PORTLAND PRESS

Research Article

Genome survey and microsatellite motif identification of *Pogonophryne albipinna*

Euna Jo[1,2,*], YII Hwan Cho[1,*], Seung Jae Lee[1], Eunkyung Choi[1], Jinmu Kim[1], Jeong-Hoon Kim[2],
Young Min Chi[1] and Hyun Park[1]

[1]Department of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Korea; [2]Division of Life Sciences, Korea Polar Research Institute (KOPRI), Yeonsu-gu, Incheon 21990, Korea

Correspondence: Hyun Park (hpark@korea.ac.kr)

The genus *Pogonophryne* is a speciose group that includes 28 species inhabiting the coastal or deep waters of the Antarctic Southern Ocean. The genus has been divided into five species groups, among which the *P. albipinna* group is the most deep-living group and is characterized by a lack of spots on the top of the head. Here, we carried out genome survey sequencing of *P. albipinna* using the Illumina HiSeq platform to estimate the genomic characteristics and identify genome-wide microsatellite motifs. The genome size was predicted to be ~883.8 Mb by K-mer analysis (K = 25), and the heterozygosity and repeat ratio were 0.289 and 39.03%, respectively. The genome sequences were assembled into 571624 contigs, covering a total length of ~819.3 Mb with an N50 of 2867 bp. A total of 2217422 simple sequence repeat (SSR) motifs were identified from the assembly data, and the number of repeats decreased as the length and number of repeats increased. These data will provide a useful foundation for the development of new molecular markers for the *P. albipinna* group as well as for further whole-genome sequencing of *P. albipinna*.

Introduction

The genus *Pogonophryne* Regan, 1914 is the most species-rich group among the perciform suborder Notothenioidei, with 28 species reported to date [1,2]. They inhabit coastal or deep waters of the Southern Ocean off Antarctica [2]. Recently, several species have been newly discovered during longlining of the Antarctic toothfish, *Dissostichus mawsoni* [1–7], but their morphological and molecular identification is still complicated.

Taxonomically, the genus *Pogonophryne* is one of the complex taxa distinguished from other taxa by slight meristic differences, and their key diagnostic character, namely the mental barbell, is highly variable in some species [6,8]. It is difficult to compare the morphology of the species from this genus because many of them were described based on only a few specimens from a single sampling site [9,10]. Accordingly, taxonomists have divided the genus *Pogonophryne* into five species groups: *P. mentella*, *P. scotti*, *P. barsukovi*, *P. marmorata*, and *P. albipinna* groups [5,11].

Phylogenetic studies have been carried out on these groups using several mitochondrial and nuclear markers, and the monophyly of these five species groups was supported by mitochondrial NADH dehydrogenase subunit 2 (ND2) and cytochrome *c* oxidase I (COI) gene markers [5,10]. However, molecular identification at the species level showed poor resolution due to low genetic variations related to a very recent divergence of the genus *Pogonophryne*, as is the case with other species in the family Artedidraconidae [10,12–14]. Therefore, it is necessary to develop markers with improved discriminatory ability for genome-wide analyses, such as microsatellite and single nucleotide polymorphism (SNP) markers. In particular, microsatellites, also termed simple sequence repeats (SSRs), have already been validated for their effectiveness in fish species delimitation [15].

*These authors contributed equally to this work.

1

Table 1 Statistics of the genome survey sequencing data of *P. albipinna*

Raw data (bp)	Total reads	Q20 (%)	Q30 (%)	GC content (%)
57104280342	378174042	96.6	91.8	41.7

The molecular data on *Pogonophryne*, mostly mitochondrial ND2 and COI, are available from the NCBI Gen-Bank database [2,5] for less than half of the species (13 out of 28). Among these species, *P. albipinna* has been reported recently with its complete mitochondrial genome sequence [16], and this is the first genome survey study of *Pogonophryne*. *Pogonophryne albipinna*, also known as white-fin plunderfish, belongs to the *P. albipinna* group, which is the most deep-living group of the genus and is mainly characterized by an absence of dark spots on the top of the head [1,5,11].

In the present study, based on next-generation sequencing (NGS), we estimated the genomic characteristics of *P. albipinna* and identified genome-wide SSR motifs. The present study can be used as a basis for further whole-genome sequencing of *P. albipinna* and the development of new molecular markers for distinguishing between *Pogonophryne* species.

Materials and methods
Sample preparation and genome survey sequencing

Sample of *P. albipinna* was collected from the Ross Sea (77°05′S, 170°30′E on CCAMLR Subarea 88.1), Antarctica and frozen while being transferred to the laboratory. The frozen sample was dissected to obtain muscle tissue samples, which were used to extract genomic DNA following the traditional phenol-chloroform method. DNA quantity and quality were checked using a Qubit fluorometer (Invitrogen, Life Technologies, CA, U.S.A.) and a fragment analyzer (Agilent Technologies, CA, U.S.A.). Species were identified by morphology as well as using mitochondrial COI markers [17]. The DNA was randomly fragmented into 350-bp fragments using a Covaris M220 focused-ultrasonicator (Covaris, MA, U.S.A.). A paired-end DNA library was prepared and sequenced on the Illumina HiSeq 2000 platform according to the manufacturer's protocol.

