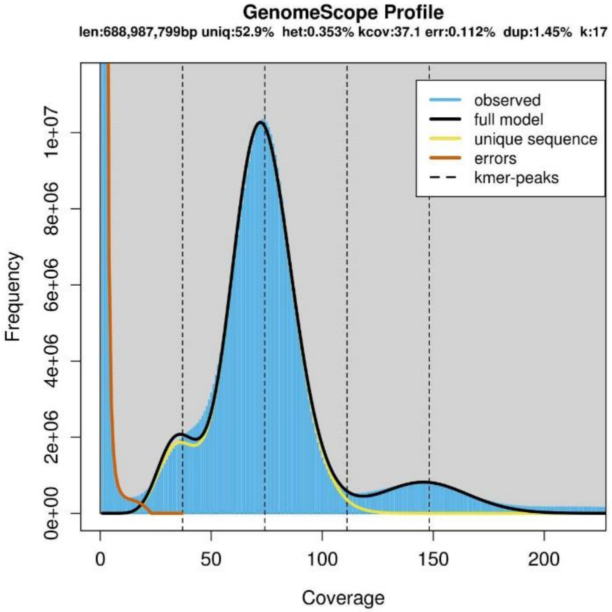
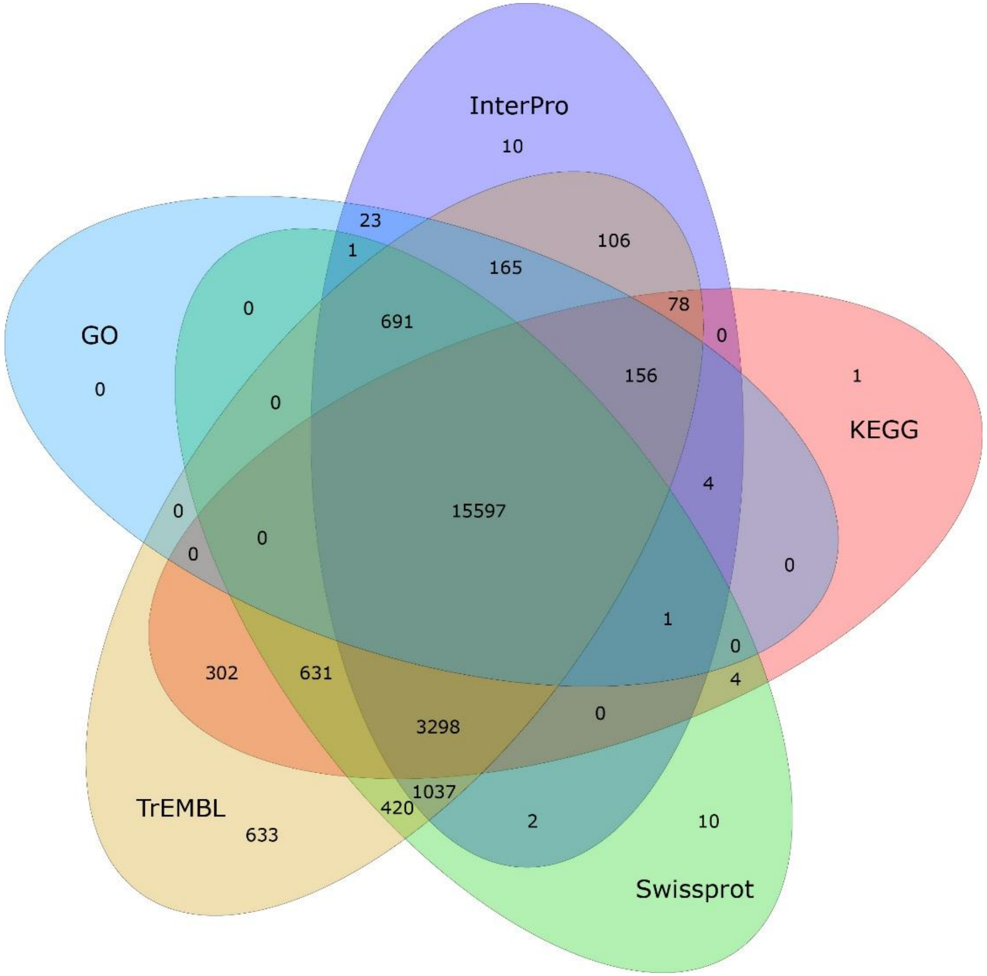
Supplementary Materials

