Research article

Full-length transcriptome and analysis of bmp-related genes in *Platypharodon extremus*

Xiaoyun Wu, Quan Gong, Yeyu Chen, Ya Liu, Mingjiang Song, Feiyang Li, Pengcheng Li, Jiansheng Lai [*]

The Fishery Institute of the Sichuan Academy of Agricultural Sciences, Chengdu 611730, China

A B S T R A C T

Platypharodon extremus is an endemic species on the Qinghai–Tibet Plateau. As a secondary protected species in China, the basic genomic information of this species has not yet been reported. Here, through third-generation sequencing, the full-length transcriptome of *P. extremus* was obtained. We identified 323,290 CCS sequences, and a total of 50,083 unigenes were extracted after correction with second-generation sequencing data and the removal of redundant reads. A total of 50,067 transcripts were annotated with the various databases. Based on the sequence information, three members in the bone morphogenetic proteins (bmps) family and their receptors, were identified. We found that the special structures of these proteins (zinc-dependent metalloproteinase domain, CUB domains, EGF-like domains and TGF-β domain) are highly conserved in fish and that they are closely evolutionarily related to the bmps and bmp receptors of Cyprinidae fishes. This is the first study to sequence the full-length transcriptome of *P. extremus*, which will help us to further understand its biology.

1. Introduction

Platypharodon extremus is a cold-water fish belonging to Cypriniformes, Cyprinidae, Schizothoracinae, *Platypharodon*; it is a second-class protected fish in China and is listed as an endangered species [1]. *P. extremus* is distributed in the upper reaches of the Yellow River. Over the long-term evolutionary process, this near-polar fish developed strong adaptation to the special environment of high altitude hypoxia, high radiation and low water temperature. However, similar to other Schizothoracinae fishes, it has a short growth period, slow growth rate and late sexual maturity, which greatly limit its population growth. In addition, after the 1950s, due to human population expansion, habitat destruction and overfishing, *P. extremus* resources decreased sharply. As one of the species in the Qinghai–Tibet Plateau, it plays an important role in the complex fish fauna and the biological composition of this freshwater system, and it is of great significance to the balance and stability of the plateau freshwater ecosystem. The research on *P. extremus* to date has focused mainly on its distribution, genetic structure [2], artificial breeding [3] and mitochondrial genome [4], and Zhou et al. [4] conducted a next-generation transcriptome analysis on it. Here, we conducted the transcriptome analysis of *P. extremus* using a combination of SMRT and Illumina sequencing technology.

In recent years, high-throughput third-generation sequencing (TGS) technology such as single-molecule real-time (SMRT) sequencing has been successfully applied to functional genome research and widely used in aquatic animals [5, 6, 7, 8]. Compared with second-generation sequencing (SGS), TGS provides longer read lengths, which makes it very suitable for unsolved problems in genomic, transcriptome and epigenetic research [8]. PacBio® SMRT sequencing generates high fidelity (HiFi) long reads with lengths up to 25 kb with high read accuracy (>99.9%) [9, 10]. These long-read data can cover different exon boundaries to obtain full-length transcripts [9, 11].

Bone morphogenetic proteins (bmps) belong to the TGF-β superfamily, and all BMPs except bmp1 can induce the formation of cartilage and bone [12, 13]. The function of BMPs in ectopic bone formation has been largely confirmed in mammals [14]. Bmps also play roles in cell proliferation, apoptosis and differentiation in various tissues and organs. Among these proteins, bmp1 is a member of the astacin family of metalloproteinases and is involved in activating TGF-β, thereby promoting bone formation [14], and bmp2 has been shown to stimulate bone growth [15]. Moreover, bmp5 has a distinct role in chondrogenesis and may act to modulate the transition of differentiating chondrocytes into a hypertrophic state [16]. Adult Bmp5$^{-/-}$ mice show defects in rib fracture repair [16]. Nevertheless, there are few studies on Bmp5. The earliest

---

report on Bmp5 was related to its mutation in "short ear" mice [17]. Bmps (except bmp1) transduce signals by binding to heteromeric complexes of type 1 and type 2 serine/threonine kinase receptors. In bone formation studies, accumulating evidence suggests that bmp/bmpr1 signaling is indispensable in the regulation of chondrogenesis and endochondral ossification [9]. Kobayashi et al. [18] demonstrated that overactivity of BMP signaling through caBMPR1A in chondrocytes stimulates chondrocyte maturation toward hypertrophic differentiation. Bmpr1a-null mice (Bmpr1a-cKO) show mild generalized chondrodysplasia with shortened long bones and delayed endochondral ossification, as well as smaller and flattened rib cages [14, 15]. The type I receptor is defined by the cytoplasmic GS box and a serine kinase that activates the cellular smad1 and smad5 signaling proteins. The type II receptor activates a type I receptor serine kinase by transphosphorylating the GS box segment [10]. The type II receptor phosphorylates the type I receptor on a Gly/Ser motif, called the GS region, that lies on the amino-terminal side of the kinase domain [19]. As we all know, the growth rate of most cold water fish is relatively slow. The regulatory effect of BMPs on bone is likely to be one of the breakthroughs in further understanding the growth mechanism of *P. extremus*. In this study, we produced a full-length transcriptome of *P. extremus* via SMRT sequencing technology analyzed its functional annotation. With these results, we obtained sequence and annotation information for bmp1, bmp2, bmp5, bmprIa and bmprII among the *P. extremus* transcripts. We predicted and analyzed their gene structure and phylogenetic relationship with other homologous genes. The present study provides basic genomic information for *P. extremus* and valuable analysis of the bone formation and growth of this endangered species.

2. Materials and methods

2.1. Animal welfare

All experimental protocols were approved by the Animal Care Advisory Committee of the Sichuan Academy of Agricultural Sciences (20200418001A).

2.2. Sampling

The fish was found in the White River, which is located in western Sichuan Province, China. Samples of six tissues, namely, brain, head kidney, spleen, liver, intestine and muscle tissue, were collected after anesthesia with M222. After collection, the tissue was quickly frozen in liquid nitrogen and stored in a freezer at −80 °C.

2.3. Total RNA extraction, cDNA library creation and sequencing

Total RNA was extracted with an RNAiso Pure RNA Isolation Kit (TaKaRa). RNA integrity was evaluated by 1% agarose gel electrophoresis. The RNA purity was evaluated according to the OD260/OD280 ratio. The 2100 Agilent Bioanalyzer system (Agilent Technologies, CA, USA) was used to further determine the integrity of the RNA. After extracting from 6 tissues, RNA were mixed and sequenced. The Iso-Seq library was prepared according to the Iso-Seq using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03).

The full-length transcriptome library was constructed by oligo(dT) magnetic bead enrichment. First, oligo(dT) was used to enrich poly(A)-containing mRNA; then, PCR was used to amplify and enrich the synthesized cDNA, with the optimal PCR conditions determined by loop optimization; large-scale PCR was carried out using magnetic beads to obtain a sufficient quantity of cDNA; the full-length cDNA was submitted to end repair, terminal repair, and connection of the SMRT dumbbell junction to construct a full-length transcriptome library; exonuclease digestion was used to remove the unconnected DNA; and finally, a complete SMRT bell library was constructed by combining the prepared template with primers and DNA polymerase.

