

Supplemental Information for

De novo assembly of a Chromosome-Level Reference Genome of Red Spotted

Grouper (*Epinephelus akaara*) using Nanopore sequencing and Hi-C

Hui Ge^{1,3#}, Keping Lin^{1#}, Mi Shen^{2#}, Shuiqing Wu¹, Yilei Wang^{3*}, Ziping Zhang⁶, Zhiyong

Wang³, Yong Zhang⁴, Zhen Huang⁵, Chen Zhou¹, Qi Lin¹, Jianshao Wu¹, Lei Liu², Jiang Hu^{2*},

Zhongchi Huang^{1*}, Leyun Zheng^{1*}

Table S1 Summary of clean reads generated from genome and transcriptome sequencing by Illumina Hiseq Xten

Application	Sample	Clean reads	Clean Bases(bp)	Q30
Genome Survey and Polish	D1703681-WY	362,233,076	54,697,194,476	0.93
RNA-Seq	Fbrain	110,890,780	16,596,641,393	0.97
	Gill	108,185,650	16,193,976,325	0.97
	Gonad	125,851,234	18,798,867,506	0.93
	Heart	104,714,712	15,661,477,121	0.97
	Kidney	126,727,978	18,971,576,481	0.97
	Liver	102,898,460	15,403,334,426	0.97
	Mbrain	97,238,278	14,553,759,266	0.97
	Muscle	106,207,434	15,895,544,144	0.97
	Spleen	113,457,912	16,987,591,796	0.97
Hi-C	D1703681-WY	769,598,162	112,826,974,259	0.95

Table S2 Estimation of genome size based on 17-mer statistics

K-mer Value	K-mer Number	K-mer Depth	Genome Size (Gb)	Heterozygosity Rate (%)
17	48,893,735,532	44	1.111	0.375

Table S3 Sequencing statistics from the Oxford Nanopore GridION X5 platform

Sample	Cells	Pass Reads Bases(Gb)	Pass Reads Mean Length(kb)	Pass Reads N50 Length(kb)	Pass Reads Max Length(kb)
19		106.29	18.35	26	184

Table S4 Comparisons of statistics among the well-constructed fish and *E. akaara* genome assembly

Species	Assembled genome size (Mb)	Contig Number	Contig N50 (Kb)	Scaffold Number	Scaffold N50 (Mb)
<i>Epinephelus akaara</i>	1,135	2,055	5,194	24	46.0
<i>Oreochromis niloticus</i>	1,003	2,960	3,325	-	-
<i>Amphiprion percula</i>	909	1,045	3,120	365	38.4
<i>Lates calcarifer</i>	587	3,917	1,066	24	25.8
<i>Amphiprion ocellaris</i>	881	7,810	324	6,404	0.4
<i>Monopterus albus</i>	635	117,579 (> 100 bp)	22	62,978 (> 100 bp)	2.1
<i>Acanthopagrus schlegelii</i>	688	115,091 (> 100 bp)	17	31,359 (> 100 bp)	7.6
<i>Gadus morhua</i> NEWB454	656	-	6	-	1.3
<i>Oryzias melastigma</i>	779	-	-	8,775	4.4

Table S5 Assessment of genome completeness by BUSCO

Type	Number	Percentage (%)
Complete BUSCOs (C)	4,437	96.8
Complete and single-copy BUSCOs (S)	4,271	93.2
Complete and duplicated BUSCOs (D)	166	3.6
Fragmented BUSCOs (F)	69	1.5
Missing BUSCOs (M)	78	1.7
Total BUSCO groups searched	4,584	-

Table S6 Summary of the transcriptomes and their mapping rate on the genome assembly

Sample	Total Reads Number	Mapped Reads Number	Mapping Rate (%)
Gill	108185650	97475941	90.1%
Spleen	113457912	104544451	92.1%
Kidney	126727978	116711449	92.1%
Muscle	106207434	100875022	95.0%
Fbrain	110890780	101239917	91.3%
Heart	104714712	86527230	82.6%
Mbrain	97238278	87390917	89.9%
Liver	102898460	96202059	93.5%
Gonad	125851234	119382396	94.9%
CD40B_L6_I034	64620162	55420144	85.8%
CD40G	53635500	48114193	89.7%
CD51B_L6_I036	58997712	50852786	86.2%
CD51G	49393822	44039822	89.2%

Table S7 Statistics of chromosome length distribution based on clustered scaffolds

Chr	Size	Scaffold Num
Chr01	53,736,897	54
Chr02	53,007,317	45
Chr03	52,524,049	97
Chr04	52,478,662	56
Chr05	50,330,081	61
Chr06	50,044,055	57
Chr07	49,122,225	58
Chr08	48,070,050	42
Chr09	46,462,825	39
Chr10	46,206,632	60
Chr11	46,067,422	47
Chr12	46,028,906	34
Chr13	45,677,884	70
Chr14	45,222,162	58
Chr15	45,116,380	52
Chr16	45,021,781	64
Chr17	44,981,524	59
Chr18	43,126,265	27
Chr19	41,713,854	40
Chr20	40,445,100	81
Chr21	39,359,364	40
Chr22	38,616,317	34
Chr23	35,838,671	31
Chr24	25,806,605	66
Total	1,085,005028	1,272

