



NAZARBAYEV
UNIVERSITY

**Characterization and Comparative Analysis of Bone Morphogenetic
Proteins (BMPs) in Maral (*Cervus canadensis sibiricus*)**

Dinara Askar

BSc in Biological Sciences, Nazarbayev University

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF SCIENCE IN LIFE SCIENCES DEPARTMENT OF
BIOLOGY SCHOOL OF SCIENCE AND HUMANITIES NAZARBAYEV UNIVERSITY

2026

Student: Dinara Askar

Submission Date: 12/April/2026

Student's Supervisor/Advisor: Ferdinand Molnar

Approval Date: 10/April/2026

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.



Dinara Askar

April 12, 2026

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Ferdinand Molnar, for his continuous support, guidance and encouragement throughout the course of this research. His insights and feedback were invaluable in shaping both the direction and quality of this work.

I am also thankful to the laboratory of Ainur Akilzhanova for their contribution to the sequencing of the maral genome, which formed the foundation of this study.

Special thanks to Rakhimbek Bektayev for his guidance and assistance with bioinformatics pipelines. His expertise helped me navigate many of the technical challenges encountered during the project.

Finally, I would like to express my heartfelt appreciation to my mother and boyfriend for their constant support, patience, and belief in me.

TABLE OF CONTENTS

DECLARATION	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
1. INTRODUCTION.....	1
1.1 Background on Tissue Regeneration in Mammals	1
1.2 Antler Regeneration: A Unique Mammalian Regeneration Model	1
1.3 Bone Morphogenetic Proteins: Overview and Significance	2
1.4 Introduction to Maral (<i>Cervus canadensis sibiricus</i>).....	3
1.5 Research Gap and Thesis Aims	3
2. MATERIALS & METHODS.....	4
2.1 Sample Collection	4
2.2 Genome Assembly and Evaluation.....	4
2.3 Mitochondrial DNA-Based Phylogenetic Analysis.....	4
2.4 Extraction of Bone Morphogenetic Protein (BMP) Sequences	5
2.5 Functional Annotation and Structural Prediction	5
2.6 Comparative Sequence and Phylogenetic Analysis.....	6
2.7 Structural and Surface Property Analysis	6
3. RESULTS	7
3.1 Assembly contiguity and structural metrics (QUAST)	7
3.2 Gene-space completeness (BUSCO).....	9
3.3 Phylogenetic Placement of Maral within <i>Cervus</i>	11
3.4 Functional annotation via InterPro	11
3.5 Comparative Sequence Analysis of BMPs	14
3.6 Structural Annotation	18
3.7 Conservation Analysis	21
3.8 Surface Hydrophobicity analysis	23
4. DISCUSSION	26
5. CONCLUSION.....	28
6. REFERENCE LIST	29
7. APPENDICES	32

1. INTRODUCTION

1.1 Background on Tissue Regeneration in Mammals

Tissue regeneration represents one of the most fascinating and clinically significant biological processes in mammals. Unlike lower vertebrates such as salamanders and zebrafish, which have remarkable regenerative capabilities that allow them to regrow entire limbs and organs, most mammals have limited regenerative potential. This evolutionary constraint has profound implications for human health, as the inability to regenerate complex tissues following injury or disease requires extensive medical interventions and often results in permanent functional impairment.

Bone tissue stands as a notable exception to the regenerative capacity of mammals due to the unique ability to heal and remodel throughout life. This regeneration is driven by a coordinated set of cellular activities, growth factors, and signaling molecules that work together to recruit osteoprogenitor cells and guide their growth and differentiation (Salazar et al., 2016). Among the molecular regulators of bone regeneration, Bone Morphogenetic Proteins (BMPs) have emerged as critical mediators, functioning as potent osteoinductive factors that can initiate the entire cascade of bone formation even in non-skeletal tissues (Carreira et al., 2015). Understanding the molecular mechanisms underlying bone regeneration has become increasingly important for developing therapeutic strategies to treat bone defects, fractures, and degenerative skeletal diseases.

