) N50 statistics. A scaffold is a series of contigs that can be linked. N50 is the sequence length of the shortest contig (or contig) at 50% of the total genome length.
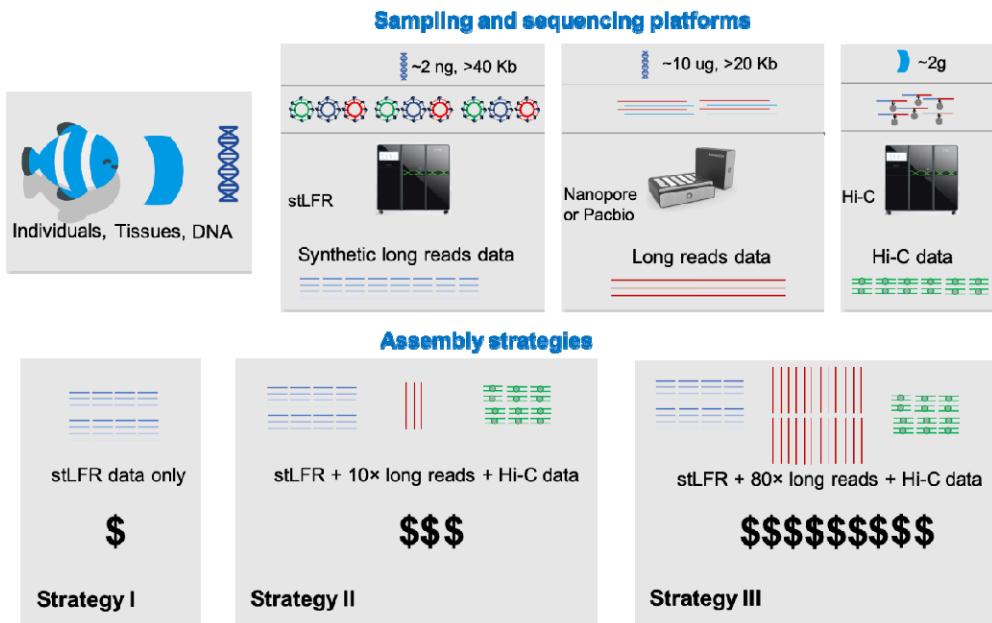
Figure 2. The sequencing and assembly strategies. In the preferred strategy (Strategy II), high-quality DNA fragments (≥40Kb) are used to construct a stLFR library which is sequenced using the DNBSEQ platform. Low-sequencing-depth long reads are only used to improve the continuity of highly complex regions (increase the contig N50). In the alternative Strategy I, high-depth long reads are used to construct contigs, while low-depth stLFR reads are used to polish the contig and link the scaffolds. Hi-C data is used to generate a chromosome-level assembly.
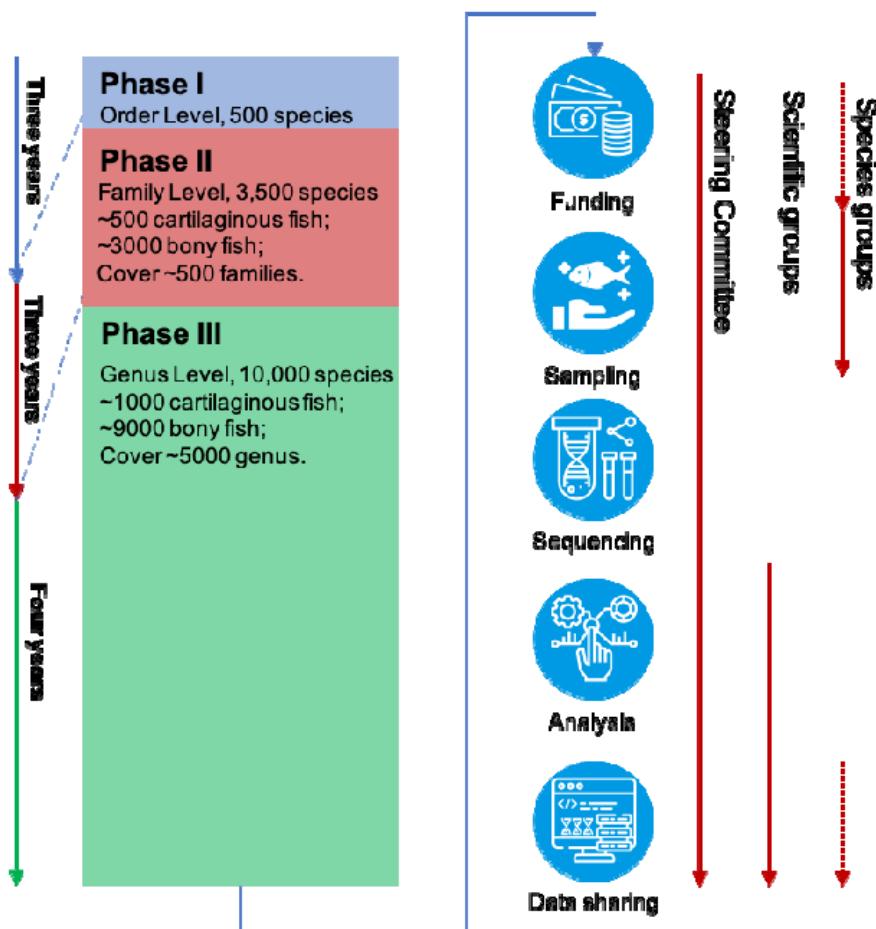
Figure 3. The roadmap and organisation of Fish10K. Fish10K is divided into three phases, based on the evolutionary relationship of fish, and three working groups (steering committee, scientific groups, and species groups).
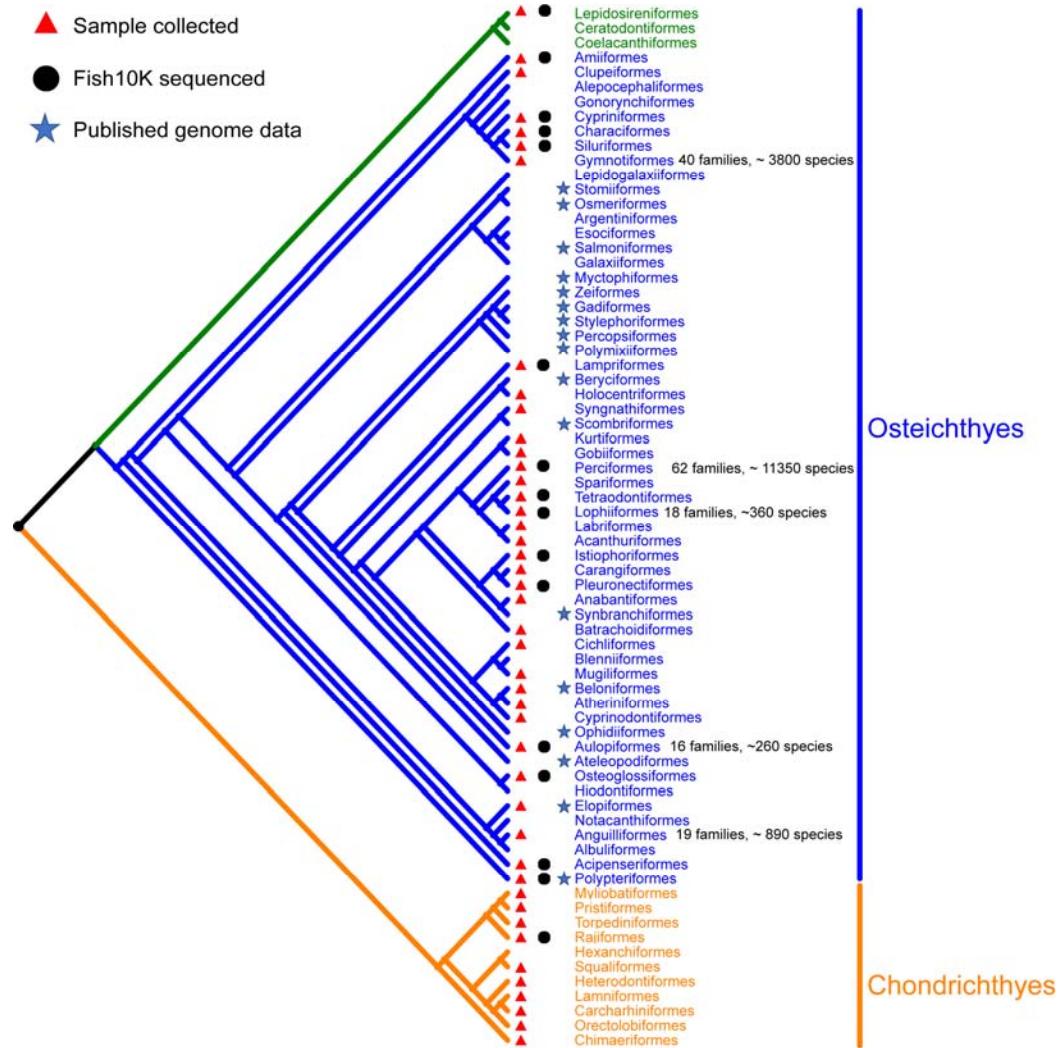
Figure 4. Phylogenetics tree of fish. Jawed vertebrates (gnathostomes) are divided into two major groups, cartilaginous fishes (Chondrichthyes; in orange) and bony vertebrates (Osteichthyes; in blue and green). Bony fishes are grouped into two subgroups (Sarcopterygii; green) and (Actinopterygii; blue). The number of families and species of the five largest orders are labeled. The remaining ten orders of bony fish (Caproiformes, Callionymiformes, Gobiesociformes, Icosteiformes, Lepisosteiformes, Moroniformes, Scombrolabraciformes, Scorpaeniformes, Trachichthyiformes and Trachiniformes) and two orders of cartilaginous fish (Rhinopristiformes and Squatiniformes) are not included in the phylogenetic tree, due to their uncertain position.