The PacBio Sequel II platform was used for sequencing according to the effective concentration of the library and our data output requirements.

2.4. Data filtering, assembly and annotation

After sequencing, the original data were trimmed to remove adapter or poly(N) sequences and reads with low quality. The output was filtered and processed by SMRTlink v8.0 software. The circular consensus (CCS) sequence was obtained by using the subreads. bam file in the offline data and the CCS algorithm, that is, self-correction of the single-molecule sequence based on multiple sequence alignment. The flnc (full-length nonchimeric) and NFL (non-full-length nonchimeric) sequences were found by detecting whether the CCS contains 5′ primer, 3′ primer and poly(A) tail sequences. Flnc sequences of the same transcript were clustered by the hierarchical n*log (n) algorithm to obtain consensus sequences. Finally, the full-length sequences were pooled, and a pooled consensus sequence was obtained for subsequent analysis.

The PacBio data were further calibrated with highly accurate Illumina RNA-Seq data (LoRDEC software). CD-HIT was used. The corrected transcripts were clustered to remove redundant transcripts with a threshold of 95% similarity between sequences. Then, genes lacking functional annotation were annotated according to the NR, NT, Pfam, KOG, SwissProt, KEGG and GO databases. Transcript analysis included transcription factor analysis, CD prediction, SSR analysis, and lncRNA prediction.

2.5. SSR analysis

The most common SSRs are dinucleotide repeats, such as $(CA)_n$. MISA software (version 1.0, default parameters) was used in our SSR analysis.

2.6. LncRNA analysis

LncRNAs (long noncoding RNAs) are RNA molecules with transcript lengths of more than 200 nt that do not encode proteins. Due to the limitation of our library building principle, we could obtain only those lncRNAs that contained poly(A) tails. To accurately predict lncRNAs, we used PLEK and CNCI software to predict the coding potential of transcripts. Then, the obtained transcript sequences were processed with CPC 2 software to evaluate their coding potential through the support vector machine classifier according to the biological sequence characteristics of each coding frame of the transcripts. Finally, the predicted transcript sequence was compared with the Pfam-A and Pfam-B databases by hmmscan. After database comparison, the coding potential was predicted more accurately, and a final set of lncRNA sequences was obtained.

2.7. Gene analysis

The sequences of bmp1, bmp2, bmp5, bmpr1a and bmpr2 were obtained according to the annotation results. The molecular weights and isoelectric points of the gene products were calculated using ExPASy tools (http://www.expasy.org/tools). The protein structures were predicted by SMART 4.0 (http://smart.embl-heidelberg.de/). Multiple alignment of the amino acid sequences was performed by using Clustal X. A phylogenetic tree was constructed by the neighbor-joining method in MEGA 6.0 software.

3. Results and discussion

3.1. Full-length transcript analysis

Iso-Seq of *P. extremus* was performed based on the PacBio sequel platform and the raw sequence was deposited into the NCBI Sequence

Table 1. The description of iso-seq of *P.extremus*.

Sample	number	Min_length	Max_length	Mean_length	N50
CCS	323,290	59	15,000	4714	5239
FLNC	240,396	50	14,908	4253	4707
Consensus	73,663	60	14,916	4250	4778
After correct	73,663	60	14,954	4250	4479
unigenes	50,083	63	14,954	4505	5069

Read Archive (SRA) under the accession number PRJNA792686. We obtained 29.46 Gb subreads, with an average sequence length of 4109 bp and an N50 of 4714 bp. In addition, 323,290 CCS were obtained (Table 1). In contrast, 615,874, 765,184, 667,681 and 481,846 CCSs were obtained for *Gymnocypris namensis* (17.65 Gb) [11], *Coilia nasus* [20] (64.18 Gb), *Acipenser dabryanus* [21] (34.26 Gb) and *Hucho bleekeri* [22] (23.1 Gb), respectively. Compared with these previous analyses, the number of CCSs identified here for *P. extremus* was lower because the sequence data were processed using SMRTlink v8.0 software; however, due to the improvements in the SMRT system, the accuracy of the results is greater.

Full-length (FL) reads were defined in SMRTlink software as sequences with primers at both the 3′ and 5′ ends and a poly(A) tail preceding the 3′ primer. A total of 240,396 full-length nonchimeric (FLNC) reads were generated (Table 1). Iso-Seq was performed using a hierarchical n*log (n) algorithm to cluster the FLNC reads of the same transcript to obtain a consensus sequence. The consensus sequence was obtained by using Arrow software, and the fused consensus sequences were obtained for subsequent analysis. Ultimately, 20,762 consensus reads were obtained. The length of the FLNC reads ranged from 50 to 14,908 bp with an average length of 4253 bp and an N50 of 4707 bp (Table 2). Compared with those of the FLNC reads of *Acipenser dabryanus* [21] (using SMRTlink 6.0), *Sebastes schlegelii* [23] (using SMRTlink 6.0) and *Paralichthys olivaceus* [24] (25.45 Gb using SMRTlink 5.1), this average length and N50 are indeed much longer.

3.2. Polymerase read statistics and transcript correction

Before correction, we obtained 73,663 consensus sequences with a mean length of 4250 bp and an N50 of 4778 bp (Table 1). Illumina RNA-Seq data were used for Iso-Seq data correction. After correction, we obtained 73,663 consensus sequences with a mean length of 4250 bp and an N50 length of 4779 bp (Table 1). There was no significant difference before and after correction, similar to what was reported for *Hucho bleekeri* [22]. Then, CD-HIT software was used to remove redundant and similar sequences by sequence alignment clustering, and a total of 50,083 unigenes with a mean length of 4505 bp were generated (Table 1). The unigenes in present study were 35,836 fewer than those of Zhou et al [25], but N50 length was 3395 bp longer than those of Zhou et al [25].

These results also confirm that the library size obtained by Iso-Seq will be smaller than that of NGS, but it has a very long reading length. From this point of view, our research is a supplement, which is based on the results of *P. extremus* [25].

3.3. Functional annotation of transcripts

The functional annotation results are shown in Figure 1: 48,716 (97.27%), 46,112 (92.07%), 48,082 (96.00%), 36,969 (73.82%), 25,632 (51.18%), 50,057 (99.94%), and 25,632 (51.18%) transcripts were matched to the NR, SwissProt, KEGG, KOG, GO, NT and Pfam databases, respectively. A total of 50,067 (99.95%) unigenes were successfully annotated in at least one database, and 20,968 (41.87%) unigenes were annotated in all the databases. As shown in Table 2, among the five fish species, the top three databases in terms of coverage were NT(99.94%), NR(97.27%) and KEGG(96.00%). The annotation results of Zhou et al. [25] are slightly different and the databases accounting for more than 80% are NR(97.13%) and KOG(88.94%). It can be seen that the results obtained by the two sequencing platforms are different.