Data analysis

The quality values of Q20 (percentage of bases whose base call accuracy exceeds 99%) and Q30 (percentage of bases whose base call accuracy exceeds 99.9%) and the GC content were evaluated from the primary Illumina paired-end data. K-mer analysis was conducted using Jellyfish 2.1.4 [18] with K-values of 17, 19, and 25. In order to estimate the genome size, heterozygosity rate and repeat content, we used GenomeScope [19] in R version 3.4.4 [20] based on the K-mer distribution (K = 25), which selected the one that the GenomeScope model showed the best match to the observed K-mer frequencies. The *de novo* draft genome was assembled using Maryland Super-Read Celera Assembler (MaSuRCA) version 3.3.4 [21], and contig-level assembly statistics were then calculated using the assemblathon_stats.pl script (available at: https://github.com/ucdavis-bioinformatics/assemblathon2-analysis/blob/master/assemblathon_stats.pl; accessed on 1 January 2021) [22]. Genome-wide identification of di- to hexanucleotide microsatellite motifs with minimum five repetitions, and primer design were performed using the pipelines of QDD version 3.1.2 [23]. Microsatellites were extracted with 200-bp flanking regions on both sides and sequences shorter than 80 were eliminated. Three QDD steps were proceeded with default parameters, and -contig 1 (step 1), -make_cons 0 (step 2) and -contig 1 (step 3) options were added. Primer pairs were selected by Primer3 software [24] to meet the following criteria: the expected PCR product size of 100–150 bp, the primer melting temperature (Tm) of 59–60°C, and the primer length of 20–25 bases.

Results and discussion
Genome size estimation and sequence assembly

The genome survey sequencing of *P. albipinna* yielded a total of ∼57.1 Gb of raw reads through the Illumina paired-end library (Table 1). The Q20 and Q30 values of the raw reads were 96.6 and 91.8%, respectively (Table 1), indicating the high quality of this genome sequencing data [25]. In addition, the GC content of the raw reads was 41.7% (Table 1). The Illumina paired-end data were then used to predict the genomic characteristics of *P. albipinna* by K-mer analysis. Based on the 25-mer frequency distribution, the genome size was estimated to be 883.8 Mb, and the heterozygous and repetitive sequence rates were 0.289 and 0.751%, respectively (Table 2, and Figure 1).

PORTLAND PRESS

Table 2 Genome estimation based on K-mer analysis of *P. albipinna*

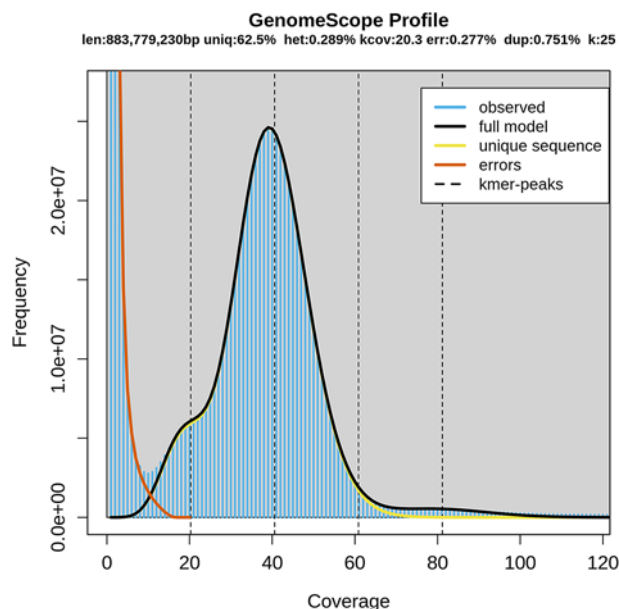| K-mer	Genome size (bp)	Heterozygosity (%)	Duplication ratio (%)
17	829857227	0.275	0.795
19	843219952	0.294	0.758
25	883779230	0.289	0.751

Figure 1. K-mer (K = 25) distribution of *P. albipinna* genome

Blue bars represent the observed K-mer distribution; black line represents the modeled distribution without the error K-mers (indicated by the red line), up to a maximum K-mer coverage specified in the model (indicated by the yellow line). Len, estimated total genome length; Uniq, unique portion of the genome (not repetitive); Het, heterozygosity rate; Kcov, mean K-mer coverage for heterozygous bases; Err, error rate; Dup, duplication rate.

Table 3 Statistics of the assembled genome sequences of *P. albipinna*

	Total length (bp)	Total number	Max length (bp)	N50 length (bp)	GC content (%)
Contig	819289238	571624	51460	2867	41.02

In earlier studies, the nuclear DNA content of *P. scotti* was measured to be 4.05 pg/diploid cell using the Feulgen staining method [26]. When this measurement is converted into the haploid genome size, it shows that the nuclear DNA content of this species is 1.98 Gb, which is more than twice as high as our estimate. Meanwhile, other research on notothenioid genome size by flow cytometry showed that their genome size was 0.78–1.43 Gb [27], and more recent studies based on NGS data indicated a genome size of 0.64–1.06 Gb [28–32]. These size ranges are comparable with those indicated by our results, suggesting that further studies are needed to acquire more accurate knowledge of *P. albipinna* genome size.