Table S8 Summary of the SSR search results

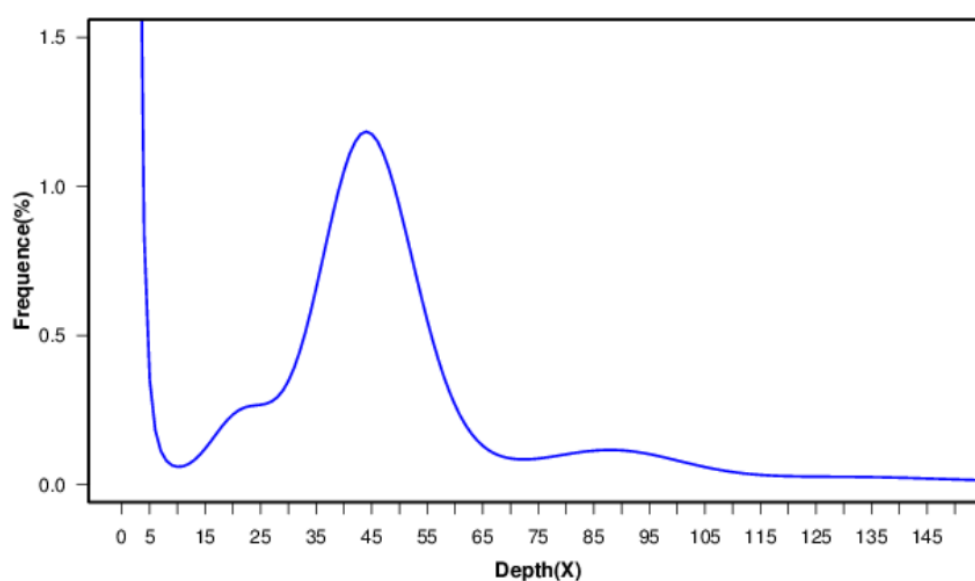
Repeat type	Unit size (repeat number)	Number
mono-nucleotide	1 (≥ 10)	190,558
Di-nucleotide	2 (≥ 6)	261,169
Tri-nucleotide	3 (≥ 5)	61,227
Tetra-nucleotide	4 (≥ 5)	25,893
Penta-nucleotide	5 (≥ 5)	5,335
Hexa-nucleotide	6 (≥ 5)	1,104
Total identified SSRs	-	545,286

Table S9 summary statistics of repeat annotation from different types

Type	Repeat Size(bp)	% of genome
Ltr_finder	64,043,800	5.64
Rebase	152,946,049	13.47
RepeatModeler	336,694,250	29.65
RepeatProteinMask	71,391,956	6.29
Trf	60,867,392	5.36
Total	488,505,111	43.02

Table S10 Summary statistics of gene sets compared with other species

Specie	Total number of gene	Average gene length(bp)	Average CDS length(bp)	Average exons number per gene	Average exon length(bp)	Average intron length(bp)
<i>E. akaara</i>	23,923	23,162.39	1,791.15	10.9	164.27	2,157.86
<i>C.idellus</i>	32,703	10,362.66	1,385.42	7.65	181.13	1,350.19
<i>C.semilaevis</i>	21,253	11,471.66	1,791.67	10.66	168.01	1,001.64
<i>D. rerio</i>	24,884	25,370.93	1,651.52	9.49	174.06	2,794.47
<i>G. aculeatus</i>	20,765	8,459.55	1,549.16	10.41	148.88	734.70
<i>L. calcarifer</i>	25,114	12,662.32	1,773.31	10.47	169.41	1,150.14
<i>O. niloticus</i>	28,892	14,577.62	1,722.03	9.55	180.24	1,502.88
<i>O. latipes</i>	19,686	12,154.68	1,516.24	10.26	147.77	1,148.79
<i>T. rubripes</i>	18,508	7,499.15	1,694.14	11.11	152.55	574.45

17-mer Depth Distribution Curve**Figure S1. Frequency distribution of the 17-mer graph analysis used to estimate the size of the *E. akaara* genome.**

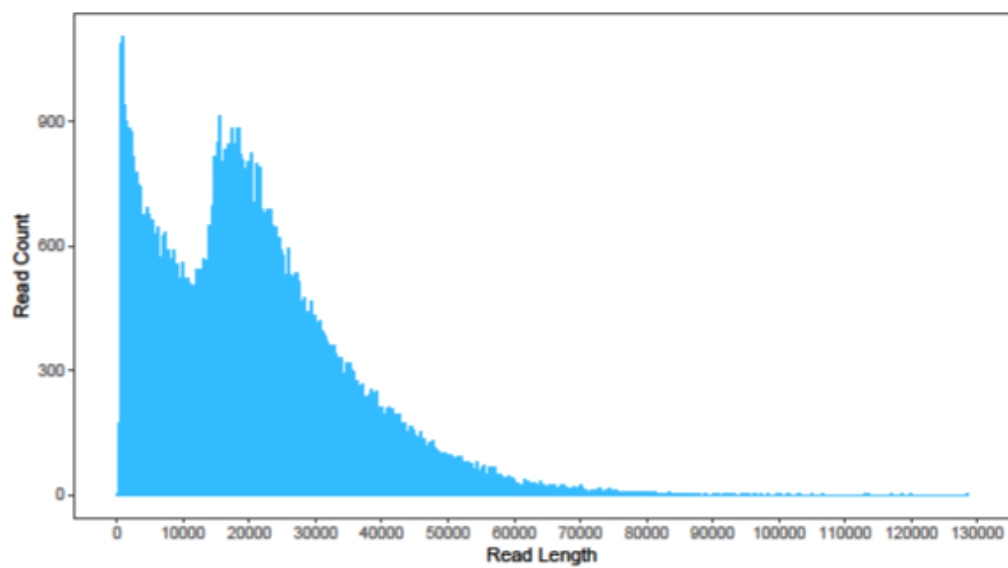


Figure S2. Passed read length distribution of Oxford Nanopore sequencing data