In comparisons with the NR database, the top three species with the most top matches were *Sinocyclocheilus rhinocerous* (14,252, 29.28%), *Sinocyclocheilus anshuiensis* (12,435, 25.54%), and *Sinocyclocheilus grahami* (8870, 18.22%) (Table 3). After GO (Gene Ontology) annotation, the unigenes were divided into three categories: cellular component, molecular function and biological process (Figure 2A). The KOG is a database of eukaryote genes, divided into 26 classifications according to their functions. The top three KOG classifications were signal transduction mechanisms (T) (9057, 24.50%); general function prediction only (R) (7773, 21.04%); and posttranslational modification, protein turnover, and chaperones (O) (3251, 8.79%) (Figure 2B). KEGG (Kyoto Encyclopedia of Genes and Genomes) integrates genomic, chemical molecule and biochemical system data, including metabolic pathways (KEGG PATHWAY), drugs (KEGG DRUG), diseases (KEGG DISEASE), functional models (KEGG MODULE), gene sequences (KEGG GENES) and genomes (KEGG ORTHOLOG). The KO (KEGG ORTHOLOG) system links the various KEGG annotation systems together. After KO annotation of the unigenes, they were divided into six classifications according to the KEGG metabolic pathways they participate in, including cellular processes, environmental information processing, genetic information processing, human diseases, and metabolism and organismal systems (Figure 2C).

3.4. SSRs (simple sequence repeats)

SSRs are simple sequence repeats, also known as short tandem repeats or microsatellite markers. SSRs consist of tens of nucleotides composed of repeat units of several nucleotides (1–6) and are widely and evenly distributed in eukaryotic genomes. Due to the different nucleotide sequences of the repeat units and the different numbers of repeats, SSR

Table 2. The statistics of transcriptome annotation of five fish species in public database, NR, Swiss prot, KEGG, KOG, GO, NT and Pfam. (Non-Redundant Protein Database; Swiss prot; Kyoto Encyclopedia of Genes and Genomes; Cluster of Orthologous Groups of proteins; Gene Ontology; Non-Redundant Nucleotide Database; Protein family).

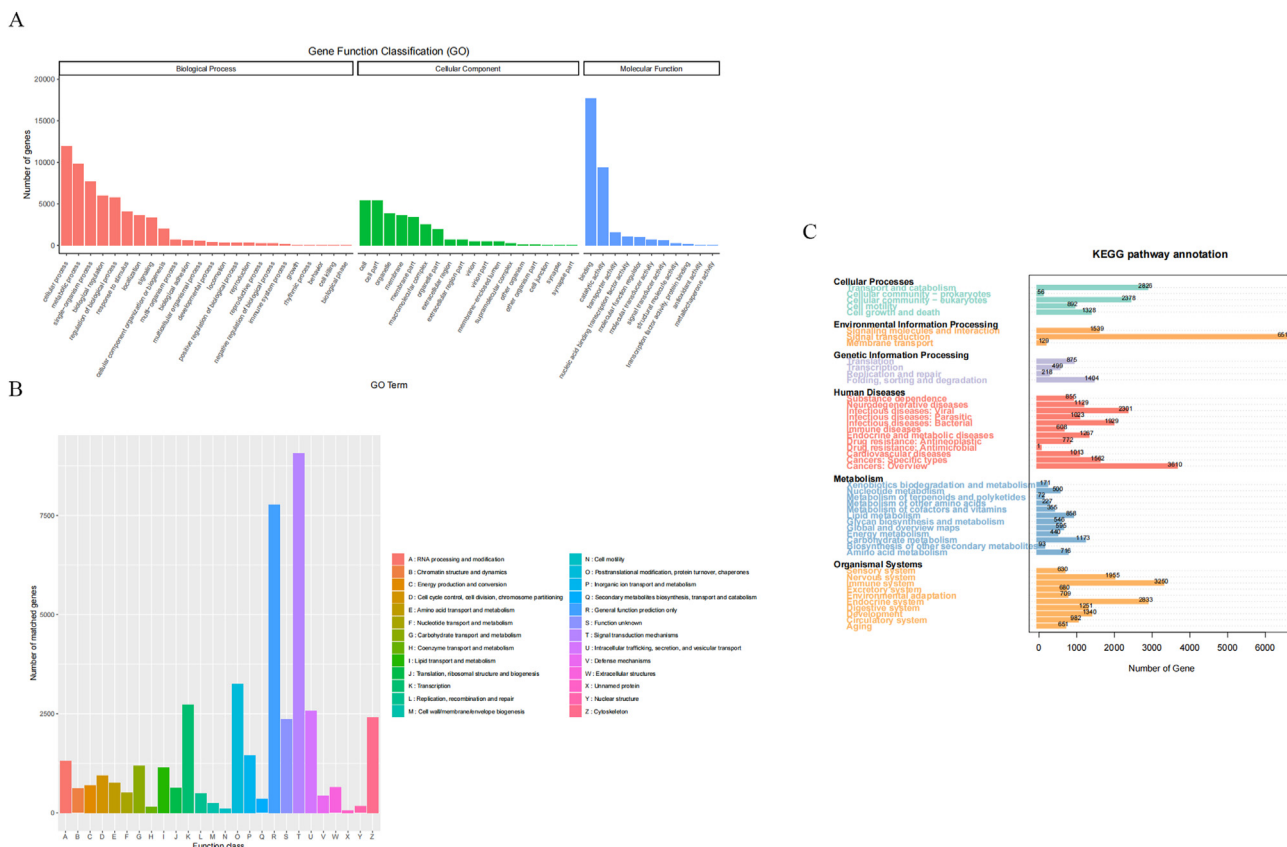| item	method	NR	SwissProt	KEGG	KOG	GO	NT	Pfam
Platypharodon extremus	ISo-Seq NGS	48,716 (97.27%)	46,112 (92.07%)	48,082 (96.00%)	36,969 (73.82%)	25,632 (51.18%)	50,057 (99.94%)	25,8632 (51.18%)
Platypharodon extremus [25]	NGS	46,694 (97.13%)	32,136 (66.85%)	31,231 (46.96%)	42,756 (88.94%)	30,766 (64.00%)	-	-
Coilia nasus [20]	ISo-Seq	84,188 (99.49%)	57,414 (67.85%)	56,530 (66.81%)	63,968 (75.60%)	56,022 (66.21%)	-	72,637 (85.84%)
Hucho bleekeri [22]	ISo-Seq NGS	86,460 (92.64%)	77,768 (83.33%)	83,323 (89.28%)	59,434 (59.8%)	55,808 (59.80%)	91,919 (98.49%)	55,808 (59.8%)
Gymnocypris namensis [11]	ISo-Seq NGS	103,213 (82.31%)	22,110 (17.63%)	69,446 (55.38%)	49,138 (39.19%)	63,926 (50.98%)	-	-
Acipenser dabryanus [21]	ISo-Seq	118,031 (75.98%)	107,468 (69.18%)	115,301 (74.22%)	83,078 (53.48%)	64,693 (41.64%)	108,717 (69.98%)	64,693 (41.64%)

A

B

Figure 1. Annotation in seven databases. Statistical chart of annotation results of seven databases (A); Gene function annotation Venn diagram (B).

Table 3. Non-Redundant Protein Database(NR) database annotation statistics.