Furthermore, the Illumina paired-end sequences of *P. albipinna* were assembled into contigs using MaSuRCA. We obtained 571624 contigs with a total length of 819289238 bp. The maximum and N50 contig lengths were 51460 and 2867 bp, respectively, with a GC content of 41.02% (Table 3). These results of genome survey sequencing provide useful preliminary data for further whole-genome studies to achieve more thorough assembly and chromosomal-level scaffolding using novel state-of-the-art genetic techniques.
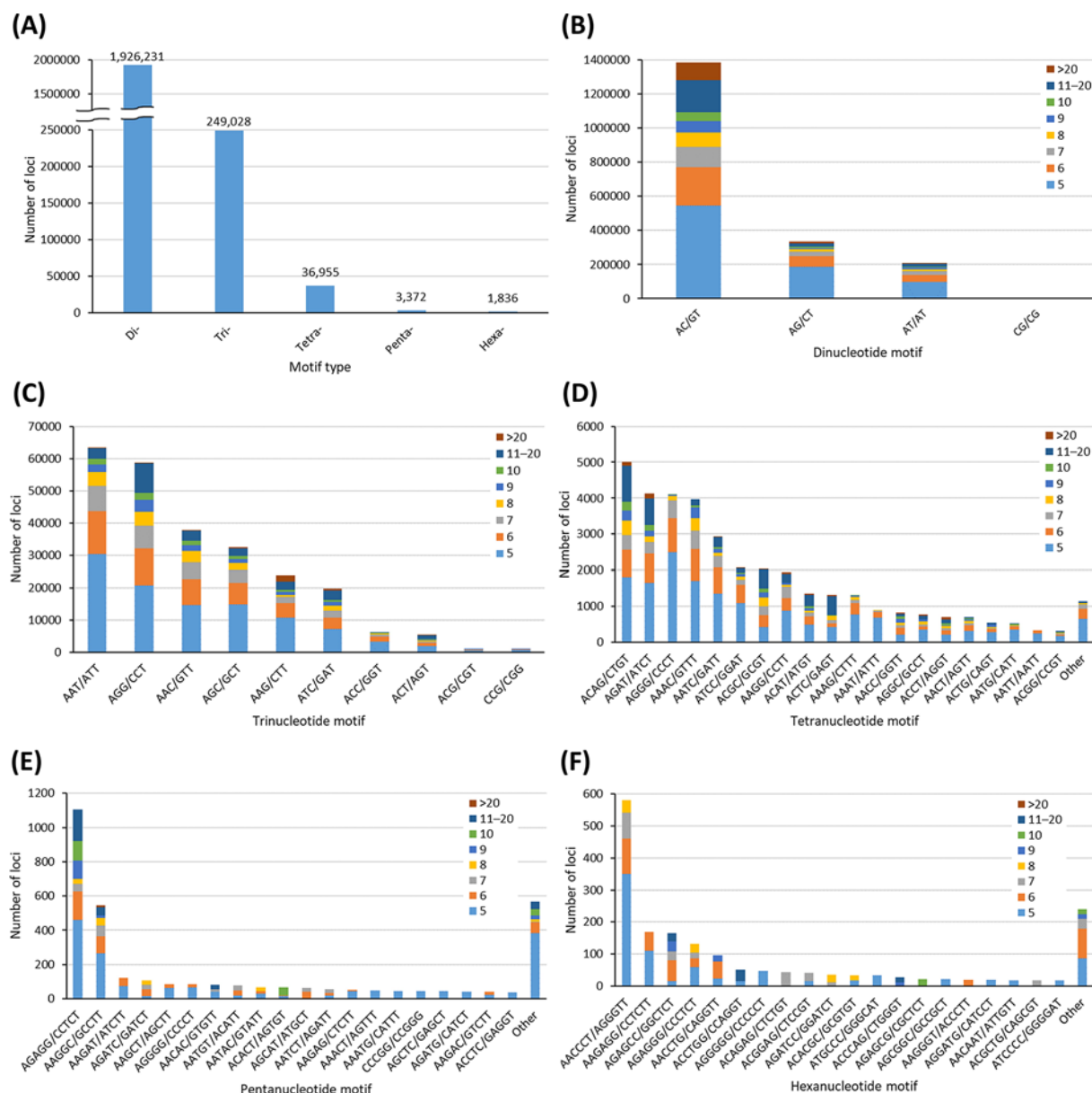
Figure 2. Type and frequency of microsatellite motifs in *P. albipinna* genome
(**A**) Frequency of different microsatellite motif types. (**B**) Frequency of different dinucleotide microsatellite motifs. (**C**) Frequency of different trinucleotide microsatellite motifs. (**D**) Frequency of different tetranucleotide microsatellite motifs. (**E**) Frequency of different pentanucleotide microsatellite motifs. (**F**) Frequency of different hexanucleotide microsatellite motifs.