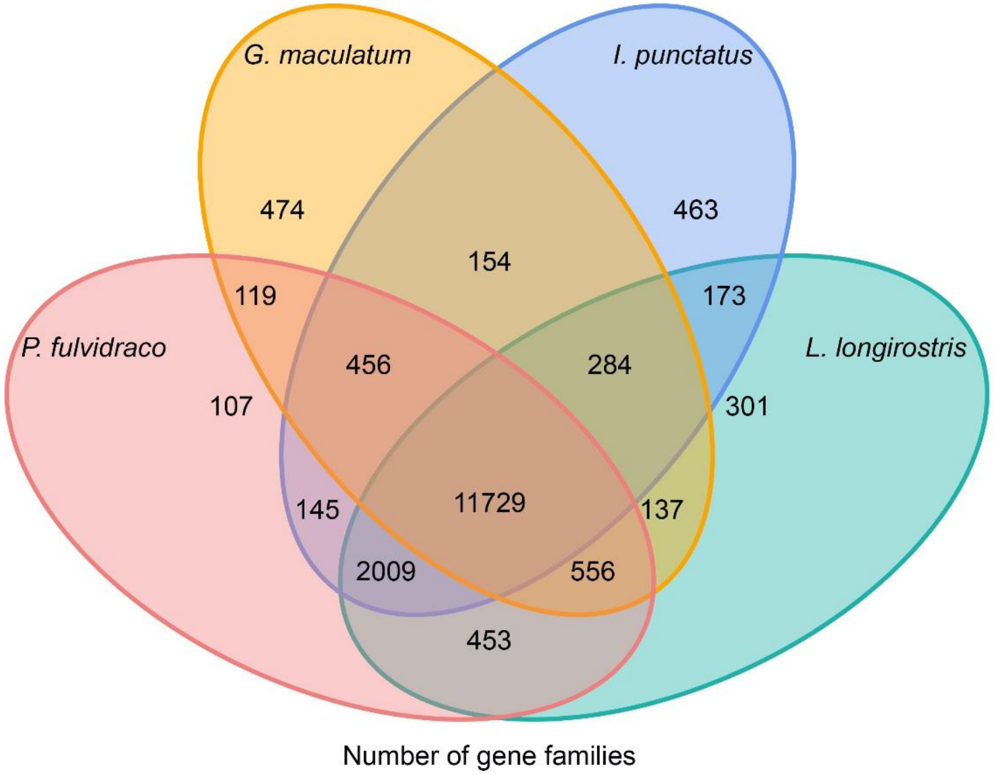
Supplementary Figures



Supplementary Figure S1 *K*-mer frequency distribution of *L. longirostris*



Supplementary Figure S2 Venn diagram of gene function annotation



Supplementary Figure S3 Venn diagram of orthologous genes

Supplementary Tables

Supplementary Table S1 Sequencing data for *L. longirostris* genome assembly

Types	**Sequencing strategy**	**Insert size (bp)**	**Raw data (Gb)**	**Clean data (Gb)**	**Mean read length (bp)**	**Sequence coverage (X)**
Genome	BGI-SEQ500	200-300	76.93	64.11	100	93.05
	Nanopore	-	43.23	-	24,300	61.48
	Hi-C	-	243.13	126.35	100	179.63
Transcriptome	Illumina	250-300	9.83	9.65	150	-

Supplementary Table S2 Assembly statistics of *L. longirostris* genome

Genome assembly	Contig Length (bp)	Contig number
Total	703,192,559	389
Max	16,905,276	
Number≥2kb		
N50	4,293,955	
N60		
N70		
N80		
N90	998,020	

Supplementary Table S3 Summary of 26 assembled chromosomes in *L. longirostris* genome