NR	Species	Gene number	%
1	*Sinocyclocheilus rhinocerous*	14,252	29.28
2	*Sinocyclocheilus anshuiensis*	12,435	25.54
3	*Sinocyclocheilus grahami*	8870	18.22
4	*Cypinus carpio*	8195	16.83
5	*Dario rerio*	2241	4.60
6	others	2687	5.52

length is highly variable. Excluding single nucleotide repeats, a total of 63,223 SSRs were detected in the transcriptome of *P. extremus*. The most abundant type of repeat motif was the dinucleotide type (52,090,

82.39%), followed by trinucleotide (7900, 12.50%), tetranucleotide (2765, 4.37%), pentanucleotide (258, 0.40%) and hexanucleotide (210, 0.32%) repeats (Figure 3). Similar phenomena were observed in *Gymnocypris namensis* [11] and *Hucho bleekeri* [22].

3.5. LncRNAs

LncRNAs are RNA molecules with transcriptional lengths of more than 200 nt that do not encode proteins. LncRNAs were initially considered 'junk sequences' or 'transcriptional noise', but a growing body of evidence has demonstrated that lncRNAs play pivotal roles in various biological processes, especially in gene expression regulation, including transcriptional regulation, posttranscriptional control and

A

B

C

Figure 2. Gene function annotation in Go, KOG and KEGG databases Go database annotation statistics(A), KOG database annotation statistics(B), Gene function annotation KEGG metabolic pathway classification map(C).
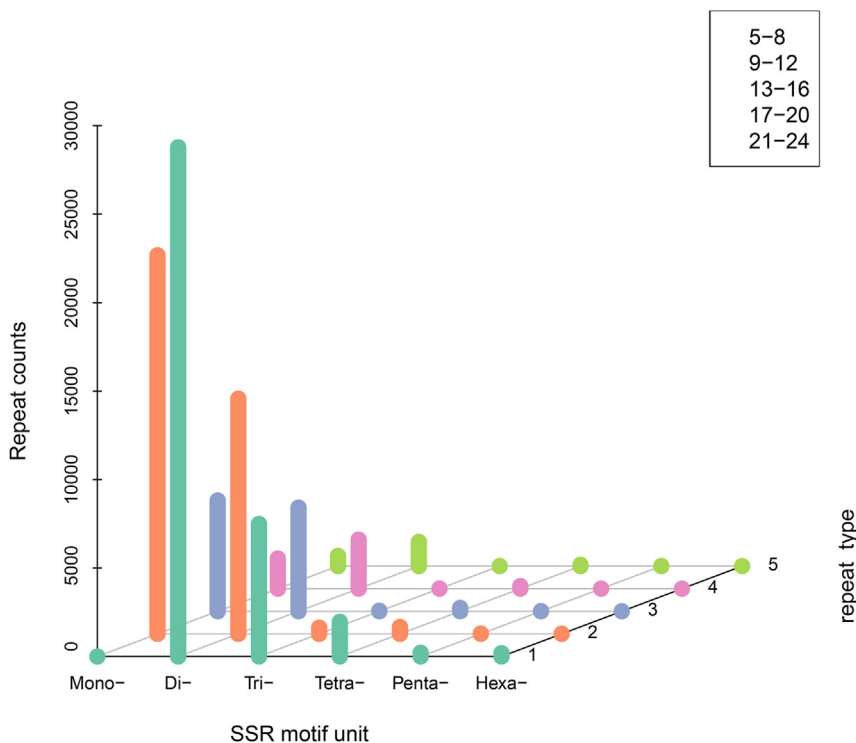
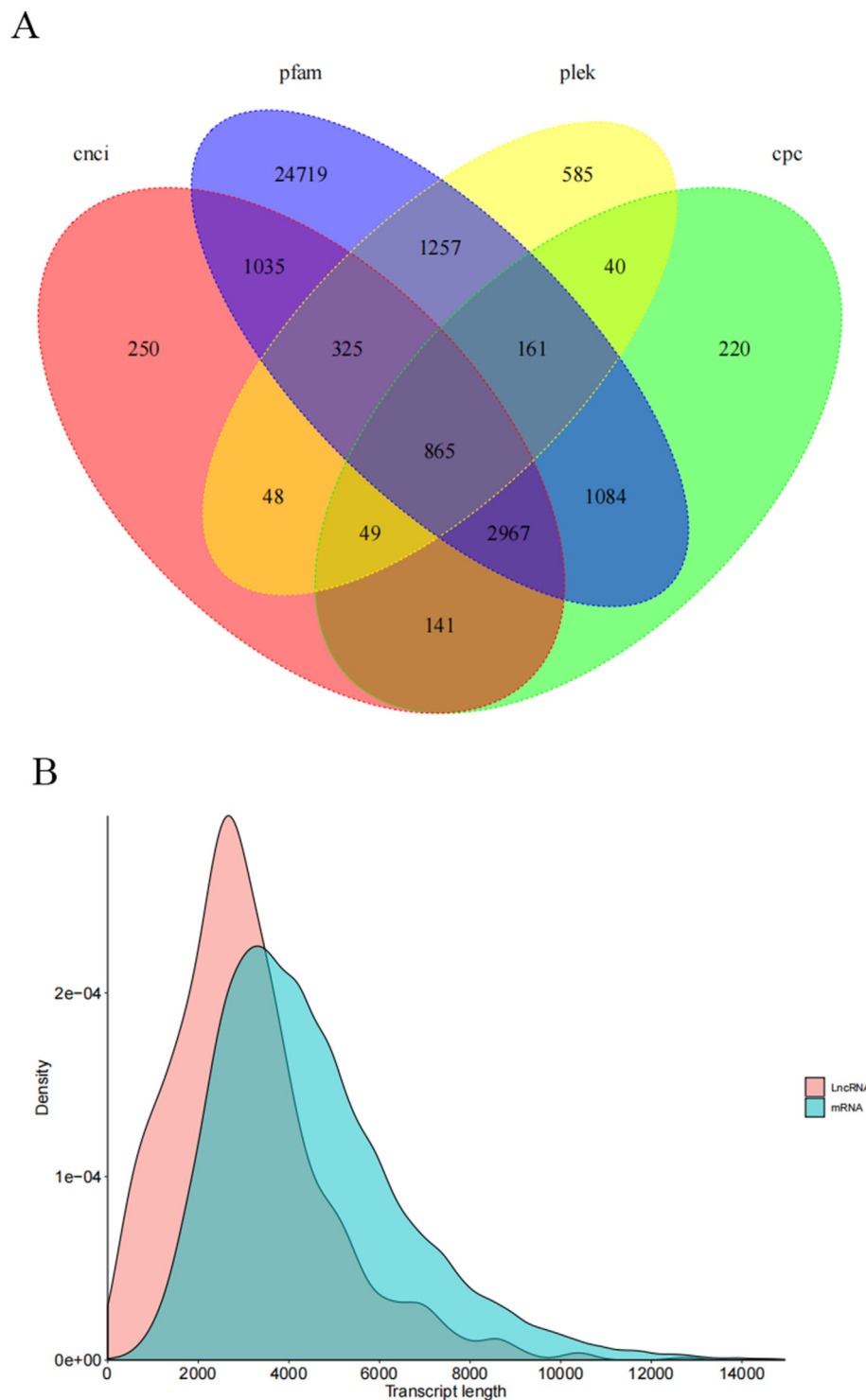# Distribution of SSR Motifs

Figure 3. SSR distribution map.