Table 4 Statistics of SSR for *P. albipinna*

Statistics	Di-	Tri-	Tetra-	Penta-	Hexa-	Total
SSR number	1926231	249028	36955	3372	1836	2217422
Percentage	86.87	11.23	1.67	0.15	0.08	-

Microsatellite motif identification

A total of 2217422 microsatellite motifs were identified from the genome assembly of *P. albipinna*. Among them, dinucleotide motifs were the most prevalent (1926231; 86.87%), followed by trinucleotides (249028; 11.23%), tetranucleotides (36955; 1.67%), pentanucleotides (3372; 0.15%), and hexanucleotides (1836; 0.08%) (Table 4 and Figure

2A). The tendency of the motif frequency in the studied species was similar to that in other fish species, with the dinucleotide motif being predominant [33,34]. In the dinucleotides, the most frequent motif was AC/GT (71.84%), followed by AG/CT (17.29%), AT/AT (10.82%), and CG/CG (0.05%) (Figure 2B). In the trinucleotides, the most frequent motif was AAT/ATT (25.43%), followed by AGG/CCT (23.57%), and AAC/GTT (15.09%) (Figure 2C). The most abundant motifs in the tetra-, penta-, and hexanucleotides were ACAG/CTGT (13.53%), AGAGG/CCTCT (32.80%), and AACCCT/AGGGTT (31.92%), respectively (Figure 2D–F). Information on 99 pairs of microsatellite marker is presented in Supplementary Table S1. To ensure the usability of the microsatellite markers, subsequent validation studies are required. Moreover, if these markers are applied for studying the *P. albipinna* group, more meaningful results could be obtained and interspecific variation could be explained better than when using conventional mitochondrial markers.

Conclusion

In the present study, genome survey sequencing of *P. albipinna* was conducted to investigate its genomic characteristics and identify microsatellite motifs. The genome size estimated by K-mer analysis (K = 25) was 883.8 Mb, and the heterozygosity and duplication rates were 0.289 and 0.751%, respectively. The assembled genome had a total size of 819.3 Mb, with an N50 of 2867 bp and a GC content of 41.02%. A total of 2217422 SSR motifs were identified from the genome data, among which dinucleotide motifs accounted for the majority of repeat motifs (86.87%). These data will be a useful basis for novel molecular marker development as well as for further whole-genome sequencing of *P. albipinna*.

Data Availability

The *P. albipinna* genome project has been registered in NCBI under the BioProject number PRJNA697561. The whole-genome sequence has been deposited in the Sequence Read Archive (SRA) database under accession numbers: SRS13617358 and SAMN17672856.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

CRediT Author Contribution

Euna Jo: Data curation, Writing—original draft, Writing—review and editing. **YIl Hwan Cho:** Data curation, Writing—original draft. **Seung Jae Lee:** Data curation, Software, Formal analysis. **Eunkyung Choi:** Data curation, Software, Formal analysis. **Jinmu Kim:** Data curation, Software, Formal analysis. **Jeong-Hoon Kim:** Resources, Data curation. **Young Min Chi:** Conceptualization, Data curation. **Hyun Park:** Conceptualization, Data curation, Writing—original draft, Writing—review and editing.

Ethics Approval

Ethical approval was not required for the present study because no endangered or alive animals were involved. The specimen used in the present study was caught by line and hook fishing and was dead when collected. The present study including sample collection and experimental research conducted on these animals was according to the law on activities and environmental protection to Antarctic approved by the Minister of Foreign Affairs and Trade of the Republic of Korea (MOFA2794).

Abbreviations

COI, cytochrome *c* oxidase I; MaSurCA, Maryland Super-Read Celera Assembler; ND2, NADH dehydrogenase subunit 2; NGS, next-generation sequencing; SSR, simple sequence repeat.

References

1 Balushkin, A.V. and Spodareva, V.V. (2015) New species of the toad plunderfish of the "albipinna" group, genus Pogonophryne (Artedidraconidae) from the Ross Sea (Antarctica). *J. Ichthyol.* **55**, 757–764, https://doi.org/10.1134/S003294521506003X

2 Shandikov, G.A. and Eakin, R.R. (2013) Pogonophryne neyelovi, a new species of Antarctic short-barbeled plunderfish (Perciformes, Notothenioidei, Artedidraconidae) from the deep Ross Sea. *ZooKeys* **296**, 59–77, https://doi.org/10.3897/zookeys.296.4295

3 Balushkin, A.V. and Spodareva, V.V. (2013) Pogonophryne sarmentifera sp. nov. (Artedidraconidae; Notothenioidei; Perciformes)—the deep-water species of Antarctic plunderfishes from the Ross Sea (Southern Ocean). *Tr. Zool. Inst.* **317**, 275–281

4   Balushkin, A.V. (2013) A new species of Pogonophryne (Perciformes: Notothenioidei: Artedidraconidae) from the deep Ross Sea, Antarctica. *Tr. Zool. Inst.* **317**, 119–124

5   Eakin, R.R., Eastman, J.T. and Near, T.J. (2009) A new species and a molecular phylogenetic analysis of the Antarctic fish genus Pogonophryne (Notothenioidei: Artedidraconidae). *Copeia* **2009**, 705–713, https://doi.org/10.1643/CI-09-024

6   Shandikov, G.A., Eakin, R.R. and Usachev, S. (2013) Pogonophryne tronio, a new species of Antarctic short-barbeled plunderfish (Perciformes: Notothenioidei: Artedidraconidae) from the deep Ross Sea with new data on Pogonophryne brevibarbata. *Polar Biol.* **36**, 273–289, https://doi.org/10.1007/s00300-012-1258-4

7   Balushkin, A., Petrov, A. and Prutko, V. (2010) Pogonophryne brevibarbata sp. nov.(Artedidraconidae, Notothenioidei, Perciformes)–a new species of toadlike plunderfish from the Ross Sea, Antarctica. *Proc. Zool. Inst. Russ. Acad. Sci.* **314**, 381–386