1.2 Antler Regeneration: A Unique Mammalian Regeneration Model

Antlers represent the only mammalian structures capable of complete annual regeneration, making them an exceptional model for studying tissue regeneration in mammals. During the growth phase, antlers can elongate at rates exceeding 2 centimeters per day, representing one of the fastest-growing tissues in the animal kingdom (Feleke et al., 2021a).

Antlers are also biologically distinctive in their tissue composition and growth environment. Unlike horns, which are generally permanent cranial appendages characterized by a keratin covering, antlers are bony structures that regrow from permanent cranial outgrowths called pedicles. During growth, antlers are living organs supported by “velvet,” a specialized integument that is richly vascularized and innervated and that supplies oxygen, nutrients, and sensory support during rapid tissue expansion (Nasoori, 2020).

Regeneration initiates at the pedicle after casting. Wound closure at the pedicle stump is followed by establishment of a proliferative growth center and activation of regeneration-competent stem and progenitor populations, rather than thickening of the previously mineralized hard antler. Antler renewal is therefore increasingly characterized as a stem cell–based epimorphic process. (Kierdorf & Kierdorf, 2011)

At the tissue level, antler elongation largely follows endochondral ossification. A distal cartilaginous template forms and is progressively replaced by bone as chondrocytes mature, blood vessels invade, and osteoblast-lineage cells deposit and mineralize extracellular matrix (Feleke et al., 2021a).

1.3 Bone Morphogenetic Proteins: Overview and Significance

Bone Morphogenetic Proteins constitute a family of highly conserved secreted signaling molecules belonging to the Transforming Growth Factor- β (TGF- β) superfamily. Originally discovered by Marshall Urist in 1965 through their ability to induce ectopic bone formation when implanted in muscle tissues, BMPs have since been recognized as multifunctional regulators involved in diverse developmental and physiological processes (Sampath & Reddi, 2020). The BMP family consists of more than 20 members in mammals, classified into several subfamilies based on structural and functional similarities (Carreira et al., 2015).

BMPs exert their biological effects by binding to heteromeric complexes of type I and type II serine/threonine kinase receptors on the cell surface. This leads to phosphorylation of SMAD1/5/9, which subsequently form complexes with Smad4. These complexes translocate to the nucleus, where they regulate the expression of genes involved in mesenchymal cell fate, proliferation, and differentiation. (Lowery & Rosen, 2018).

The osteoinductive properties of BMPs have been extensively validated in preclinical and clinical studies, leading to the approval of recombinant human BMP2 (rhBMP2) and BMP7 (rhBMP7) by regulatory agencies for specific orthopedic applications, including spinal fusion, tibial nonunion fractures, and maxillary sinus augmentation (McKay et al., 2007).

Within antler regeneration, BMP signaling is consistently implicated as part of a broader osteochondral program that supports rapid endochondral progression and ossification. Cross-species transcriptomic analyses and antler growth-center multi-omics studies commonly report enrichment of pathways related to chondrogenesis and osteogenesis that include BMP pathway components (Ba et al., 2022). Understanding how BMPs orchestrate these complex processes in the context of antler regeneration may provide insights into enhancing bone regeneration in clinical settings.

1.4 Introduction to Maral (*Cervus canadensis sibiricus*)

The Maral (*Cervus canadensis sibiricus*), also known as the Siberian wapiti or Altai wapiti, is a subspecies of elk native to the mountainous regions of Central Asia, including Kazakhstan, Mongolia, and parts of Russia and China.

Beyond its biological significance, the maral holds ecological, cultural, and economic importance in the region. In Kazakhstan, maral farming has developed as a specialized agricultural practice, largely driven by the demand for “pant deer velvet”, a soft, living tissue that covers the growing antler of deer before they harden into bone. Velvet antlers are harvested before mineralization and are widely used in traditional medicine, particularly in East Asia, due to their reported regenerative and pharmacological properties (Sui et al., 2014). The velvet covering the growing antlers is richly vascularized and innervated, providing the necessary nutrients and growth factors to support the underlying cartilage and bone formation.