Figure 4. LncRNA and mRNA length distribution(A), LncRNA and mRNA length distribution(B).

epigenetic processes [26]. A study in *Cynoglossus semilaevis* [27] suggested that lncRNAs may be involved in the immune response of this fish. Another study also found that in rainbow trout (*Oncorhynchus mykiss*), intestinal long noncoding RNAs were modulated by the consumption of functional diets formulated with pre- and probiotics [28]. All these results show that the study of lncRNAs is necessary. We used CNCI, PLEK, and CPC2 software and the Pfam database to predict the coding potential of the transcripts in our PacBio sequencing data (Figure 4A). A total of 865 lncRNAs were identified by all four prediction methods. The longest predicted lncRNA was 12,735 nt long; most of the lncRNAs ranged from 1000 nt to 4000 nt (75.93%) (Figure 4B).

3.6. BMPs and BMPRs

BMP1 can participate in osteogenesis and activates certain TGF-β proteins [29], but the role of BMP1 in bone formation and organogenesis is unclear [30]. As reported by Hartigan et al. [31], BMP1 is encoded by mammalian tolloid (mTLD), which is similar to *Drosophila* tolloid (TLD) and contains five CUB domains and two EGF-like domains at its C-terminus in addition to a zinc-dependent metalloproteinase domain. Our sequencing results indicate that bmp1 from *P. extremus* has a similar structure to its ortholog in zebrafish [32]. The BMP1 gene encoded a protein of 974 amino acids (aa), with a molecular weight of 109,557.55
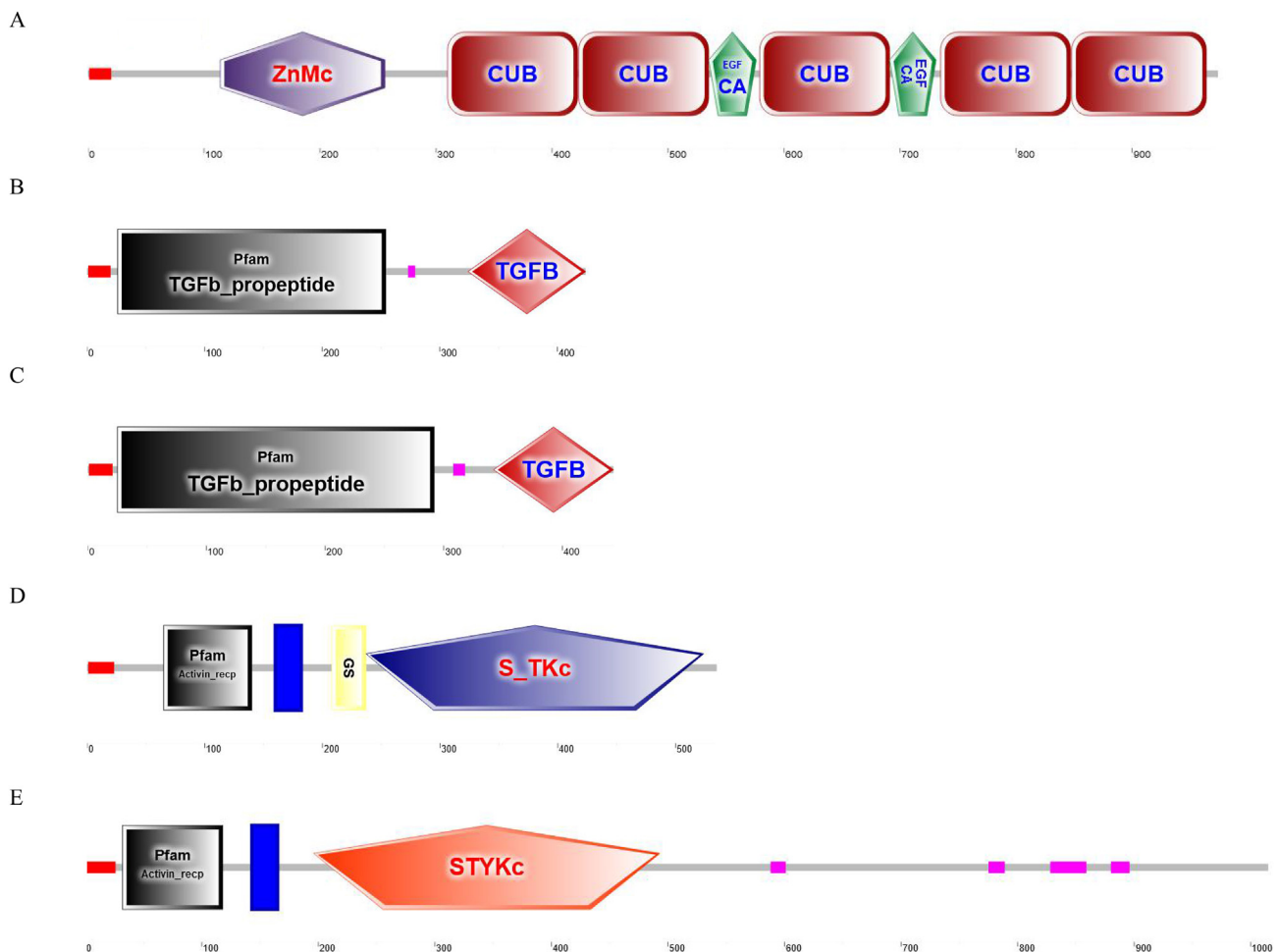
Figure 5. Schematic depicting the domains present in bmp1(A), bmp2(B), bmp5(C), bmpr1a(D) and bmpr2(E) according to SMART analysis.

Da and PI of 5.64. The signal peptide of BMP5 is located in the 1–20 aa region; the zinc-dependent metalloprotease domain is from 114–256 aa; the five CUB domains are at 310–422, 423–535, 579–691, 735–847 and 848–964 aa; and the EGF-like domains are at 535–576 and 691–731 aa (Figure 5A). The zinc-dependent metalloprotease sequence is most similar to the HEXXHXXGXXH ([33]) zinc-binding site/active site. The structure of *P. extremus* BMP1 is shown in Figure S1. The CUB domain (for complement C1r/C1s, Uegf, Bmp1) is a structural motif of approximately 110 residues found almost exclusively in extracellular and plasma membrane-associated proteins, many of which are developmentally regulated [34]. In addition, research in humans found that each CUB domain in BMP-1 has a different function, mainly related to the secretion of BMP1 and the maintenance of its protease activity [31]. This suggests that the CUB in bmp1 of *P. extremus* may have a similar function. The calcium-binding EGF-like domain has Ca^{2+}-binding activity and is followed by another CUB domain [35]. The two EGF-like domains in *P. extremus* bmp1 are interspersed with different CUB domains, similar to the case in zebrafish [32]. In the phylogenetic tree (Figure 6), it is clear that bmp1 clusters separately from bmp2 and bmp5 and thus does not belong to the same family as the other two genes, suggesting that there are differences in function.