8   Eakin, R.R., Eastman, J.T. and Jones, C.D. (2001) Mental barbel variation in Pogonophryne scotti Regan (Pisces: Perciformes: Artedidraconidae). *Antarct. Sci.* **13**, 363–370, https://doi.org/10.1017/S0954102001000517

9   Eakin, R. (1990) Artedidraconidae. In *Fishes of the Southern Ocean* (Gon, O. and Heemstra, P.C., eds), pp. 332–356, JLB Smith Institute of Ichthyology, Grahamstown

10 Smith, P., Steinke, D., Dettai, A., McMillan, P., Welsford, D., Stewart, A. et al. (2012) DNA barcodes and species identifications in Ross Sea and Southern Ocean fishes. *Polar Biol.* **35**, 1297–1310, https://doi.org/10.1007/s00300-012-1173-8

11 Balushkin, A. and Eakin, R. (1998) A new toad plunderfish Pogonophryne fusca sp. nova (Fam. Artedidraconidae: Notothenioidei) with notes on species composition and species groups in the genus Pogonophryne Regan. *J. Ichthyol.* **38**, 574–579

12 Dettai, A., Lautredou, A.-C., Bonillo, C., Goimbault, E., Busson, F., Causse, R. et al. (2011) The actinopterygian diversity of the CEAMARC cruises: barcoding and molecular taxonomy as a multi-level tool for new findings. *Deep Sea Res. Part II* **58**, 250–263, https://doi.org/10.1016/j.dsr2.2010.05.021

13 Lecointre, G., Gallut, C., Bonillo, C., Couloux, A., Ozouf-Costaz, C. and Dettaï, A. (2011) The antarctic fish genus Artedidraco is paraphyletic (Teleostei, Notothenioidei, Artedidraconidae). *Polar Biol.* **34**, 1135–1145, https://doi.org/10.1007/s00300-011-0974-5

14 Near, T.J., Dornburg, A., Kuhn, K.L., Eastman, J.T., Pennington, J.N., Patarnello, T. et al. (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3434–3439, https://doi.org/10.1073/pnas.1115169109

15 Vanhaecke, D., De Leaniz, C.G., Gajardo, G., Young, K., Sanzana, J., Orellana, G. et al. (2012) DNA barcoding and microsatellites help species delimitation and hybrid identification in endangered galaxiid fishes. *PLoS ONE* **7**, e32939, https://doi.org/10.1371/journal.pone.0032939

16 Tabassum, N., Alam, M.J., Kim, J.-H., Lee, S.R., Lee, J.-H., Park, H. et al. (2020) Characterization of complete mitochondrial genome of Pogonophryne albipinna (Perciformes: Artedidraconidae). *Mitochondrial DNA Part B.* **5**, 156–157, https://doi.org/10.1080/23802359.2019.1698361

17 Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R. and Hebert, P.D. (2005) DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 1847–1857, https://doi.org/10.1098/rstb.2005.1716

18 Marçais, G. and Kingsford, C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011

19 Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J. et al. (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204, https://doi.org/10.1093/bioinformatics/btx153

20 R Core Team (2017) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria

21 Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L. and Yorke, J.A. (2013) The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677, https://doi.org/10.1093/bioinformatics/btt476

22 assemblathon2-analysis/assemblathon_stats.pl.
    https://github.com/ucdavis-bioinformatics/assemblathon2-analysis/blob/master/assemblathon_stats.pl

23 Meglécz, E., Pech, N., Gilles, A., Dubut, V., Hingamp, P., Trilles, A. et al. (2014) QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: Experimental validation of variables determining genotyping success rate. *Mol. Ecol. Resour.* **14**, 1302–1313, https://doi.org/10.1111/1755-0998.12271

24 Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols*, pp. 365–386, Springer

25 Li, G.-Q., Song, L.-X., Jin, C.-Q., Li, M., Gong, S.-P. and Wang, Y.-F. (2019) Genome survey and SSR analysis of Apocynum venetum. *Biosci. Rep.* **39**, BSR20190146, https://doi.org/10.1042/BSR20190146

26 Morescalchi, A., Morescalchi, M.A., Odierna, G., Sitingo, V. and Capriglione, T. (1996) Karyotype and genome size of zoarcids and notothenioids (Taleostei, Perciformes) from the Ross Sea: cytotaxonomic implications. *Polar Biol.* **16**, 559–564, https://doi.org/10.1007/BF02329052

27 Detrich, H.W., Stuart, A., Schoenborn, M., Parker, S.K., Methé, B.A. and Amemiya, C.T. (2010) Genome enablement of the notothenioidei: genome size estimates from 11 species and BAC libraries from 2 representative taxa. *J. Exp. Zool. B Mol. Dev. Evol.* **314**, 369–381, https://doi.org/10.1002/jez.b.21341

28 Shin, S.C., Ahn, D.H., Kim, S.J., Pyo, C.W., Lee, H., Kim, M.K. et al. (2014) The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome Biol.* **15**, 1–14, https://doi.org/10.1186/s13059-014-0468-1

29 Ahn, D.H., Shin, S.C., Kim, B.M., Kang, S., Kim, J.H., Ahn, I. et al. (2017) Draft genome of the Antarctic dragonfish, Parachaenichthys charcoti. *Gigascience* **6**, gix060, https://doi.org/10.1093/gigascience/gix060