1.5 Research Gap and Thesis Aims

While BMPs have been extensively studied in model organisms such as mice, zebrafish, and humans, and their roles in bone formation and regeneration are well established, there is little information regarding BMP genes and proteins in species with exceptional regenerative capabilities, particularly cervids like the maral.

The primary objective of this thesis is to characterize and comparatively analyze Bone Morphogenetic Proteins in the maral (*Cervus canadensis sibiricus*), with a focus on understanding their molecular features and evolutionary relationships to BMPs in other mammalian species. This research aims to address the knowledge gap regarding maral BMPs and contribute to our understanding of the molecular basis of antler regeneration.

The specific aims of this study are:

1. To sequence the whole maral genome and perform genome annotation with the goal of identifying bone morphogenetic proteins.
2. To characterize maral BMPs and perform a comprehensive comparative analysis with BMPs from humans and other relevant species.
3. To identify and analyze unique features of maral BMPs that may contribute to their exceptional antler regeneration capabilities.

2. MATERIALS & METHODS

2.1 Sample Collection

Peripheral blood samples (2 mL each) were collected from four maral individuals at a maral farm located in the eastern region of Kazakhstan. The samples were treated with EDTA as an anticoagulant and stored at 4 °C until further processing. Genomic DNA was extracted using the Illustra Blood Kit (GE Healthcare/Cytiva) according to the manufacturer's protocol. The concentration and purity of the extracted DNA were evaluated using a Qubit fluorometer (Invitrogen, USA), a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA), and 1% agarose gel electrophoresis. High-quality, high-molecular-weight DNA was then used for library preparation and Oxford Nanopore sequencing.

2.2 Genome Assembly and Evaluation

Draft genome assemblies were generated using Flye v2.9.2 (Kolmogorov et al., 2019) and subsequently polished with Oxford Nanopore Technologies (ONT) reads using Medaka v1.11.1. Assembly quality was evaluated with QUAST v5.2.0 (Gurevich et al., 2013) by aligning the results against the CerElai1.1 reference genome. Genome completeness was further assessed using BUSCO v5.8.3 (Simão et al., 2015) with the *laurasiatheria_odb10* ortholog database, which contains 12,234 conserved single-copy genes representative of Laurasiatherian mammals.

2.3 Mitochondrial DNA-Based Phylogenetic Analysis

To investigate the evolutionary relationships of the maral, a phylogenetic tree was reconstructed using mitochondrial DNA (mtDNA) sequences. The complete mtDNA sequence of the maral was assembled from whole-genome sequencing data using a reference-guided approach. A high-quality mitochondrial reference genome of a closely related *Cervus elaphus* was downloaded from the NCBI database and used to identify homologous regions via BLASTn. Identified regions were extracted and concatenated using BEDTools and custom scripts. For phylogenetic comparison, mitochondrial genomes of multiple cervid species were retrieved from GenBank. All sequences were aligned using MAFFT v7.505 under default parameters. A maximum likelihood phylogenetic tree was constructed in MEGA11 v.11.0.13 (Tamura et al., 2021) using the Tamura-Nei model with 1000 bootstrap replicates to assess node support.

2.4 Extraction of Bone Morphogenetic Protein (BMP) Sequences

BMP gene sequences were identified and extracted from the maral genome assembly using Exonerate v2.4.0, which performs spliced alignments of protein queries against a genome assembly. Protein sequences of BMP2 (UniProt: O19006) and BMP4 (UniProt: Q29607) were obtained from UniProtKB using reviewed Swiss-Prot entries. For BMP3 (NCBI: XP_043762314.1), BMP7 (NCBI: XP_043740578.1), and BMP9 (NCBI: XP_061019055.1), sequences were retrieved from the NCBI GenBank database due to the limited availability of reviewed UniProt entries for cervid species. The selected BMP set was chosen to represent both classical osteogenic BMPs (BMP2, BMP4, BMP7) and functionally distinct members (BMP3 and BMP9), enabling comparative analysis across conserved and divergent BMP subgroups. All selected sequences were used as query inputs, while the maral genome assembly served as the target database.