Bone morphogenetic protein (BMP)-2 plays a central role in bone tissue engineering because of its potent bone induction capacity [36]. BMP2 can induce osteogenic differentiation and endochondral ossification in mesenchymal stem cells [37], as well as bone healing [38]. Our sequencing results showed that bmp2 encoded a protein of 424 aa with a molecular weight of 47,805.04 Da and a PI of 8.87. SMART predicted that the signal peptide of BMP2 is located at 1–20 aa (20 aa), the TGF-β propeptide domain is at 26–254 aa (229 aa), and the TGF-β domain is located at 324–424 aa (101 aa) (Figure 5B and Figure S2). The coding sequence of *bmp2* in *Sparus aurata* [39] is 436 aa in length and contains a signal peptide of 23 aa, a TGF-β prodomain of 256 aa, and a TGF-β domain of 101 aa. In a study in *Hemibarbus labeo* [40], bmp2a had seven conserved cysteine residues; similar results were found in *Cyprinus carpio* var. Jian [41], *Chlamys farreri* [42] and our study. This shows that the structure of *bmp*2 is highly conserved.

Bmp5 belongs to the 60A subgroup of BMPs and is a member of a large family of secreted signaling molecules that play key roles in axis formation, tissue differentiation, mesenchymal-epithelial interactions, and skeletal development [43]; bmp5 also plays a role in regulating the dendritic growth of sympathetic neurons [44]. Another report showed that *bmp5* is expressed in precise domains in developing muscle masses and in autopodial tendons [45]. As reported by Zuzarte-Luís et al. [45], the application of BMP5 in differentiating cartilage promotes dramatic growth in fertilized chicken embryos. In the present study, *P. extremus* bmp5 encoded a protein of 443 aa with a molecular weight of 49,984.99 Da and a PI of 9.31. The signal peptide of bmp5 is located at 1–21 aa, the TGF-β propeptide domain is at 25–292 aa, and the TGF-β domain is located at 342–443 aa (Figure 5C and Figure S3). Similar to bmp2, it has a conserved structure of the TGF beta family. There are few studies on bmp5 in aquatic animals. Huang et al [46] showed that zebrafish bmp5 is expressed in neural crest progenitor cells, which give rise to many ectodermal and mesodermal derivatives, and that it activates the Smad and Erk signaling pathways to regulate cell survival and proliferation, respectively. On the other hand, its receptor plays a key role in the process of BMP signal transduction.
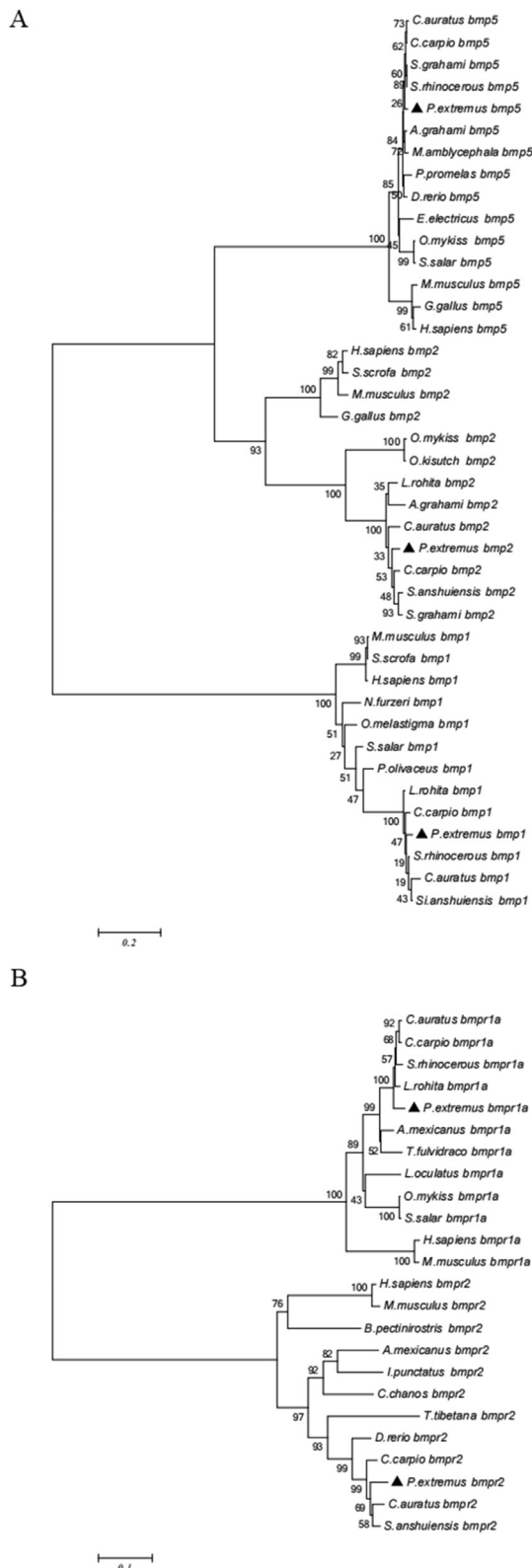
A

B

Figure 6. Phylogenetic tree of bmp1, bmp2, bmp5, bmpr1a and bmpr2. (A: phylogenetic tree analysis of bmp1, bmp2, bmp5; B: phylogenetic tree analysis of bmpr1a and bmpr2).

BMPs are members of the TGF-β superfamily and transduce signals by binding to heteromeric complexes of type 1 and type 2 serine/threonine kinase receptors. Accumulating evidence suggests that BMP/BMPR1 signaling is indispensable in the regulation of chondrogenesis and endochondral ossification [47]. Kobayashi et al. [48] demonstrated that overactivity of BMP signaling through caBMPR1A in chondrocytes stimulates chondrocyte maturation toward hypertrophic differentiation. Bmpr1a-null mice (Bmpr1a-cKO) show mild generalized chondrodysplasia with shortened long bones and delayed endochondral ossification, as well as smaller and flattened rib cages [47, 49]. *P. extremus* bmp1ra encodes a protein of 535 aa with a molecular weight of 59,980.86 Da and a PI of 7.08. The bmp1ra protein contains a signal peptide located at 1–23 aa, a Pfam activin receptor type domain at 65–140 aa, a transmembrane region at 160–182 aa, a GS domain at 208–237 aa and a serine/threonine protein kinase catalytic domain at 237–524 aa (Figure 5D and Figure S4). *P. extremus* bmpr2 encodes a protein of 1015 aa with a molecular weight of 110,720.48 Da and a PI of 5.72. The bmpr2 protein contains a signal peptide at 1–26 aa, a Pfam activin receptor type domain at 32–118 aa, a transmembrane region at 143–165 aa, and a serine/threonine protein kinase catalytic domain at 196–493 aa (Figure 5E and Figure S5). The type 1 receptor is defined by the cytoplasmic GS box and a serine kinase that activates cellular smad1 and smad5 signaling proteins. Both smad1 and smad5 have been shown to play important roles in the osteogenesis of C2C12 cells [50]. The type 2 receptor phosphorylates the type 1 receptor at the GS region [51, 52, 53]. BMPs combined with BMPR2 can initiate the smad signaling pathway [51]. There have been few reports on BMPs and their receptors in fish. Nine bmp genes of blunt snout bream (*Megalobrama amblycephala*) were cloned by Zhang [54], and their expression patterns in four different growth and development stages of intermuscular bone were analyzed.