30 Kim, B.M., Amores, A., Kang, S., Ahn, D.H., Kim, J.H., Kim, I.C. et al. (2019) Antarctic blackfin icefish genome reveals adaptations to extreme environments. *Nat. Ecol. Evol.* **3**, 469–478, https://doi.org/10.1038/s41559-019-0812-7

31 Chen, L., Lu, Y., Li, W., Ren, Y., Yu, M., Jiang, S. et al. (2019) The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonian robalo genomes. *GigaScience* **8**, giz016, https://doi.org/10.1093/gigascience/giz016

32 Lee, S.J., Kim, J.H., Jo, E., Choi, E., Kim, J., Choi, S.G. et al. (2021) Chromosomal assembly of the Antarctic toothfish (Dissostichus mawsoni) genome using third-generation DNA sequencing and Hi-C technology. *Zool. Res.* **42**, 124–129, https://doi.org/10.24272/j.issn.2095-8137.2020.264

33 Chen, B., Sun, Z., Lou, F., Gao, T.-X. and Song, N. (2020) Genomic characteristics and profile of microsatellite primers for Acanthogobius ommaturus by genome survey sequencing. *Biosci. Rep.* **40**, BSR20201295, https://doi.org/10.1042/BSR20201295

34 Xu, S.-Y., Song, N., Xiao, S.-J. and Gao, T.-X. (2020) Whole genome survey analysis and microsatellite motif identification of Sebastiscus marmoratus. *Biosci. Rep.* **40**, BSR20202252, https://doi.org/10.1042/BSR20192252

Supplementary Table S1. Information of SSR primers of *Pogonophryne albipinna*

Primer No.	Motif	PCR product size (bp)	Forward primer Sequence (5' → 3')	Tm (℃)	Reverse primer Sequence (5' → 3')	Tm (℃)
1	(AC)15	140	TCTAGTCGAACAGCTCAGCC	59.2	AGTGGAATGAGGATGGGAGC	59.2
2	(AC)11	139	CATGTTTATTCCCACAGATGGCA	59.3	AATGTGAGGAGTTCAGGCCC	59.7
3	(AC)10	150	GGCAAGCTTTGTGTTCAGCT	59.6	ATGGACAATGAAAGGCGGGA	59.7
4	(AG)12	110	TCCCTCAACTTGTTCTTCTGGT	59.2	TCTAAGACCATGCTGACGGC	59.8
5	(AC)14	125	TCCATTCATCCTTGCCTGCT	59.4	ACAGCAGCCACATTAGACCA	59.3
6	(AC)14	142	ATTGATCAGCATTGCACCGC	59.9	CTCACTGCCAACATCAGGGT	60.0
7	(AC)10	133	GGAGCCTACAGTCCAATCGG	59.9	ACCAAGGTTTCAGAGCAGCA	59.8
8	(AC)10	135	GAGCCTACAGTCCAATCGGG	59.9	AGCACCAAGGTTTCAGAGCA	59.8
9	(AC)10	148	ACGTTTGAACCTGAAATGGGC	59.7	TGGCTCATTAGTGGTGCAGA	59.0
10	(AC)12	142	GCCTTACTTTCTGACAGCAGC	59.5	ACTGAGCTCATGTGAGGACG	59.5
11	(AC)11	150	AGTAAACTGCTGCCAAGGCT	59.9	TATCTGTCTGGCACCACCTG	59.1
12	(AC)11	135	CGTTTGATCTTGCCTGCAGG	59.8	CACCAGGTAACCTCAGCAGA	59.0
13	(AC)16	150	AGACACTCAAACTCTTAGACGCT	59.4	TCTCCGAGTGCAATGAGACG	59.8
14	(AC)10	137	CCACCGTGTTAGAAACGCAG	59.5	TTCACTGCTAAGCCAAGGCA	59.9
15	(AC)11	140	GCACCAGGATAGTCAGCACA	59.8	AGCACCTTTCACCTTGAGGG	59.9
16	(AC)11	108	ATCACCCAGGATAACCACGC	59.8	GTGGGCCTTACATCTGCTCA	59.7
17	(AG)10	118	CACGGTGGTTAAATCAGATTGGG	59.9	TAGGCCATGCAGTCATCACC	59.8
18	(AC)10	136	GCAGGATATGGGTGTTTGCG	59.6	TAGCGTCTTTGCTCAGTGCT	59.7
19	(AC)18	141	GGTATCGCTAACACCACCCA	59.5	CCGTTTGGTTAATGTCTTCGCT	59.5
20	(AC)11	140	CCAGAGCTGCTATCAGTGCA	59.8	TAGAGATGAGAGGCGGTGGT	59.7
21	(AC)15	116	GAAGCTTAGGTTTGCCTGGC	59.5	TTTAGCTCCTTCGCACCAGT	59.3
22	(AC)12	141	GCTCCATGGTGAACTTCTGC	59.2	CAGCAGAATTGAGTTAACGGCA	59.5
23	(AC)12	140	TCCATGGTGAACTTCTGCGT	59.6	ACAGCAGAATTGAGTTAACGGC	59.5
24	(AC)13	142	AGGTTTAAAGTTGTACATTTCCGCA	59.7	GGAGGACATACTGTGGTGCA	59.4
25	(AT)15	150	ATACGAGTCAGTGTGAGGCC	59.2	GGCTGGATCTCAGAAAGGTCA	59.4
26	(AC)11	106	TGGGATGGACTCAGAGCTGA	60.0	CTGATCAGCTGCTGAGACGT	59.8
27	(AC)11	118	TCACACCCTCATTTGCTCCC	60.0	CTGACAACACAGGAGCGACT	60.0
28	(AG)16	127	TCTGTGTTCGTGTTTGTGCG	59.6	TGTATAGCCCGCAAACTGCT	59.7
29	(AC)13	150	TTCTGGTCAAAGGCAACGGA	59.8	GGCTGATGGAAGATGACAGGT	59.8
30	(AC)11	123	GCTCCGTCTCTTTGAGCTCA	59.8	CAACCGGCTGATAAACACCG	59.6
31	(AC)11	142	TGTCCGTTACAACTGAGTACTGG	60.0	TCCACTCCAATGAGCTCTGC	59.7
32	(AC)10	123	CTGCAATAACAGGCCAGCAC	59.8	TTTGGAGGCACAGAGACCAC	59.9
33	(AC)11	119	GTCTCAGATACAGAAAGACAGGCT	59.8	TGTTGATTGGAGGGAGCAGG	59.7
34	(AC)10	138	ACCATGCCATGTTAGCGCTA	59.8	TGGACATTCCCACATGCACA	59.9
35	(AC)17	129	CATGGCTTCTGAAGGAGGCT	59.7	CCCTGTTAGATTGATGGGAGGG	59.9
36	(AC)11	143	CCCTTTCATCCGTTACATGCTG	59.6	TCTGAGATCTCTAGTCAGGGCA	59.5
37	(AG)16	139	TGTCTATCTGCCTTTCTGCCA	59.1	TCACCGTCATGCTAATCCTGG	59.9