Table 1. Assembly statistics of the ten released genome assemblies. An asterisk (*) denotes a chromosome-level genome assembly (Hi-C data generated).

Strategy	Species	Estimated genome size (Mb)	Assembly size (Mb)	Scaffold N50 (bp)	Contig N50 (bp)	BUSCO (%)	Anchored (%)
I	*Diodon holocanthus*	722.9	643.4	6,098,089	2,149,931	95.7	--
	*Heterotis niloticus**	743.4	669.7	9,615,753	2,307,881	97.6	96.8
	Oxyeleotris marmorata	589.7	502.6	13,190,768	1,270,297	92.9	--
	Datnioides pulcher	623.1	595.7	9,741,635	2,175,996	97.2	--
	Chaetodon trifasciatus	698.5	668.3	9,974,986	1,859,054	97.3	--
II	*Naso vlamingii*	961.4	861.3	5,736,754	182,642	97.80	--
	*Chelmon rostratus**	711.4	638.9	2,627,953	294,414	98.40	94.40
	*Helostoma temminckii**	729.7	635.4	913,351	95,536	96.30	91.80
III	*Pseudobrama simoni*	940.9	929.1	13,799,189	13,799,189	95.70	--
	Rhodeus ocellatus	850.5	902.4	4,198,183	4,198,183	94.50	--

18

Table 2. Sample collection template.

Species	Length (cm)	Weight (g)	Sex	Meta	Intestinal	Muscle	Liver	Time	Place	longitude and latitude	Photo	Sampling person	Identification person	Status
Sebastiscus marmoratus	11.5	27	♂	√	√	√	√	2019 0421	Xiamen	N24° 11′ 59.58″ E118° 25′ 1.92″	√	Dr. Meng	Prof. He	Living
Pisodonophis cancrivorus	12	31	♂	√	√	√	√	2019 0421	Ningde	N24° 11′ 59.58″ E118° 25′ 1.92″	√	Dr. Meng	Prof. He	Fresh
Odontobutis obscura	13.3	35	♀	√	√	√	√	2019 0421	Hangzhou	N24° 11′ 59.58″ E118° 25′ 1.92″	√	Dr. Meng	Prof. He	Frozen

Supplementary information

Supplementary Table 1. Assembly information of publicly available fish genomes.

Extant fishes are divided into three major classes, among which Actinopterygii (ray-finned fishes) account for more than half.