4. Conclusion

This study is the first to analyze the isoform sequencing of *P. extremus*. To a large extent, this is a supplement to the research of Zhou et al. [25]. This analysis will provide us with the biological data necessary to further understand this fish. Bmp is an important factor regulating bone growth, and we hope to explore and utilize the functions of these genes to determine the growth mechanism of this cold-water fish in the alpine region.

Declarations

Data availability statement

Data associated with this study has been deposited at NCBI under the accession number PRJNA792686.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e10783.

References

[1] S. Wang, G.M. Zheng, Q.S. Wang, China Red Data Book of Endangered Animals: Pisces, 1998.

[2] J. Su, W. Ji, Y. Zhang, D.M. Gleeson, Z. Lou, J. Ren, Y. Wei, Genetic diversity and demographic history of the endangered and endemic fish (*Platypharodon extremus*): implications for stock enhancement in Qinghai Tibetan Plateau, Environ. Biol. Fish. 98 (2015) 763–774.

[3] Y.P. Zhang, Z.Y. Lou, J.H. Su, W.L. Jiao, T. Wang, X.F. Gong, S.H. Qi, J.P. Wu, Y.M. Wei, Artificial propagation of *Platypharodon extremus*, J. Fish. China 34 (2010) 1698–1703.

[4] C. Li, W. Xiong, Y. Chen, D. He, C. Li, W. Xiong, Y. Chen, D. He, The complete mitochondrial genome sequence of *Platypharodon extremus* (Cypriniformes: Cyprinidae), Mitochondrial DNA 27 (2016) 1430–1431.

[5] X. Zhang, L. Pan, R. Tong, Y. Li, Y. Tian, D. Li, L. Si, PacBio full length transcript sequencing and Illumina transcriptome insight into immune defense mechanism of Litopenaeus vannamei under ammonia-N stress, Aquaculture 536 (2021), 736457.

[6] A. Yx, B. Yla, B. Xla, L.A. Chao, Full-length transcriptome sequencing from multiple immune-related tissues of Paralichthys olivaceus, Fish Shellfish Immunol. 106 (2020) 930–937.

[7] R. Mehjabin, X. Lv, H. Rong, Y. Cheng, C. Geng, Full-length transcriptome sequencing and the discovery of new transcripts in the unfertilized eggs of zebrafish (*Danio rerio*), G3-Genes Genomes Genet. 9 (2019) 1831–1838.

[8] S. Yi, X. Zhou, J. Li, M. Zhang, S. Luo, Full-length transcriptome of Misgurnus anguillicaudatus provides insights into evolution of genus Misgurnus, Sci. Rep. 8 (2018), 11699.

[9] A. Rhoads, K.F. Au, PacBio sequencing and its applications, Dev. Reprod. Biol. 13 (2015) 278–289.

[10] L.L. Wong, S.A. Razali, Z.M. Deris, M. Danish-Daniel, Application of second-generation sequencing (SGS) and third generation sequencing (TGS) in aquaculture breeding program, Aquaculture 548 (2022), 737633.

[11] H. Luo, Full-length transcript sequencing accelerates the transcriptome research of *Gymnocypris namensis*, an iconic fish of the Tibetan Plateau, Sci. Rep. 10 (2020).

[12] Wei-Zhuo Zhang, Tian Lan, Chun-Hong Nie, Ning-Nan Guan, Ze-Xia Gao, Characterization and spatiotemporal expression analysis of nine bone morphogenetic protein family genes during intermuscular bone development in blunt snout bream, Gene 642 (2018) 116–124.

[13] W. Mei, C. Xu, BMP signaling in skeletal development, Biochem. Biophys. Res. Commun. 328 (2005) 651–657.

[14] Y. Nakamura, S. Wakitani, J. Nakayama, S. Wakabayashi, H. Horiuchi, K. Takaoka, Temporal and spatial expression profiles of BMP receptors and Noggin during BMP-2-induced ectopic bone formation, J. Bone Miner. Res. 18 (2010) 1854–1860.

[15] J. Gill, P. Connolly, M. Roth, S.H. Chung, D.S. Geller, The effect of bone morphogenetic protein-2 on osteosarcoma metastasis, PLoS One 12 (2017), e0173322.

[16] S. Loughlin, 179 Characterizing the function of bone morphogenetic protein-5(BMP5) in chondrocytes, Osteoarthritis Cartilage 15 (2007) C107.

[17] BMP5 and the molecular, skeletal, and soft-tissue alterations in short ear mice, Dev. Biol. 166 (1994) 112–122.

[18] F. Delolme, C. Anastasi, L.B. Alcaraz, V. Mendoza, C. Moali, Proteolytic control of TGF-β co-receptor activity by BMP-1/tolloid-like proteases revealed by quantitative iTRAQ proteomics, Cell. Mol. Life Sci. Cmls 72 (2015) 1009–1027.

[19] A.M. Wenger, P. Peluso, W.J. Rowell, P.C. Chang, M.W. Hunkapiller, Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome, Nat. Biotechnol. 37 (2019) 1155–1162.

[20] B. Jza, B. Sga, C. Ys, C. Yy, B. Qla, Full-length transcriptome of anadromous *Coilia nasus* using single molecule real-time (SMRT) sequencing, Aquacul. Fish. (2020).

[21] Y. Chen, X. Wu, J. Lai, Y. Liu, M. Song, F. Li, Q. Gong, Full-length transcriptome sequencing and identification and immune response of TRIM genes in Dabry's sturgeon (*Acipenser dabryanus*), Aquaculture 538 (2021) 736599.

[22] Y. Chen, H. Yang, Y. Chen, M. Song, H. Li, Full-length transcriptome sequencing and identification of immune-related genes in the critically endangered Hucho bleekeri, Dev. Comp. Immunol. 116 (2021), 103934.

[23] M. Cao, M. Zhang, N. Yang, Q. Fu, C. Li, Full length transcriptome profiling reveals novel immune-related genes in black rockfish (*Sebastes schlegelii*), Fish Shellfish Immunol. 106 (2020) 1078–1086.

[24] A. Yx, B. Yla, B. Xla, L.A. Chao, Full-length transcriptome sequencing from multiple immune-related tissues of Paralichthys olivaceus, Fish Shellfish Immunol. 106 (2020) 930–937.

[25] C.W. Zhou, S.J. Xiao, Y.C. Liu, Z.B. Mou, J.S. Zhou, Y.Z. Pan, C. Zhsng, J. Wang, X.X. Deng, M. Zou, H.P. Liu, Comprehensive transcriptome data for endemic Schizothoracinae fish in the Tibetan Plateau, Sci. Data 28 (2020).

[26] N. Yang, B. Wang, Z. Yu, X. Liu, Q. Fu, M. Cao, T. Xue, Y. Ren, F. Tan, C. Li, Characterization of a novel lncRNA (SETD3-OT) in turbot (*Scophthalmus maximus* L.), Fish Shellfish Immunol. 102 (2020) 145–151.