38	(AT)10	141	TCTGCACATTGGTCTTGCAAC	59.7	TCCTCGTATGTGTAAACCTACCTG	59.6
39	(AC)11	122	TGGCTCTAGATCTACTCCGGG	59.9	GCATTACAGGGCATGAAGGC	59.6
40	(AC)16	138	GTTCAGAGGCAGACAGAGGG	59.8	AGGTGAACACAGACTCTGGC	59.6
41	(AC)10	139	AGCAGTGTTTACAGGCTCCC	60.0	ACCATGTCTCATCTGGACACC	59.4
42	(AC)10	140	ATCAGAAACACATTTGCCGCT	59.1	TGCAGCATTGAACACAATGGG	60.0
43	(AC)17	125	TCTGTGAGGTCGTTCAGCTG	59.7	CCTGCATGTCTGTGTTTGTCC	59.7
44	(AC)12	131	AGGCACATTTATCCACCGCT	59.7	ACAGGCCTTTCACATGCTCA	59.9
45	(AG)26	150	ACTCTGAAATGACACCGTGCT	59.9	AGCATGTGCCAATCAAACGG	59.8
46	(AC)26	143	CAGGGTACTCTCCATGGTGC	59.8	GGCAACTCTAATAGATGACTCATGC	59.6
47	(AC)11	123	CCTCAGATTGGATTCCAGAGCA	59.8	CTTGGTGTTGGAGAGGAGGG	59.7
48	(AC)12	134	GCTTCATGACCACCAAGGGA	60.0	AGGAGAGCCAATGTTAGCCG	59.8
49	(AT)11	123	CAACCAACAAGAGGCACGTC	59.7	TCGCAAGAGGAGATGATGCC	59.9
50	(AC)17	123	ACAAGGACTACTGGCCAAGC	60.0	TAATTCATCGCCATGCACGC	59.7
51	(AC)25	140	ATTTCAGCCATCCATCAGCCT	59.8	GATTCAACGCAGGTCTGACG	59.3
52	(AC)11	142	TATAAGCCCTGCAGCACGAG	59.9	AATGAGCTTTACGATGGCGC	59.3
53	(AC)14	136	GTGCATTCTGTCCTCTGTTGC	59.8	CGTGCTAATCTCCCAGAGGG	59.6
54	(AC)10	118	GTCCAGACACTGAGAGGCTG	59.8	ACGTGTAAGTGCACTCGTCT	59.3
55	(AC)11	150	CAGGCAAACATAGTCCTGCAG	59.3	ACGTGCATCTGATAAGGCCT	59.2
56	(AC)11	136	CGCCTTAATTGTTTCACGCAC	59.0	CGCCCATATCCAGTTCTGCT	59.9
57	(AC)11	128	GCATTTCTTTAGCCTAAAGCCGA	59.6	TGGATTAAGCGGATAAGGCCC	59.9
58	(AC)10	117	TCTCCTGTGGTCTTACCCGT	59.9	AGTACTTGTCCTCCAGGTGC	59.0
59	(AC)10	128	CGACAGTTACACCTACGCTCA	59.8	CAACGCACATCCTTCATGGG	59.5
60	(AC)17	125	ACACTTCACTGGTCTGGGTG	59.5	TTTCCCTGCCATGACTCGAG	59.8
61	(AT)10	147	CCTTCCTCTGTAAGCGCTGT	59.8	TTGTTGTTCTGCCCTCCTGT	59.4
62	(AC)11	147	AACGCTATTAGCACCCTGGG	59.8	CCCTGTGAGGAGACTTCAGC	59.8
63	(AT)10	137	TCTCCATCCATCCTAAATCAGCA	59.0	AGCAAGGTCCAATAAAGAAACCG	59.5
64	(AC)11	140	ACATCTGCTTGTGTAGGCGT	59.7	CTCCACAATCCATCATGCGC	59.7
65	(AC)11	144	GGCCGCCTCTTAGTATGACC	60.0	AAGCAGAAAGAGACAAGCCCA	59.9
66	(AC)10	141	GGAGACTGAGTTTGAATCGGGA	59.8	CCCGATCTGAACAAAGATTGCC	59.9
67	(AT)11	140	TGTGATGCTAGAGTGACTCGTG	59.8	CATTGTTTACACCCGTATGAGCT	59.1
68	(AT)11	140	GTGATGCTAGAGTGACTCGTGA	59.6	ACATTGTTTACACCCGTATGAGC	59.3
69	(AC)14	140	CAATAACCTACAGCAGTGACCTG	59.1	TGCATCAGGATCTTTAACGCAC	59.3
70	(AC)14	140	AATAACCTACAGCAGTGACCTGT	59.4	CTGCATCAGGATCTTTAACGCA	59.1
71	(AC)10	141	TGCTACGGCAACTAACACCT	59.3	GAACATTCCGTACAACCAGGC	59.5
72	(AC)14	142	GTCCCAGAAAGAGCTCTGCA	59.7	TCACATTAGGAGCTACTGTAACCC	59.6
73	(AC)21	136	AGAAGGGAGCAAGGTTGAGC	60.0	GTCCAAAGAGGCTGTCGCTA	59.8
74	(AC)22	142	GGCTGTCATTCCCTTTCTGC	59.2	GCTGATCGGAGAACACAGGT	59.8
75	(AC)10	148	AAATCTCCTCCACGTGGACC	59.4	TTTGGGCTGGACCTAACTGG	59.6
76	(AC)16	142	ACACAGAGGGACAAACGTCA	59.2	ATGAGTTACGTCACCAGCCC	59.8
77	(AGC)7	144	TAGCAGAACGGTTAGCTCGG	59.5	GGGTAATGGTGACTCTGCCA	59.4
78	(AGG)8	142	ATCGATCGACAGGTCAAGGC	59.9	GGGTACTCCGCTCTAGTTGC	59.9