Species	NCBI Taxonomy ID	Assembly level	Total sequence length (bp)	Number of scaffolds	Scaffold N50 (bp)	Number of contigs	Contig N50 (bp)	Accession Number
Acanthochaenus luetkenii	473344	Scaffold	545,759,480	91,087	8,444	131,532	5,636	GCA_900312575.1
Acanthochromis polyacanthus	80966	Scaffold	991,584,656	30,414	334,400	159,493	16,099	GCF_002109545.1
Acipenser ruthenus	7906	Scaffold	1,732,545,901	215,913	219,750	446,905	24,377	GCA_004119895.1
Ageneiosus marmoratus	2066578	Scaffold	1,030,000,983	16,063	223,139	169,048	7,741	GCA_003347165.1
Amphilophus citrinellus	61819	Scaffold	844,902,565	6,637	1,216,136	67,543	23,456	GCA_000751415.1
Amphiprion ocellaris	80972	Scaffold	880,720,895	6,405	401,715	7,803	324,210	GCF_002776465.1
Amphiprion percula	161767	Chromosome	908,955,932	366	38,416,550	1,048	3,123,421	GCA_003047355.2
Anabarilius grahami	495550	Scaffold	991,887,266	80,398	4,459,447	131,192	36,058	GCA_003731715.1
Anarrhichthys ocellatus	433405	Scaffold	612,774,492	10,816	5,717,598	33,924	41,757	GCA_004355925.1
Anguilla anguilla	7936	Scaffold	1,018,701,900	501,148	59,657	865,467	2,544	GCA_000695075.1
Anguilla japonica	7937	Scaffold	966,917,315	83,292	36,264,158	276,102	11,014	GCA_003597225.1
Anguilla rostrata	7938	Scaffold	1,413,032,609	79,209	86,641	307,316	7,355	GCA_001606085.1
Anoplogaster cornuta	88656	Scaffold	404,844,788	108,625	4,562	128,159	3,719	GCA_900683385.1
Anoplopoma fimbria	229290	Contig	699,326,415			208,506	5,156	GCA_000499045.1
Antennarius striatus	241820	Scaffold	441,856,641	70,398	9,774	103,068	6,086	GCA_900303275.1
Aphyosemion australe	52653	Scaffold	868,348,903	12,236	1,435,212	30,782	119,465	GCA_006937985.1
Arapaima gigas	113544	Scaffold	667,351,951	60,052	285,171	70,752	64,954	GCA_007844225.1
Archocentrus centrarchus	63155	Chromosome	932,930,362	188	35,590,001	926	2,146,538	GCF_007364275.1
Arctogadus glacialis	185735	Scaffold	428,791,846	139,389	3,702	151,920	3,282	GCA_900303235.1

Astatotilapia burtoni	8153	Scaffold	831,411,547	8,001	1,194,190	69,074	21,886	GCF_000239415.1
Astatotilapia calliptera	8154	Chromosome	880,445,564	249	38,669,361	739	4,438,245	GCF_900246225.1
Astyanax mexicanus	7994	Chromosome	1,335,239,194	2,415	35,377,769	3,030	1,767,240	GCF_000372685.2
Austrofundulus limnaeus	52670	Scaffold	866,963,281	29,785	1,098,383	168,369	8,097	GCF_001266775.1
Bathygadus melanobranchus	630650	Scaffold	431,202,967	92,290	6,483	112,630	4,956	GCA_900302375.1
Benthosema glaciale	125796	Scaffold	676,314,385	143,923	6,111	188,319	4,393	GCA_900323375.1
Beryx splendens	88663	Scaffold	533,267,752	117,400	5,987	151,933	4,286	GCA_900312565.1
Betta splendens	158456	Chromosome	441,388,503	70	20,129,463	398	2,497,747	GCF_900634795.2
Boleophthalmus pectinirostris	150288	Scaffold	955,752,150	16,620	2,375,582	108,947	20,437	GCF_000788275.1
Boreogadus saida	44932	Scaffold	412,070,465	137,701	3,572	147,911	3,221	GCA_900302515.1
Borostomias antarcticus	473354	Scaffold	430,362,827	104,734	5,368	133,045	3,928	GCA_900323325.1
Bregmaceros cantori	630652	Scaffold	1,144,104,325	258,566	5,922	319,170	4,452	GCA_900302395.1
Brosme brosme	81638	Scaffold	412,731,310	114,910	4,650	136,096	3,682	GCA_900302425.1
Brotula barbata	432164	Scaffold	485,061,200	29,854	45,752	59,402	17,578	GCA_900303265.1
Callopanchax toddi	60409	Scaffold	853,385,936	9,724	1,656,350	45,972	52,012	GCA_006937965.1
Carapus acus	1491482	Scaffold	387,834,307	46,699	16,922	70,747	9,554	GCA_900312935.1
Carassius auratus	7957	Chromosome	1,820,635,050	6,216	22,763,433	8,463	821,153	GCF_003368295.1
Coryphaenoides rupestris	163118	Scaffold	829,208,733	47,680	159,738	82,633	20,848	GCA_002895965.1
Cottoperca gobio	56716	Chromosome	609,391,784	322	25,156,145	766	6,330,900	GCF_900634415.1
Cottus rhenanus	446433	Scaffold	563,609,416	164,693	7,249	490,620	2,129	GCA_001455555.1
Cynoglossus semilaevis	244447	Chromosome	470,199,494	31,181	509,861	62,912	27,008	GCF_000523025.1
Cyprinodon variegatus	28743	Scaffold	1,035,184,475	9,259	835,301	110,959	20,803	GCF_000732505.1
Cyprinus carpio	7962	Chromosome	1,713,658,011	9,378	7,828,959	53,088	75,080	GCF_000951615.1
Cyttopsis rosea	1176755	Scaffold	546,506,150	111,974	7,082	150,231	4,843	GCA_900302355.1
Danio rerio	7955	Chromosome	1,373,454,788	1,917	7,379,053	19,725	1,422,317	GCF_000002035.6
Danionella dracula	623740	Scaffold	665,208,374	996	10,287,669	1,611	2,300,191	GCA_900490495.1
Danionella translucida	623744	Scaffold	735,303,417	27,639	340,819	36,005	133,131	GCA_007224835.1
Denticeps clupeoides	299321	Chromosome	567,401,054	460	22,793,177	924	3,059,612	GCF_900700375.1