The protein2genome model of Exonerate was employed with the `–showtargetgff` and `–showalignment` options enabled to retrieve both the predicted gene structure and corresponding alignment. Output files were parsed to extract genomic coordinates and coding regions from gff file. BEDTools and Seqtk were used to extract the matched genomic sequences, which were then translated into protein sequences using Biopython.

The extracted BMP protein sequences were verified via pairwise alignment against their reference counterparts using EMBOSS Needle to confirm the accuracy of the predicted open reading frames and to assess conservation.

2.5 Functional Annotation and Structural Prediction

Functional annotation of predicted maral BMP protein sequences was performed using InterProScan, which integrates multiple protein signature databases (Pfam, SMART, SUPERFAMILY, and PROSITE) to identify conserved domains, motifs, and functional sites. Particular attention was given to detection of the canonical TGF- β superfamily domain and cystine-knot cytokine motif characteristic of BMP ligands.

Three-dimensional structures of selected BMP proteins were predicted using ColabFold, an accelerated implementation of AlphaFold2. Default parameters were applied, and multiple sequence alignments were generated automatically within the ColabFold pipeline using MMseqs2. Structural models were evaluated using per-residue confidence scores (predicted Local Distance Difference Test, pLDDT). Residues with pLDDT scores above 90 were considered highly reliable, 70–90 moderately reliable, and below 70 indicative of lower

structural confidence. Structural models were visualized and rendered in PyMOL for downstream analysis.

2.6 Comparative Sequence and Phylogenetic Analysis

BMP protein sequences from maral and representative mammalian species (including cervids, bovids, primates, rodents, and carnivores) were retrieved from UniProt and NCBI databases. Multiple sequence alignments were generated using MAFFT v7.505 with default parameters. Alignments were manually inspected to ensure correct positioning of conserved cysteine residues forming the cystine-knot motif.

Pairwise sequence identity and similarity were calculated using EMBOSS Needle, implementing the Needleman–Wunsch global alignment algorithm with default substitution matrices and gap penalties.

Residue-level conservation scores were calculated in Jalview v2.11 using the built-in conservation scoring method based on physicochemical similarity and sequence identity across aligned positions. The conservation index for each residue position was exported and plotted as a histogram using standard data visualization tools in Python, allowing visualization of conservation distribution across the full-length protein.

2.7 Structural and Surface Property Analysis

Predicted BMP protein structures were analyzed in PyMOL v2.x to examine domain organization and surface properties. Surface hydrophobicity was calculated using the Eisenberg normalized consensus hydrophobicity scale, implemented through the `color_h` PyMOL script (*Color h - PyMOL Wiki*, n.d.). Amino acid residues were colored according to hydrophobicity values, with strongly hydrophobic residues (Ile, Leu, Val, Phe) displayed in red and hydrophilic or charged residues (Arg, Lys, Asp, Glu) shown in lighter shades.

3. RESULTS

3.1 Assembly contiguity and structural metrics (QUAST)

To evaluate the quality of final maral assemblies, ONT contigs were aligned to CerEla 1.1 genome assembly and assessed with QUAST v5.2.0. Assembly quality was assessed using a combination of contiguity, completeness, and compositional metrics. A summary of key QUAST statistics is provided in Table 1, while graphical representations of selected metrics are shown in Figures 1–3.

Table 1

Summary of QUAST assembly statistics for four draft maral genome assemblies (M1–M4), including contiguity, completeness, and compositional metrics

	M1	M2	M3	M4
Number of contigs	26228	16213	22644	31407
Contig N50 (bp)	298,723	1,873,284	370,881	199,215
Largest contig (bp)	2,607,845	16,947,203	3,126,970	1,604,360
Total length (bp)	3,410,255,336	3,186,643,300	3,376,042,163	3,724,359,404
GC content (%)	42.4	42.73	42.32	42.1
Genome fraction (%)	89.056	90.717	89.621	87.803

The four assemblies differed substantially in overall quality (Table 1). The number of contigs ranged from 16,213 in M2 to 31,407 in M4, indicating varying degrees of fragmentation. In terms of contig N50 values, M2 achieved the highest N50 (1,873,284 bp), compared to 298,723 bp in M1, 370,881 bp in M3, and 199,215 bp in M4. Similarly, the longest contig was observed in M2 (16,947,203 bp), substantially exceeding those of the other assemblies.