79	(AAT)8	144	CCACCAACATCTGTCTCCGT	59.7	TTCACCATGTAAAGCGGCCT	60.0
80	(AAT)8	144	ACCACCAACATCTGTCTCCG	59.7	TCACCATGTAAAGCGGCCTT	60.0
81	(ATC)7	138	AGCCCAATCTGAAACAGGAGG	60.0	CGTGGCTGATGTTCTTGCTG	59.8
82	(AGG)7	121	CACAGTTGACAAGGCACAGC	60.0	AGAGGAACAGGATAGGATGGGA	59.5
83	(AAT)12	139	TGTAAAGCGACCTTGGGTGA	59.2	TCGGAGTGTAGTGAGTCACCT	59.9
84	(AGG)12	141	ACCCTCTTGTCACTGACAGC	59.6	AGCTACAACTGCTGTTAGGGT	59.0
85	(AGG)12	140	CCCTCTTGTCACTGACAGCA	59.6	AGCTACAACTGCTGTTAGGGTT	59.6
86	(AAC)8	118	CTCCAAAGCCTTGTGAAGCG	59.8	TCATTGTCAGTGCTGGTCCC	60.0
87	(AGC)7	133	ACAAGATGAAGAGGCAGGCT	59.0	TTCCCATCCTTCACCAGCAC	60.0
88	(AAC)7	121	ACATGATCTCTGCAGCTGCT	59.5	AGGTAATGACACCATGCAGCA	60.0
89	(ACC)9	136	ATCCACAGACTGATCCAGCG	59.5	TCTGACATCACCTCTGCCAG	59.1
90	(AAC)8	138	TCTCCAGTCAGCTCAACACG	59.7	GAGGGATGCTCTGACTTGCA	59.7
91	(AAC)8	131	ATGTGAAGACCCTGACCTGC	59.7	ATTGTGGTGGGTGAGACAGG	59.6
92	(AAG)10	140	TTTGAAGCTGCTAACACGCAA	59.3	AGCAGTAACGTTAGAGGGCAG	59.8
93	(AGG)8	133	TCAATGCCTTATTTAACCAATGGGA	59.0	TCTGAAAGGCATCTTTGAGGCT	60.0
94	(AGG)7	143	TCCTTGTGAATGACAGCGCT	60.0	TGCATTGTTATGTTATGGTGCCA	59.2
95	(AGC)8	141	GTCACACTCAGACAGACGCT	59.7	GTGTTCTCACCTGTCTGAGCA	59.9
96	(AGG)7	140	CCTACAGTCATGGCTCAGCA	59.5	TGGAGACAGTAAAGCAGCAGG	60.0
97	(AAT)7	141	GTTTCCTAGGCCAGTCAGCT	59.4	AATGCGACCTTGAGAGCCTT	59.7
98	(AGG)9	144	GAGGTAGCAAGAGGATGGAGG	59.3	TAACTGACCTGCCACTCACC	59.3
99	(AGC)10	138	AGAGCCAAGACTAGCAGTGC	59.8	TGACGGATATCAGCTGTGGC	59.9