21

Dicentrarchus labrax	13489	Scaffold	675,917,103	25	26,439,989	37,781	54,134	GCA_000689215.1
Diretmoides pauciradiatus	1415272	Scaffold	672,603,014	147,686	6,032	204,934	4,028	GCA_900660315.1
Diretmus argenteus	88682	Scaffold	302,363,458	107,968	3,367	162,384	2,172	GCA_900660295.1
Echeneis naucrates	173247	Chromosome	544,229,245	38	23,287,306	178	12,371,513	GCF_900963305.1
Electrophorus electricus	8005	Scaffold	551,880,868	8,786	613,956	47,652	37,141	GCF_003665695.1
Epinephelus lanceolatus	310571	Chromosome	1,087,399,367	4,200	46,227,939	22,225	159,800	GCA_005281545.1
Erpetoichthys calabaricus	27687	Chromosome	3,811,038,701	1,885	199,226,436	7,498	1,143,051	GCF_900747795.1
Esox lucius	8010	Chromosome	940,906,975	811	37,550,661	1,395	3,396,779	GCF_004634155.1
Fundulus heteroclitus	8078	Scaffold	1,021,898,560	10,180	1,252,252	120,723	16,688	GCF_000826765.1
Gadiculus argenteus	185737	Scaffold	396,767,394	123,363	3,951	137,724	3,379	GCA_900302595.1
Gadus chalcogrammus	1042646	Scaffold	448,868,398	130,159	4,335	149,286	3,603	GCA_900302575.1
Gambusia affinis	33528	Scaffold	598,663,367	2,943	6,651,460	73,682	17,511	GCA_003097735.1
Gasterosteus aculeatus	69293	Scaffold	467,452,432	10,242	3,715,221	32,646	38,090	GCA_006229165.1
Gephyroberyx darwinii	334984	Scaffold	535,045,983	55,277	18,145	102,041	8,514	GCA_900660455.1
Gouania willdenowi	441366	Chromosome	937,150,793	441	38,978,045	1,594	1,838,341	GCF_900634775.1
Guentherus altivela	1263181	Scaffold	539,598,795	189,411	3,201	205,243	2,928	GCA_900312595.1
Haplochromis nyererei	303518	Scaffold	830,133,247	7,236	2,525,540	68,053	22,622	GCF_000239375.1
Hippocampus comes	109280	Scaffold	493,775,940	37,377	2,034,572	60,478	39,546	GCF_001891065.1
Holocentrus rufus	722565	Scaffold	649,757,301	58,113	21,389	113,697	9,243	GCA_900302615.1
Hoplostethus atlanticus	96778	Scaffold	520,173,038	58,279	15,562	103,281	7,812	GCA_900660355.1
Hucho hucho	62062	Scaffold	2,487,549,814	71,639	287,338	221,746	37,639	GCA_003317085.1
Hypophthalmichthys molitrix	13095	Scaffold	1,104,676,189	107,095	314,181	1,528,765	2,130	GCA_004764525.1
Hypophthalmichthys nobilis	7965	Scaffold	1,012,063,666	121,326	83,012	742,098	4,840	GCA_004193235.1
Hypoplectrus puella	146810	Scaffold	612,290,098	14,375	24,210,077	64,238	22,581	GCA_900610375.1

Species								
Ictalurus punctatus	7998	Chromosome	1,002,389,428	3,163	26,676,597	5,816	2,695,784	GCA_004006655.2
Kryptolebias hermaphroditus	1747188	Chromosome	683,986,837	5,211	27,459,464	32,872	46,411	GCA_007896545.1
Kryptolebias marmoratus	37003	Chromosome	1,002,389,428	3,163	26,676,597	5,816	2,695,784	GCA_004006655.2
Labeo rohita	84645	Scaffold	1,484,730,970	13,623	1,959,535	42,076	522,833	GCA_004120215.1
Labrus bergylta	56723	Scaffold	805,480,521	13,466	794,648	13,723	703,847	GCF_900080235.1
Laemonema laureysi	1784819	Scaffold	306,494,646	86,525	4,715	108,340	3,431	GCA_900303225.1
Lampris guttatus	81370	Scaffold	849,277,706	208,230	5,222	253,854	4,051	GCA_900302545.1
Lamprogrammus exutus	1592065	Scaffold	492,850,272	120,937	5,470	145,536	4,213	GCA_900312555.1
Larimichthys crocea	215358	Chromosome	657,939,657	9,998	27,037,660	16,979	277,487	GCF_000972845.2
Lates calcarifer	8187	Scaffold	668,481,366	3,808	1,191,366	3,918	1,066,117	GCF_001640805.1
Lepisosteus oculatus	7918	Chromosome	945,878,036	2,106	6,928,108	45,200	68,337	GCF_000242695.1
Lesueurigobius sanzi	1365564	Scaffold	810,626,388	130,360	11,480	186,717	6,729	GCA_900303255.1
Leuciscus waleckii	155063	Scaffold	752,538,629	4,888	21,959,719	38,277	38,877	GCA_900092035.1
Liparis tanakae	230148	Scaffold	498,979,456	27,878	375,216	97,972	9,903	GCA_006348945.1
Lota lota	69944	Scaffold	397,499,185	106,616	4,892	128,281	3,803	GCA_900302385.1
Maccullochella peelii	135761	Scaffold	633,241,041	18,198	109,974	31,008	70,439	GCA_002120245.1
Macquaria australasica	135765	Scaffold	675,976,139	2,962	845,515	3,369	678,975	GCA_005408345.1
Macrourus berglax	473319	Scaffold	399,875,629	118,318	4,291	142,500	3,353	GCA_900302365.1
Malacocephalus occidentalis	630739	Scaffold	350,339,566	95,829	4,932	116,459	3,697	GCA_900312585.1
Mastacembelus armatus	205130	Chromosome	591,935,101	122	25,090,313	358	8,014,513	GCA_900324485.2
Maylandia zebra	106582	Chromosome	957,485,262	1,690	32,660,920	2,331	1,407,748	GCF_000238955.4
Melanogrammus aeglefinus	8056	Scaffold	652,790,733	8,420	209,126	15,188	77,605	GCA_900291075.1
Melanonus zugmayeri	181410	Scaffold	432,902,915	82,409	7,633	123,515	4,562	GCA_900302345.1
Merlangius merlangus	8058	Scaffold	423,942,190	122,642	4,444	144,924	3,538	GCA_900323355.1
Merluccius capensis	89947	Scaffold	414,317,329	110,925	4,774	131,629	3,792	GCA_900312945.1
Merluccius	8063	Scaffold	401,034,705	102,914	5,117	133,682	3,670	GCA_900312545.1