Chromosomes	**Length (Mbp)**	**Percentage**
Chr1	43.97	6.25
Chr2	40.42	5.75
Chr3	36.30	5.16
Chr4	33.36	4.74
Chr5	33.31	4.74
Chr6	32.52	4.62
Chr7	31.99	4.55
Chr8	30.13	4.28
Chr9	29.33	4.17
Chr10	29.08	4.13
Chr11	28.03	3.98
Chr12	27.85	3.96
Chr13	25.36	3.61
Chr14	24.16	3.43
Chr15	24.11	3.43
Chr16	23.71	3.37
Chr17	21.98	3.12
Chr18	21.32	3.03
Chr19	20.55	2.92
Chr20	19.22	2.73
Chr21	18.90	2.69
Chr22	18.86	2.68
Chr23	18.31	2.60
Chr24	17.81	2.53
Chr25	17.40	2.47
Chr26	17.36	2.47
Linked Total	685.35	97.44
Unlinked Total	18.03	2.56

Supplementary Table S4-1 Classification of repeat elements in *L. longirostris* genome

Type	Repbase TEs		*De novo* TEs		TE protiens		Combined TEs	
	Length (Mb)	% in genome	Length (Mb)	% in genome	Length (Mb)	% in genome	Length (Mb)	% in genome
DNA	80.23	11.41	107.28	15.25	948.20	0.13	146.40	20.81
LINE	30.71	4.37	64.31	9.14	19.22	2.73	82.19	11.68
SINE	15.34	2.18	6.28	0.89	0	0.00	20.52	2.92
LTR	19.17	2.73	68.24	9.7	8.42	1.2	78.65	11.18
Others	0.03	0.00	0	0.00	0	0.00	0.03	0.00
Unknown	0	0.00	1.96	0.28	0	0.00	1.96	0.28
Total	129.07	18.35	214.87	30.55	28.53	4.06	239.11	33.99

Supplementary Table S4-2 Proportion of repetitive elements in partial teleost genomes

Species	% in genome	References
Leiocassis longirostris	33.99	This study
Danio rerio	52.2	Howe et al., 2013
Onychostoma macrolepis	46.23	Sun et al., 2020
Bagarius yarrelli	35.26	Jiang et al., 2019
Glyptosternon maculatum	33.96	Liu et al., 2018
Ictalurus punctatus	32.56	Chen et al., 2016
Cyprinus carpio	31.3	Xu et al., 2014
Oryzias latipes	30.68	Kasahara et al., 2007
Larimichthys crocea	26.13	Chen et al., 2019
Gasterosteus aculeatus	13.48	Jones et al., 2012
Fugu rubripes	7.1	Aparicio et al., 2002
Tetraodon nigroviridis	5.7	Van de Peer, 2004

REFERENCES

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., … Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, *496*(7446), 498–503. doi: 10.1038/nature12111

Sun, L., Gao, T., Wang, F., Qin, Z., Yan, L., Tao, W., … Wang, D. (2020). Chromosome-level genome assembly of a cyprinid fish Onychostoma macrolepis by integration of nanopore sequencing, Bionano and Hi-C technology. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.13190

Jiang, W., Lv, Y., Cheng, L., Yang, K., Bian, C., Wang, X., … Shi, Q. (2019). Whole-Genome Sequencing of the Giant Devil Catfish, Bagarius yarrelli. *Genome Biology and Evolution*, *11*(8), 2071–2077. doi: 10.1093/gbe/evz143

Liu, H., Liu, Q., Chen, Z., Liu, Y., Zhou, C., Liang, Q., … Mou, Z. (2018). Draft genome of Glyptosternon maculatum, an endemic fish from Tibet Plateau. *Gigascience*, *7*(9). doi: 10.1093/gigascience/giy104

Chen, X., Zhong, L., Bian, C., Xu, P., Qiu, Y., You, X., … Bian, W. (2016). High-quality genome assembly of channel catfish, Ictalurus punctatus. *Gigascience*, *5*, 39. doi:

10.1186/s13742-016-0142-5

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., … Sun, X. (2014). Genome sequence and genetic diversity of the common carp, Cyprinus carpio. *Nature Genetics*, *46*(11), 1212–1219. doi: 10.1038/ng.3098

