Supplementary information of Chromosome-level assembly of *Gymnocypris eckloni* genome

Table s1. Statistics of genome size estimation by 17-mer analysis.

Kmer	*Kmer* number	*Kmer* Depth	Genome size (Mb)	Heterozygosity	Repeat_rate(%)
17	169,021,371,761	181	927.13	1.82%	50.5

Table s2. Statistics of the chromosome assemblies using Hi-C data.

Pseudo-Chromosomes	Sequence IDs	Cluster Numbers	Sequences Lengths
Chr01	Hic_asm_0	170	44,161,858
Chr02	Hic_asm_1	20	25,669,045
Chr03	Hic_asm_2	102	34,715,927
Chr04	Hic_asm_3	125	61,913,669
Chr05	Hic_asm_4	6	28,665,261
Chr06	Hic_asm_5	66	43,543,958
Chr07	Hic_asm_6	154	61,756,641
Chr08	Hic_asm_7	118	47,622,867
Chr09	Hic_asm_8	62	27,783,661
Chr10	Hic_asm_9	69	32,896,108
Chr11	Hic_asm_10	82	21,739,851
Chr12	Hic_asm_11	29	32,200,481
Chr13	Hic_asm_12	42	29,129,546
Chr14	Hic_asm_13	56	28,882,651
Chr15	Hic_asm_14	232	70,525,535
Chr16	Hic_asm_15	195	89,391,071
Chr17	Hic_asm_16	193	16,898,605
Chr18	Hic_asm_17	54	15,913,822
Chr19	Hic_asm_18	68	33,789,651
Chr20	Hic_asm_19	190	46,494,833
Chr21	Hic_asm_20	100	32,874,627
Chr22	Hic_asm_21	96	39,527,105
Chr23	Hic_asm_22	101	39,007,358

Table s3. The annotation of repeated sequences in the *G. eckloni* genome using TRF, RepeatMasker and RepeatProteinMask. Note that the total content was merged and redundancy was eliminated by each method.

Type	Repeat Size(bp)	% of genome
TRF	59,024,407	6.42
RepeatMasker	420,920,632	45.82
RepeatProteinMask	63,100,708	6.87
Total	437,590,862	47.63

Table s4. Summary statistics of repeat annotation in *G. eckloni*. Note that *De novo* + Repbase represent the result of RepeatMasker based on Repbase, RepeatModeler, RepeatScout, and LTR_FINDER; TE proteins meant the result of RepeatProteinMask based on Repbase, and the Combined TEs refer to the combined results of *De novo* + Repbase and TE proteins.

Type	Denovo+Repbase		TE Proteins		Combined TEs	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA	43,441,654	4.73	4,297,129	0.47	44,941,543	4.89
LINE	5,251,877	1.66	30,834,954	3.36	36,383,537	3.96
SINE	2,370,900	0.26	0	0	2,370,900	0.26
LTR	355,341,965	38.68	27,985,849	3.05	356,794,021	38.84
Unknown	11,690,176	1.27	0	0	11,690,176	1.27
Total	420,920,632	45.82	63,100,708	6.87	425,012,609	46.26

Table s5. The number of the annotated non-coding RNA in the *G. eckloni* genome.

Type		Number	Average length (bp)	Total length (bp)	% of genome
miRNA		1,717	116.58	200,168	0.021789
tRNA		12,157	75.00	911,763	0.099247
rRNA	rRNA	1,780	178.65	317,991	0.034614
	18S	207	437.05	90,469	0.009848
	28S	282	346.36	97,674	0.010632
	5.8S	55	117.29	6,451	0.000702
	5S	1,236	99.84	123,397	0.013432
snRNA	snRNA	1,152	133.32	153,582	0.016718
	CD-box	172	120.66	20,754	0.002259
	HACA-box	107	142.94	15,295	0.001665
	splicing	826	132.76	109,663	0.011937
	scaRNA	35	179.94	6,298	0.000686
	Unknown	12	131	1,572	0.000171

Table s6. The comparison of the gene models annotated from the homolog species of _G. eckloni_.

Species	Gene number	Average length(bp)				Exons number per gene
		Transcript	CDS	Exon	Intron	
Gymnocypris eckloni	24,430	16,219.34	1,536.71	173.00	1,862.69	8.88
Cyprinus carpio	47,513	11,443.46	1,214.22	164.14	1,598.93	7.40
Astyanax mexicanus	22,631	20,200.93	1,574.58	162.81	2,148.02	9.67
Danio rerio	26,162	24,859.47	1,620.00	173.68	2,790.70	9.33
Takifugu rubripes	21,944	8,721.04	1,804.84	167.83	709.07	10.75
Ctenopharyngodon idellus	32,659	10,458.83	1,389.10	181.01	1,358.97	7.67
Oryzias latipes	19,146	11,969.78	1,504.25	148.03	1,142.33	10.16
Ictalurus punctatus	22,494	17,637.62	1,753.46	171.17	1,718.35	10.24

Table s7. The number of genes with functional classifications in _G. eckloni_.

Database	Number	Percent (%)
SwissProt	20,432	83.60
NR	23,110	94.60
KEGG	20,593	84.30
InterPro	21,539	88.20
GO	15,281	62.60
Pfam	19,290	79.00
Annotated	23,157	94.80
Unannotated	1,273	5.20
Total	24,430	94.80

Table s8. Gene families among the 14 species.

Species	Genes Number	Single-copy genes	Multiple-copy genes	Unique	Other	Uncluster
G. eckloni	24,428	1,396	4,358	576	1,4782	3,316
T. rubripes	21,391	1,442	4,190	368	14,274	1,117
O. latipes	22,107	1,443	4,286	298	14,858	1,222
C. carpio	23,755	1,299	4,704	338	13,559	3,855
A. mexicanus	18,684	1,580	3,569	231	12,238	1,066
C. idellus	32,659	1,300	4,942	891	17,421	8,105
D. rerio	25,357	1,270	4,957	865	17,167	1,098
I. punctatus	23,599	1,333	4,674	390	15,774	1,428
O. niloticus	28,149	1,347	4,815	1,294	19,416	1,277
H. sapiens	20,155	1,652	3,141	503	14,061	798
M. musculus	22,312	1,681	3,063	1,916	14,730	922
G. gallus	16,628	1,748	2,646	587	10,621	1,026
X. tropicalis	17,204	1,828	2,659	1,007	10,709	1,001
P. marinus	10,127	2,187	1,404	388	4,234	1,914

Fig. s1. A picture of *G. eckloni* used in the genome sequencing and assembly.