23

merluccius								
Merluccius polli	89951	Scaffold	401,149,128	113,894	4,482	138,586	3,471	GCA_900312625.1
Micropterus floridanus	225391	Contig	1,001,521,525			249,768	10,978	GCA_002592385.1
Miichthys miiuy	240162	Scaffold	619,300,777	6,294	1,145,539	20,386	81,271	GCA_001593715.1
Mola mola	94237	Scaffold	639,451,992	5,552	8,766,736	51,826	23,239	GCA_001698575.1
Molva molva	163112	Scaffold	437,480,619	111,875	5,266	133,189	4,136	GCA_900323295.1
Monocentris japonica	181435	Scaffold	556,023,515	52,108	18,672	109,034	8,046	GCA_900323365.1
Mora moro	248765	Scaffold	344,961,111	100,621	4,433	125,652	3,267	GCA_900303205.1
Morone chrysops	46259	Scaffold	620,984,155	84,096	51,932	111,517	20,318	GCA_003610055.1
Morone saxatilis	34816	Scaffold	598,109,547	629	25,942,274	70,506	17,211	GCA_004916995.1
Muraenolepis marmorata	487677	Scaffold	416,390,766	138,928	3,555	152,594	3,126	GCA_900302325.1
Myoxocephalus scorpius	8097	Scaffold	520,316,443	85,863	9,473	127,441	5,716	GCA_900312955.1
Myripristis jacobus	371672	Scaffold	720,396,841	64,974	21,306	121,390	9,816	GCA_900302555.1
Myripristis murdjan	586833	Chromosome	835,254,674	87	34,950,760	340	14,475,636	GCF_902150065.1
Neogobius melanostomus	47308	Contig	1,003,738,541			1,364	2,817,412	GCA_007210695.1
Neolamprologus brichardi	32507	Scaffold	847,910,432	9,099	4,430,025	118,197	13,047	GCF_000239395.1
Nibea albiflora	240163	Scaffold	574,466,150	11,977	2,154,052	34,769	55,145	GCA_900327885.1
Nothobranchius furzeri	105023	Chromosome	1,242,518,059	6,013	15,858,201	74,941	19,950	GCF_001465895.1
Nothobranchius kuhntae	321403	Scaffold	1,122,656,415	34,756	1,178,460	101,919	22,234	GCA_006942095.1
Notothenia coriiceps	8208	Scaffold	636,613,682	38,657	217,655	72,571	17,492	GCF_000735185.1
Oncorhynchus kisutch	8019	Chromosome	2,369,932,239	22,813	1,266,128	97,074	58,118	GCF_002021735.1
Oncorhynchus mykiss	8022	Chromosome	2,178,999,613	139,800	1,670,138	559,855	13,827	GCF_002163495.1
Oncorhynchus nerka	8023	Chromosome	1,927,141,915	38,027	1,058,586	57,813	329,583	GCF_006149115.1
Oncorhynchus tshawytscha	74940	Chromosome	2,425,713,975	15,946	1,728,323	69,485	133,169	GCF_002872995.1
Ophiodon elongatus	225387	Scaffold	635,567,917	18,379	5,092,707	52,084	31,240	GCA_004358465.1
Oplegnathus fasciatus	163134	Scaffold	766,301,214	4,149	1,126,915	66,839	29,992	GCA_003416845.1
Opsanus beta	95145	Scaffold	1,028,783,780	345,629	3,335	371,656	3,062	GCA_900660325.1

Oreochromis aureus	47969	Scaffold	918,937,175	12,951	1,102,239	61,878	60,340	GCA_005870065.1
Oreochromis niloticus	8128	Chromosome	1,005,681,550	2,460	38,839,487	3,010	2,923,640	GCF_001858045.2
Oreochromis spilurus	64544	Contig	764,974,731			221,829	11,851	GCA_008269305.1
Oryzias javanicus	123683	Chromosome	809,679,899	254	35,390,520	751	3,558,013	GCA_003999625.1
Oryzias latipes	8090	Chromosome	734,057,086	25	31,218,526	516	2,530,934	GCF_002234675.1
Oryzias melastigma	30732	Scaffold	779,469,774	8,603	23,737,187	56,275	30,057	GCF_002922805.1
Osmerus eperlanus	29151	Scaffold	342,758,722	73,274	6,820	99,348	4,524	GCA_900302275.1
Oxygymnocypris stewartii	361644	Scaffold	1,849,224,471	26,281	257,093	26,283	257,093	GCA_003573665.1
Pachypanchax playfairii	52664	Scaffold	669,774,067	4,488	3,173,794	19,608	76,641	GCA_006937955.1
Pagrus major	143350	Scaffold	875,465,402	886,260	4,644	1,164,424	2,822	GCA_002897255.1
Pampus argenteus	206143	Scaffold	350,448,509	298,139	1,586	532,813	1,001	GCA_000697985.1
Pangasianodon hypophthalmus	310915	Scaffold	715,760,110	567	14,288,580	23,339	62,522	GCF_003671635.1
Parablennius parvicornis	171872	Scaffold	599,249,148	66,539	16,796	129,811	7,343	GCA_900302745.1
Paralichthys olivaceus	8255	Chromosome	545,775,252	7,202	3,817,360	38,614	30,544	GCA_001904815.2
Parambassis ranga	210632	Chromosome	551,012,959	156	22,993,012	1,677	5,080,925	GCF_900634625.1
Paramormyrops kingsleyae	1676925	Scaffold	799,421,083	4,667	1,731,158	47,999	37,656	GCF_002872115.1
Parasudis fraserbrunneri	1784818	Scaffold	707,987,062	156,389	6,391	213,462	4,177	GCA_900302295.1
Perca flavescens	8167	Chromosome	877,456,336	268	37,412,490	1,097	4,268,950	GCF_004354835.1
Perca fluviatilis	8168	Scaffold	958,225,486	31,105	6,260,519	100,821	18,196	GCA_003412525.1
Percopsis transmontana	143327	Scaffold	458,089,168	53,197	15,180	88,141	8,161	GCA_900302285.1
Periophthalmodon schlosseri	1365757	Scaffold	679,761,122	46,662	39,308	85,749	16,946	GCA_000787095.1
Periophthalmus magnuspinnatus	409849	Scaffold	701,696,780	26,060	296,161	76,770	28,254	GCA_000787105.1
Phycis blennoides	163115	Scaffold	416,766,999	62,684	10,640	132,164	4,532	GCA_900302315.1
Phycis phycis	349666	Scaffold	346,335,180	100,771	4,502	120,694	3,458	GCA_900302335.1
Pimephales promelas	90988	Scaffold	1,219,326,373	73,057	60,380	215,176	7,468	GCA_000700825.1
Planiliza	370040	Contig	747,342,729			1,453	3,973,280	GCA_005024645.1