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., … Kohara, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature*, *447*(7145), 714–719. doi: 10.1038/nature05846

Chen, B., Zhou, Z., Ke, Q., Wu, Y., Bai, H., Pu, F., & Xu, P. (2019). The sequencing and de novo assembly of the Larimichthys crocea genome using PacBio and Hi-C technologies. *Scientific Data*, *6*, 188. doi: 10.1038/s41597-019-0194-3

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. doi: 10.1038/nature10944

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J., Dehal, P., … Brenner, S. (2002). Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science*, *297*(5585), 1301–1310. doi: 10.1126/science.1072104

Van de Peer, Y. (2004). Tetraodon genome confirms Takifugu findings: Most fish are ancient polyploids. *Genome Biology*, *5*(12), 250. doi: 10.1186/gb-2004-5-12-250

Supplementary Table S5 Statistics of predicted protein-coding genes in *L. longirostris* genome

Method		**Gene numbers**	**Average gene length (bp)**	**Average CDS length (bp)**	**Average intron length (bp)**	**Average exon length (bp)**	**Average exon per gene**
Homolog	***D. rerio***	22,607	24,906.24	1,586.43	2,991.09	180.35	8.80
	G. aculeatus	19,515	14,591.05	1,522.77	1,622.74	168.20	9.05
	I. punetaus	23,916	19,607.43	1,698.77	2,068.38	175.89	9.66
	L. crocea	19,859	21,505.49	1,685.09	2,269.19	173.10	9.73
	O. niloticus	19,427	25,587.29	1,664.02	2,771.77	172.78	9.63
	O. latipe	19,122	58,907.43	1,603.53	7,358.53	182.48	8.79
	P. hypophthalmus	22,691	20,471.44	1,737.03	2,119.68	176.56	9.84
	T. fulvidraco	25,662	19,646.64	1,676.25	2,081.63	174.01	9.63
	T. rubripes	18,205	16,156.31	1,440.29	1,964.65	169.64	8.49
De novo	**Augustus**	23,758	14,112.19	1,448.53	1,734.99	174.54	8.30
Transcript	**PASA**	51,771					
Merge	**Glean**	23,708	16,546.44	1,792.67	1,547.65	170.20	10.53

Supplementary Table S6 Statistics of gene function annotation in *L. longirostris* genome

	Number	Percent (%)
Total	23,708	100
SwissProt	21,692	91.50
KEGG	20,072	84.66
TrEMBL	23,114	97.49
InterPro	21,169	89.29
GO	16,638	70.18
Annotated	23,170	97.73
Unannotated	538	2.27

Supplementary Table S7 Statistics of noncoding RNAs in *L. longirostris* genome

Type		Copy	Average length(bp)	Total length(bp)	Percentage in Genome (%)
miRNA		422	85.1	35,939	0.005109
tRNA		2,118	75.07	159,004	0.022605
rRNA	**rRNA**	919	110.83	101,867	0.014481
	18s	14	429.79	6,017	0.000855
	28s	60	184.3	11,058	0.001572
	5.8s	3	132	396	0.000056
	5S	842	100.22	84,386	0.011997
snRNA	**snRNA**	970	139.93	135,733	0.019297
	CD-box	135	122.78	16,575	0.002356
	HACA-box	57	154.98	8,834	0.001256
	splicing	763	142.23	108,519	0.015428

Supplementary Table S8 Statistics of single-copy ortholog families

Species	**Genes number**	**Unclustered genes**	**Family number**	**Unique gene families**	**Average genes per family**
L. longirostris	23,708	845	15,642	68	1.46
D. rerio	24,285	616	15,307	64	1.55
A. mexicanus	25,679	690	15,926	102	1.57
G. aculeatus	20,801	1,403	13,897	34	1.40
G. maculatum	21,755	3,173	13,909	127	1.34
I. punctatus	22,722	400	15,413	32	1.45
L. oculatus	18,199	1,213	15,760	176	1.08
O. niloticus	21,377	196	14,155	12	1.50
O. latipes	22,564	567	14,743	60	1.49
P. fulvidraco	23,258	246	15,574	23	1.48
T. rubripes	19,701	672	14,224	36	1.34