[27] S. Wei, Y. Chen, L. Huang, H. Ma, Z. Sha, Analysis of lncRNA and mRNA expression profiles in peripheral blood leukocytes of the half-smooth tongue sole (*Cynoglossus semilaevis*) treated with chitosan oligosaccharide, Dev. Comp. Immunol. 120 (2021), 104043.

[28] G. Nuez-Acua, C. Detrée, C. Gallardo-Escárate, A.T. Gonalves, Functional diets modulate lncRNA-coding RNAs and gene interactions in the intestine of rainbow trout Oncorhynchus mykiss, Mar. Biotechnol. 19 (2017) 287–300.

[29] D.R. Hopkins, S. Keles, D.S. Greenspan, The bone morphogenetic protein 1/Tolloid-like metalloproteinases, Matrix Biol. 26 (2007) 508–523.

[30] Efrat Kessler, Takahara, Kazuhiko, Bone morphogenic protein-1; The type I procollagen c-proteinase, Science 271 (1996) 360.

[31] N. Hartigan, L. Garrigue-Antar, K.E. Kadler, Bone morphogenetic protein-1 (BMP-1), J. Biol. Chem. 278 (2003) 18045–18049.

[32] R. Jasuja, N. Voss, G. Ge, G.G. Hoffman, J. Lyman-Gingerich, F. Pelegri, D.S. Greenspan, bmp1 and mini fin are functionally redundant in regulating formation of the zebrafish dorsoventral axis, Mech. Dev. 123 (2006) 548–558.

[33] P. Michael, Sarras, BMP-1 and the astacin family of metalloproteinases: a potential link between the extracellular matrix, growth factors and pattern formation, Bioessays 18 (1996) 439–442.

[34] Danilo Predes, Joo Victor R. Cruz, Jose G. Abreu, and Fabio, CUB domain-containing protein 1 (CDCP1) binds transforming growth factor beta family members and increase TGF-β1 signaling pathway - ScienceDirect. Exp. Cell Res. 383, 111499.

[35] S.D. Reynolds, D. Zhang, J.E. Puzas, R.J. O'Keefe, P.R. Reynolds, Cloning of the chick BMP1/Tolloid cDNA and expression in skeletal tissues, Gene 248 (2000) 233–243.

[36] K. Hashimoto, T. Kaito, M. Furuya, S. Seno, M. Ishii, In vivo dynamic analysis of BMP-2-induced ectopic bone formation, Sci. Rep. 10 (2020) 4751.

[37] S.Z. Karoulias, M. Pitou, R. Papi, P. Lamprou, and T. Choli-Papadopoulou, Specific amino acids from the broad C-terminal region of BMP-2 are crucial for osteogenesis. BoneKEy Rep. 14(2021), 101092.

[38] K. Tsuji, A. Bandyopadhyay, B.D. Harfe, K. Cox, S. Kakar, L. Gerstenfeld, T. Einhorn, C.J. Tabin, V. Rosen, BMP2 activity, although dispensable for bone formation, is required for the initiation of fracture healing, Nat. Genet. 38 (2006) 1424–1429.

[39] M.S. Rafael, V. Laizé, M.L. Cancela, Identification of Sparus aurata bone morphogenetic protein 2: molecular cloning, gene expression and in silico analysis of protein conserved features in vertebrates, Bone 39 (2006) 1373–1381.

[40] J. Chen, Y.P. Lv, Q.M. Dai, L. Zhang, C.B. Xu, J. Lu, Molecular characterization of a bmp2a homologue in barbel steed (*Hemibarbus Labeo*) and its involvement in intermuscular bone development, Acta Hydrobiol. Sin. 45 (2021) 8–12.

[41] L.X. Ma, Z.J. Dong, S.Y. Su, J.Q. Zhang, W. Liu, L.L. Li, X.H. Yuan, Cloning of fragments of bone morphogenetic proteigen(BMP2B) of Cyprinus carpio vat. Jian and its expression analysis, Jiangsu J. Agric. Sci. 29 (2013) 370–378.

[42] L.Y. Feng, H.H. Guo, X. Li, Q. Yu, X.L. Hu, L.L. zhang, S. Wang, Z.M. Bao, Cloning and expression analysis of bone morphogenetic protein 2 gene of *Chlamys farreri*, Periofocal Ocean Univ. China 43 (2013) 48–55.

[43] Paul Marker, C. Seung, Kwonjune, Spectrum of Bmp5 mutations from germline mutagenesis experiments in mice, Genetics 145 (1997) 435–443.

[44] H.N. Beck, K. Drahushuk, D.B. Jacoby, D. Higgins, P.J. Lein, Bone morphogenetic protein-5 (BMP-5) promotes dendritic growth in cultured sympathetic neurons, BMC Neuroence 2 (2001) 1–11.

[45] V. Zuzarte-Lu, S.J.A. Montero, J. Rodriguez-León, R. Merino, J.C. Rodr Guez-Rey, J.M. Hurlé, A new role for BMP5 during limb development acting through the synergic activation of Smad and MAPK pathways, Dev. Biol. 272 (2004) 39–52.

[46] Shih Hung-Yu, Shu-Yuan, Pin Hsu, Sheng-Jia Ouyang, Ting-Yun Lin, Chou, Bmp5 regulates neural crest cell survival and proliferation via two different signaling pathways, Stem Cells 35 (2017) 1003–1014.

[47] Huabing, Min Qi, Yaqi Jin, Xiaolan Duan, Yuanquan Du, Zhang, FGFR3 induces degradation of BMP type I receptor to regulate skeletal development, Biochim. Biophys. Acta Mol. Cell Res. 1843 (2014) 1237–1247.

[48] T. Kobayashi, K.M. Lyons, A.P. Mcmahon, H.M. Kronenberg, BMP signaling stimulates cellular differentiation at multiple steps during cartilage development, Proc. Natl. Acad. Sci. U. S. A 102 (2005) 18023–18027.

[49] B.S. Yoon, D.A. Ovchinnikov, I. Yoshii, Y. Mishina, R.R. Behringer, K.M. Lyons, Bmpr1a and Bmpr1b have overlapping functions and are essential for chondrogenesis in vivo, Proc. Natl. Acad. Sci. U. S. A 102 (2005) 5062–5067.

[50] P.T. Dijke, O. Korchynskyi, G. Valdimarsdottir, M.J. Goumans, Controlling cell fate by bone morphogenetic protein receptors, Mol. Cell. Endocrinol. 211 (2003) 105–113.

[51] T. Kirsch, BMP-2 antagonists emerge from alterations in the low-affinity binding epitope for receptor BMPR-II, EMBO J. 19 (2014) 3314–3324.

[52] G. Sanchez-Duffhues, E. Williams, M.J. Goumans, C.H. Heldin, P.T. Dijke, Bone morphogenetic protein receptors: structure, function and targeting by selective small molecule kinase inhibitors, Bone 138 (2020), 115472.

[53] J.L. Wrana, TGFß Signal Transduction. Handbook of Cell Signaling, 2003.

[54] W.Z. Zhang, Regulatory Rolas of Bone Morphogenetic Protein Family Genes during Intermuscular Bone Development in Bunt Snout Bream, HuaZhong Agricultural University, 2018.