haematocheila								
Poecilia formosa	48698	Scaffold	748,923,461	3,985	1,574,226	31,058	57,472	GCF_000485575.1
Poecilia latipinna	48699	Scaffold	815,144,743	17,988	279,200	54,625	33,278	GCF_001443285.1
Poecilia mexicana	48701	Scaffold	801,711,499	18,105	275,316	50,601	39,840	GCF_001443325.1
Poecilia reticulata	8081	Chromosome	731,622,281	3,029	5,270,359	40,144	41,908	GCF_000633615.1
Pollachius virens	8060	Scaffold	394,927,939	116,705	4,344	137,332	3,457	GCA_900312635.1
Polymixia japonica	81385	Scaffold	554,895,936	92,198	9,571	134,725	5,803	GCA_900302305.1
Poropuntius huangchuchieni	357532	Scaffold	760,177,161	625,277	2,931	821,804	2,273	GCA_004124795.1
Pseudochromis fuscus	280673	Scaffold	657,041,210	52,042	24,689	91,797	12,029	GCA_900323345.1
Pseudopleuronect es yokohamae	245875	Contig	547,831,023			525,502	1,994	GCA_000787555.1
Pungitius pungitius	134920	Scaffold	441,089,565	7,847	302,682	49,352	14,136	GCA_003399555.1
Pygocentrus nattereri	42514	Scaffold	1,285,352,492	283,518	1,440,044	325,620	57,732	GCF_001682695.1
Regalecus glesne	81389	Scaffold	656,003,707	105,196	9,773	141,447	6,781	GCA_900302585.1
Reinhardtius hippoglossoides	111784	Scaffold	677,540,803	4,453	17,640,195	5,488	775,256	GCA_006182925.2
Rhamphochromis esox	163638	Scaffold	71,295,074	55,751	1,324	78,130	1,126	GCA_000150935.1
Rondeletia loricata	88713	Scaffold	568,597,941	103,827	7,469	140,955	5,112	GCA_900302605.1
Salarias fasciatus	181472	Chromosome	797,507,141	203	32,729,575	805	2,597,836	GCF_902148845.1
Salmo salar	8030	Chromosome	2,966,890,203	241,573	1,366,254	368,060	57,618	GCF_000233375.1
Salmo trutta	8032	Chromosome	2,371,880,186	1,441	52,209,666	5,378	1,703,178	GCF_901001165.1
Salvelinus alpinus	8036	Chromosome	2,169,553,147	16,702	1,018,695	97,014	55,619	GCF_002910315.2
Sander lucioperca	283035	Scaffold	900,461,225	1,312	4,929,547	1,347	4,695,595	GCA_008315115.1
Sardina pilchardus	27697	Scaffold	949,617,276	117,259	96,617	194,510	9,398	GCA_900499035.1
Scartelaos histophorus	166764	Scaffold	695,008,792	156,044	15,105	209,353	8,806	GCA_000787155.1
Scleropages formosus	113540	Chromosome	784,563,014	72	31,084,684	217	9,102,216	GCF_900964775.1
Scophthalmus maximus	52904	Chromosome	524,979,463	22	24,811,384	21,326	54,836	GCA_003186165.1
Sebastes aleutianus	214485	Scaffold	899,650,391	10,489	340,062	110,635	10,838	GCA_001910805.2
Sebastes koreanus	290523	Contig	725,092,264			147,157	16,662	GCA_004335335.1
Sebastes minor	214483	Scaffold	681,652,711	166,448	7,676	812,852	1,901	GCA_001910765.2
Sebastes nigrocinctus	72089	Scaffold	746,044,620	15,872	116,274	89,356	13,471	GCA_000475235.3

Sebastes norvegicus	394699	Scaffold	717,740,616	75,627	16,564	117,709	9,467	GCA_900302655.1
Sebastes nudus	1617787	Contig	724,045,237			180,312	11,290	GCA_004335365.1
Sebastes rubrivinctus	72099	Scaffold	756,296,653	68,206	30,046	136,109	13,541	GCA_000475215.1
Sebastes schlegelii	214486	Contig	728,476,695			146,105	14,246	GCA_004335315.1
Sebastes steindachneri	201708	Scaffold	648,011,071	279,232	4,288	1,089,366	1,311	GCA_001910785.2
Selene dorsalis	179366	Scaffold	528,779,420	36,113	32,464	85,660	11,209	GCA_900303245.1
Seriola dumerili	41447	Scaffold	677,686,174	34,656	5,812,906	41,188	249,509	GCF_002260705.1
Seriola lalandi	302047	Scaffold	766,364,468	7,606	411,616	63,278	37,711	GCA_003054885.1
Seriola quinqueradiata	8161	Scaffold	639,269,536	384	5,610,255	1,312	872,227	GCA_002217815.1
Seriola rivoliana	173321	Scaffold	666,141,578	1,343	9,509,606	3,939	740,108	GCA_002994505.1
Simochromis diagramma	43689	Scaffold	848,827,444	823	8,960,300	1,764	2,231,376	GCA_900408965.1
Sinocyclocheilus anshuiensis	1608454	Scaffold	1,632,718,266	85,682	1,284,143	254,423	17,271	GCF_001515605.1
Sinocyclocheilus grahami	75366	Scaffold	1,750,287,761	31,277	1,156,368	168,074	29,353	GCF_001515645.1
Sinocyclocheilus rhinocerous	307959	Scaffold	1,655,786,410	164,173	945,738	314,963	18,758	GCF_001515625.1
Sparus aurata	8175	Chromosome	833,595,063	176	35,791,275	1,224	2,862,625	GCF_900880675.1
Sphaeramia orbicularis	375764	Chromosome	1,342,662,642	340	57,165,184	2,183	2,360,121	GCF_902148855.1
Spondyliosoma cantharus	50595	Scaffold	680,472,139	47,064	28,198	97,735	11,633	GCA_900302685.1
Stegastes partitus	144197	Scaffold	800,491,834	5,818	411,659	42,060	43,010	GCF_000690725.1
Stylephorus chordatus	409996	Scaffold	488,488,587	128,468	4,684	170,584	3,373	GCA_900312615.1
Symphodus melops	171736	Scaffold	533,823,763	50,156	21,275	95,080	9,362	GCA_900323315.1
Syngnathus acus	161584	Chromosome	324,331,233	87	14,974,571	130	11,959,915	GCA_901709675.1
Tachysurus fulvidraco	1234273	Scaffold	713,810,725	663	3,653,474	2,402	980,445	GCF_003724035.1
Takifugu bimaculatus	433685	Chromosome	371,675,663	22	16,786,025	1,055	1,398,332	GCA_004026145.1
Takifugu flavidus	433684	Chromosome	366,286,831	867	15,676,631	1,111	4,357,567	GCA_003711565.2
Takifugu rubripes	31033	Chromosome	384,126,662	128	16,705,553	530	3,136,617	GCF_901000725.2
Tenualosa ilisha	373995	Scaffold	815,647,530	124,209	188,026	131,117	129,889	GCA_003651195.1
Thalassoma bifasciatum	76338	Scaffold	1,095,910,316	379,332	155,821	397,893	122,955	GCA_008086565.1
Thunnus	8236	Scaffold	728,212,003	38,995	46,920	84,919	16,808	GCA_900302625.1

albacares								
Thunnus orientalis	8238	Contig	684,497,465			133,062	8,235	GCA_000418415.1
Thunnus thynnus	8237	Scaffold	648,208,697	354,425	3,045	450,338	2,430	GCA_003231725.1
Thymallus thymallus	36185	Chromosome	1,564,834,359	3,831	32,985,317	204,386	31,774	GCA_004348285.1
Trachinotus ovatus	173339	Scaffold	648,062,395	138	29,494,812	749	1,846,793	GCA_900607315.1
Trachyrincus murrayi	241836	Scaffold	452,416,606	40,927	19,998	114,476	6,231	GCA_900323305.1
Trachyrincus scabrus	562814	Scaffold	369,861,760	80,958	6,379	119,164	3,900	GCA_900303215.1
Triplophysa siluroides	422203	Scaffold	583,428,323	1,002	2,872,994	1,039	2,549,348	GCA_006030095.1
Trisopterus minutus	80722	Scaffold	334,717,091	106,116	3,976	122,084	3,248	GCA_900302415.1
Typhlichthys subterraneus	940470	Scaffold	555,559,596	84,841	9,654	106,331	7,314	GCA_900302405.1
Xiphophorus couchianus	32473	Chromosome	688,541,509	68	30,550,352	297	15,315,838	GCF_001444195.1
Xiphophorus hellerii	8084	Chromosome	733,126,988	85	26,465,357	561	7,119,835	GCA_003331165.1
Zeus faber	64108	Scaffold	610,433,400	135,758	6,332	172,424	4,642	GCA_900323335.1

