

Table 1 Statistics for the sequencing data of burbot genome

Type	Sequencing technology	Sequencing platform	Library size (bp)	Clean data (Gb)	Coverage*
Genome	illumina	Illumina NovaSeq-6000	150	79	143.64
Genome	PacBio	PacBio Sequel II	20,000	95.24	173.16
Genome	Hi-c	Illumina NovaSeq-6000	150	69.51	126.38
RNA	illumina	Illumina NovaSeq-6000	150	12.28G	/

*Coverage was calculated using an estimated genome size of 550Mb.

Table 2 Statistics of the burbot genome assembly before Hi-C correction

Mode	Total length (Mb)	Total number	max length(bp)	N50 (bp)	N90 (bp)
assembly	576.93	868	17,815,485	2,149,355	229,676
assembly+arrow	576.20	868	17,791,745	2,148,783	228,945
assembly+arrow+pilon	575.83	868	17,782,207	2,146,850	228,790

Table 3 BUSCO analysis result of the burbot genome

	Gene number	Percentage
Complete BUSCOs (C)	4344	94.76
Complete and single-copy BUSCOs (S)	4214	91.93
Complete and duplicated BUSCOs (D)	130	2.84
Fragment BUSCOs (F)	69	1.51
Missing BUSCOs (M)	171	3.73
Total BUSCOs groups searched	4584	100

Table 4 Statistics of the pseudochromosome assemblies using Hi-C data

Chr	Chr length (bp)	Contig number	Gene number
chr1	51,800,751	32	1909
chr2	25,925,363	27	1111
chr3	25,585,349	35	946
chr4	25,498,223	25	1050
chr5	24,788,504	33	863
chr6	24,176,936	19	1017
chr7	23,797,670	37	990
chr8	23,199,215	17	953
chr9	23,064,643	14	931
chr10	22,651,864	14	772
chr11	22,097,186	32	773
chr12	21,949,514	21	1083
chr13	21,705,196	34	862
chr14	21,607,725	23	983
chr15	21,604,800	19	762
chr16	20,600,739	18	825
chr17	20,263,900	40	661
chr18	19,816,954	27	823
chr19	19,424,836	14	756
chr20	18,715,696	21	597
chr21	17,145,549	21	759
chr22	15,180,800	19	610

Table 5 Statistics of the Hi-C assembly of the burbot genome

Genome size (Mb)	Percent assembled	Contig number	Contig N50 (Mb)	Scaffold number
575.92	94.83	868	2.1	287

Table 6 Statistics of repetitive sequences in the burbot genome.

Type	Repeat size(bp)	% of genome
Trf	89,079,979	15.47
Repeatmasker	106,671,932	18.52
Proteinmask	16,479,266	2.86
De novo	309,097,264	53.68
Total	384,291,034	66.74

Table 7 Statistics on transposable elements in burbot genome

	RepBase TEs		TE proteins		De novo		Combined TEs	
	Length (bp)	%in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA TE	73,063,918	12.69	4,437,703	0.77	245,833,034	42.69	289,324,656	50.24
LINE	14,751,054	2.56	3,457,054	0.60	17,837,406	3.10	30,964,639	5.38
SINE	1,173,013	0.20	0	0.00	1,888,174	0.33	2,970,446	0.52
LTR	29,625,915	5.14	8,586,143	1.49	45,057,876	7.82	66,948,775	11.63
Satellite	12,885,992	2.24	0	0.00	10,691,818	1.86	23,170,736	4.02
Simple_repeat	0	0.00	0	0.00	0	0.00	0	0.00
Other	3,520	0.00	0	0.00	0	0.00	3,520	0.00
Unknown	1,287,223	0.22	10,638	0.00	6,292,305	1.09	7,586,982	1.32
Total	106,671,932	18.52	16,479,266	2.86	298,405,446	51.82	352,349,656	61.19

Table 8 Statistics of gene predictions in the burbot genome

	Gene set	Protein coding gene number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
denovo	Genscan	13,337	24,305	1,405	7.71	182.24	3,412
	AUGUSTUS	23,738	8,372	954.69	4.68	203.97	2,015
Homolog	<i>Acanthochromis polyacanthus</i>	57,870	6,446	829.94	4.38	189.61	1,663
	<i>Oryzias latipes</i>	47,042	6,847	994.59	4.95	201.03	1,483
	<i>Amphiprion ocellaris</i>	51,189	7,883	950.76	4.93	192.75	1,763
	<i>Anabas testudineus</i>	52,035	8,460	951.79	4.90	194.09	1,923
	<i>Astatotilapia calliptera</i>	52,021	7,911	994.61	4.85	205.24	1,798
	<i>Astyanax mexicanus</i>	67,301	6,618	810.26	3.85	210.52	2,038
	<i>Austrofundulus limnaeus</i>	62,600	6,159	853.22	4.05	210.81	1,741
	<i>Gadus morhua</i>	72,746	6,286	827.65	4.00	207.03	1,821
	<i>Lepisosteus oculatus</i>	38,891	7,748	954.07	5.34	178.68	1,566
	<i>Notothenia coriiceps</i>	49,513	5,600	795.50	4.18	190.15	1,509
RNAseq	Transdecoder	16,207	16,722	1,769	11.59	325.21	1,223
BUSCO		4,576	12,155	2,018	13.42	150.44	816.44
MAKER		20,312	15,221	1,395	8.39	275.15	1,748
HiCESAP		21,664	14,606	1,658	10.45	292.38	1,223

Table 9 Summary of functional annotations for predicted genes

Annotation database	Annotated number of predicted genes	Percent (%)
InterPro	19058	87.97
GO	14369	66.33
KEGG	13247	61.15
Swissprot	18619	85.94
TrEMBL	20540	94.81
NR	20576	94.98
KOG	17187	79.33
All annotated	20658	95.36
Predicted genes	21664	