28

Supplementary Table 2. List of the 93 species sequenced in the Fish10K pilot study. An asterisk (*) denotes that a species could not be classified by morphological classification and need further evidence (i.e. genomics data) to be classified.

Species	stLFR (Gb)	Hi-C (Gb)	ATAC-seq (Gb)	RNA-seq (Gb)	Nanopore (Gb)
Mormyrus caschive Linnaeus	119.39	142.63	12.56	18.32	9.61
Campylomormyrus elephas	121.14	40.10	12.38	in prep.	in prep.
Semaprochilodus insignis	120.65	78.58	11.26	Fail	in prep.
Panaque nigrolineatus	118.51	in prep.	13.78	18.14	11.33
Glossolepis incisus	121.70	in prep.	10.80	19.56	9.80
Kuhlia marginata	114.67	71.48	12.84	in prep.	in prep.
Distichodus sexfasciatus	125.09	in prep.	1.88	17.58	14.12
Megalops atlanticus	117.61	in prep.	9.06	22.36	in prep.
*Nemacheilidae**	115.22	143.26	in prep.	in prep.	in prep.
Helostoma temminckii	117.99	98.46	8.98	Fail	11.11
Asterophysus batrachus	319.01	in prep.	in prep.	in prep.	in prep.
*Callichthyidae**	114.60	in prep.	in prep.	in prep.	in prep.
Syin prep.dontis nigriventris	114.76	in prep.	in prep.	in prep.	in prep.
Sahyadria denisonii	116.50	in prep.	in prep.	in prep.	in prep.
Iriatherina werneri	123.63	in prep.	in prep.	in prep.	in prep.
Pantodon buchholzi	99.81	in prep.	in prep.	in prep.	in prep.
Boulengerella lucia	122.82	68.36	11.12	22.80	in prep.
Crossocheilus oblongus	122.12	in prep.	in prep.	in prep.	in prep.
Kryptopterus bicirrhis	123.83	in prep.	in prep.	in prep.	in prep.
Acestrorhynchus altus	125.46	sequencing	6.50	19.40	12.06
Moin prep.tremus palembangensis	121.83	82.08	17.54	18.36	in prep.
Tetraodon fluviatilis	122.42	sequencing	10.12	19.32	in prep.
Hydrocynus vittatus	92.93	56.64	5.14	21.30	12.75
Tetraodon palembangensis	120.40	19.00	11.12	20.64	12.32

Malapterurus electricus	118.87	124.02	8.96	16.22	9.47
Naso vlamingii	131.47	sequencing	in prep.	18.86	12.13
Centropyge bicolor	130.47	134.61	in prep.	in prep.	in prep.
Synchiropus splendidus	130.25	sequencing	in prep.	12.44	in prep.
Corythoichthys haematopterus	364.22	in prep.	in prep.	in prep.	in prep.
Diodon holocanthus	127.52	sequencing	in prep.	22.86	11.66
Chelmon rostratus	134.64	134.22	in prep.	in prep.	in prep.
Auchein prep.glanis occidentalis	128.79	19.24	5.82	17.84	10.15
Heterotis niloticus	119.30	21.24	10.06	23.02	10.46
Crenicichla johanna	120.14	136.03	8.56	19.88	10.79
Hydrolycus armatus	120.76	54.46	7.80	18.48	in prep.
Macrochirichthys macrochirius	114.39	66.38	9.72	21.38	10.27
Hepsetus odoe	118.61	66.86	2.68	22.26	11.73
Salminus brasiliensis	120.54	73.60	10.96	25.50	12.05
Metynnis lippincottianus	341.02	in prep.	in prep.	in prep.	12.33
Polypterus endlicheri	336.58	in prep.	in prep.	in prep.	in prep.
Brachyplatystoma tigrinum	94.44	in prep.	in prep.	in prep.	in prep.
Syin prep.dontis flavitaeniatus	73.88	in prep.	in prep.	in prep.	in prep.
Hexanematichthys leptaspis	85.30	in prep.	in prep.	in prep.	in prep.
Leporinus frederici	302.09	in prep.	in prep.	in prep.	in prep.
Garra orientalis	99.14	in prep.	in prep.	in prep.	in prep.
*Gyriin prep.cheilus**	98.65	137.06	8.36	22.88	11.19
Aplocheilus lineatus	97.10	in prep.	in prep.	in prep.	in prep.
Phenacogrammus interruptus	104.41	in prep.	in prep.	in prep.	in prep.
Apteroin prep.tus albifrons	108.95	in prep.	in prep.	in prep.	10.00
Hoplias aimara	95.89	in prep.	in prep.	in prep.	in prep.
*Lethrinus**	126.19	in prep.	in prep.	in prep.	11.38
Therapon oxyrhynchus	124.76	in prep.	in prep.	in prep.	18.53
*Eleotridae**	125.74	in prep.	in prep.	in prep.	19.64

Datnioides pulcher	122.38	in prep.	in prep.	in prep.	11.24
Rhiin prep.gobius giurinus	133.18	in prep.	in prep.	in prep.	11.60
Chaetodon trifasciatus	128.29	in prep.	in prep.	in prep.	10.32
Cobitis sinensis	130.76	in prep.	in prep.	in prep.	in prep.
Eleotris oxycephala	137.64	in prep.	in prep.	in prep.	in prep.
Harpadon nehereus	131.32	in prep.	in prep.	in prep.	in prep.
Sillago japonica	129.92	in prep.	in prep.	in prep.	in prep.
Brachirus orientalis	137.34	in prep.	in prep.	in prep.	in prep.
Apogonidae*	131.97	in prep.	in prep.	in prep.	in prep.
Triacanthus biaculeatus	134.93	in prep.	in prep.	in prep.	in prep.
Anabas testudineus	130.97	in prep.	in prep.	in prep.	in prep.
Lutjanus fulviflamma	134.98	in prep.	in prep.	in prep.	in prep.
Bostrychus sinensis	135.54	in prep.	in prep.	in prep.	in prep.
Urain prep.scopiformes	134.82	in prep.	in prep.	in prep.	in prep.
Arothron hispidus	133.51	in prep.	in prep.	in prep.	in prep.
Abudefduf vaigiensis	132.93	in prep.	in prep.	in prep.	in prep.
Gymin prep.thorax	140.25	in prep.	in prep.	in prep.	in prep.
Moin prep.pterus albus	136.21	in prep.	in prep.	in prep.	in prep.
Eleotris oxycephala	125.41	in prep.	in prep.	in prep.	in prep.
Sillaginidae*	115.00	in prep.	in prep.	in prep.	in prep.
Gerres acinaces	118.53	in prep.	in prep.	in prep.	in prep.
Platycephalidae	131.16	in prep.	in prep.	in prep.	in prep.
Sillaginidae*	122.42	in prep.	in prep.	in prep.	in prep.
Inimicus japonicus	126.58	in prep.	in prep.	in prep.	in prep.
Platycephalidae	118.44	in prep.	in prep.	in prep.	in prep.
Johnius grypotus	120.79	in prep.	in prep.	in prep.	in prep.
Gobiidae*	131.75	in prep.	in prep.	in prep.	in prep.
Pisodoin prep.phis cancrivorus	126.37	in prep.	in prep.	in prep.	in prep.
Leiognathus brevirostris	121.58	in prep.	in prep.	in prep.	in prep.

Hemibarbus labeo	130.99	in prep.	in prep.	in prep.	in prep.
Zebrasoma veliferum	139.93	in prep.	in prep.	in prep.	in prep.
Siganus unimaculatus	139.93	in prep.	in prep.	in prep.	in prep.
Chaetodon auriga	138.54	in prep.	in prep.	in prep.	in prep.
Carassioides cantonensis	121.21	in prep.	in prep.	in prep.	in prep.
Varicorhinus lepturus	119.60	in prep.	in prep.	in prep.	in prep.
Carassioides cantonensis	124.74	in prep.	in prep.	in prep.	in prep.
Balitora brucei	96.75	in prep.	in prep.	in prep.	in prep.
Opsariichthys bidens	141.76	in prep.	in prep.	in prep.	in prep.
Chain prep.dichthys erythropterus	141.25	in prep.	in prep.	in prep.	in prep.

Supplementary Table 3. Assembly statistics of other 38 fish genome assemblies.
An asterisk (*) denotes cartilaginous fish. A scaffold is a series of contigs that can be linked. N50 is the sequence length of the shortest contig (or contig) at 50% of the total genome length. BUSCO (Benchmarking Universal Single-Copy Orthologs) is a standard qualitative measure for assessing genome quality and completeness based on evolutionarily informed expectations of gene contents from near-universal single-copy orthologs. The highest BUSCO score is 100%. A higher BUSCO score indicates a better quality of the annotated genome.

Strategy	Species	Assembly size (bp)	Scaffold N50 (bp)	Contig N50 (bp)	BUSCO
I	*Raja porosa**	3,047,746,089	2,182,378	16,150	---
	Acestrorhynchus altus	955,799,133	9,632,287	48,769	94.0%
	Monotremus palembangensis	359,992,882	5,706,744	213,105	96.3%
	Tetraodon fluviatilis	326,757,569	7,010,509	47,398	96.0%
	Hydrocynus vittatus	1,007,380,147	3,263,334	25,092	88.7%
	Tetraodon palembangensis	359,651,202	7,038,803	218,260	96.2%
	Malapterurus electricus	798,763,704	3,588,439	48,402	90.3%
	Centropyge bicolor	681,516,296	4,420,410	115,524	96.2%
	Synchiropus splendidus	535,219,047	7,262,416	117,146	96.3%
	Auchenoglanis occidentalis	767,889,523	6,633,917	110,792	93.5%
	Crenicichla johanna	826,044,049	2,355,403	60,889	96.6%
	Hydrolycus armatus	806,097,948	3,692,170	20,392	86.0%
	Macrochirichthys macrochirius	835,452,375	6,246,214	56,429	87.8%
	Hepsetus odoe	860,429,938	5,143,489	43,947	93.8%
	Salminus brasiliensis	942,321,778	2,364,359	39,978	89.3%
	Metynnis lippincottianus	1,347,978,869	1,154,248	21,525	89.1%
	Therapon oxyrhynchus	724,925,999	11,684,578	50,940	90.5%
	Rhinogobius giurinus	973,105,242	4,063,184	25,681	84.8%

	Mormyrus caschive Linnaeus	819,947,615	1,532,218	48,123	89.6%
	Megalops atlanticus	1,033,524,629	7,259,301	65,702	95.4%
	Kuhlia marginata	674,579,433	4,757,558	40,209	92.8%
	Semaprochilodus insignis	1,178,289,156	1,780,909	29,765	80.9%
	Distichodus sexfasciatus	814,422,444	1,048,871	31,524	89.4%
	Glossolepis incisus	889,156,884	1,422,316	32,279	91.0%
	Campylomormyrus elephas	905,010,418	1,613,581	30,544	92.5%
II	*Parabramis pekinensis*	984,227,144	---	5,486,023	95.90
	Pseudolaubuca sinensis	806,127,291	---	11,561,046	96.00
	Redfin culter	1,102,622,456	---	2,126,716	94.60
	Toxabramis swinhonis	1,038,263,990	---	1,754,830	93.70
	Squaliobarbus curriculus	1,066,071,259	---	1,664,874	95.40
	Ochetobibus elongatus	929,258,157	---	3,217,609	95.90
	Opsariichthys bidens	841,061,991	---	5,688,910	94.60
	Nicholsicypris normalis	854,847,741	---	2,288,299	95.50
	Yaoshanicus arcus	835,190,039	---	5,992,616	96.40
	Xenocypris argentea	1,033,647,724	---	5,841,742	96.30
	Gymnocypris scolistomus	969,320,259	---	1,907,468	92.00

Supplementary Table 4. Statistics of the existing fish species. Based on data from FishBase web site (accessed 10 September 2019) and "Fishes of the world" (5th ed.). Actinopterygii belong to Osteichthyes, whose number of species accounts for most all fishes.

Type	Class	Order	Family	Species
Chondrichthyes	Elasmobranchii	12	51	1,190
	Holocephali	1	3	56
Osteichthyes	Actinopterygii	67	475	32,869
Total		80	529	34,115

Supplementary Table 5. An Excel file with all the species list of Phase I.