Total assembly lengths ranged from approximately 3.19 Gb to 3.72 Gb. Although M4 produced the largest total assembly size, this was not accompanied by improved contiguity or completeness, suggesting potential redundancy or assembly artifacts. In contrast, M2 achieved a more compact assembly size while maintaining superior structural continuity.

Assembly contiguity was further examined using Nx statistics, which describe the contig length at which a given percentage of the total assembly length is reached. The Nx plot (Figure 1) shows that M2 consistently exhibits higher Nx values, including the highest N50, indicating that 50% of the total assembly length is contained in relatively long contigs. This reflects reduced fragmentation and improved long-range continuity.

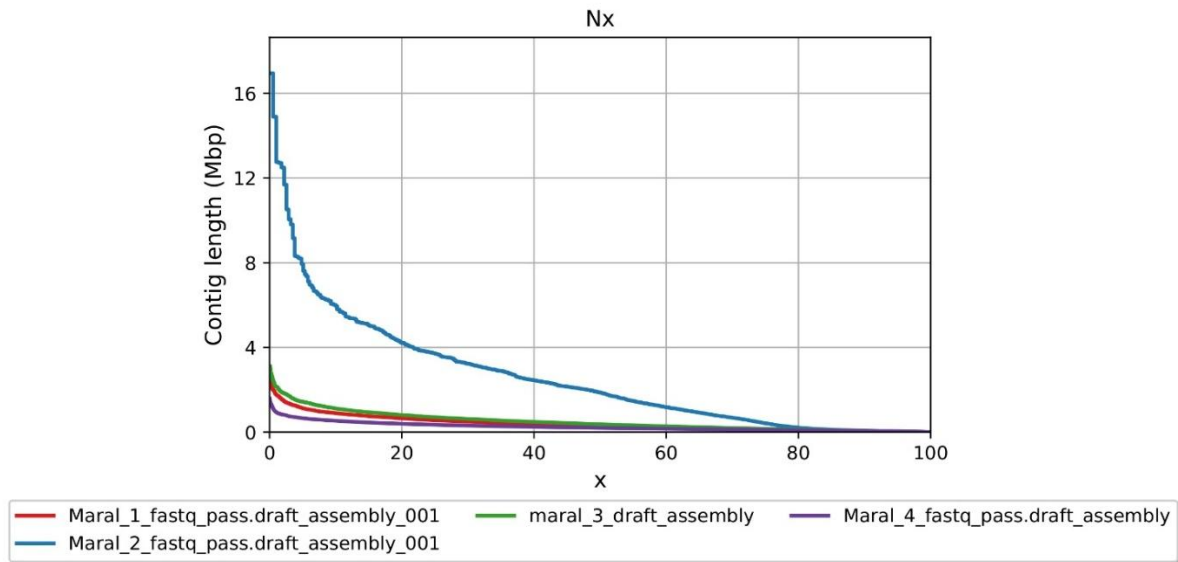


Figure 1. Nx plot comparing contiguity of the four draft maral genome assemblies

Alignment-based contiguity metrics (NAX and NGAX) showed similar trends (see Appendix A, Figure A1-A2), confirming that the improved contiguity of M2 was maintained when contigs were aligned to the reference genome.

Assembly completeness was assessed using the genome fraction metric, which represents the percentage of the reference genome covered by the assembly. M2 achieved the highest genome fraction (90.7%), indicating the most comprehensive coverage of the reference genome (Figure 2). M3 and M1 showed slightly lower values (89.6% and 89.1%, respectively), while M4 exhibited the lowest genome fraction (87.8%). These results indicate that M2 provides the most comprehensive representation of the maral genome relative to the reference.

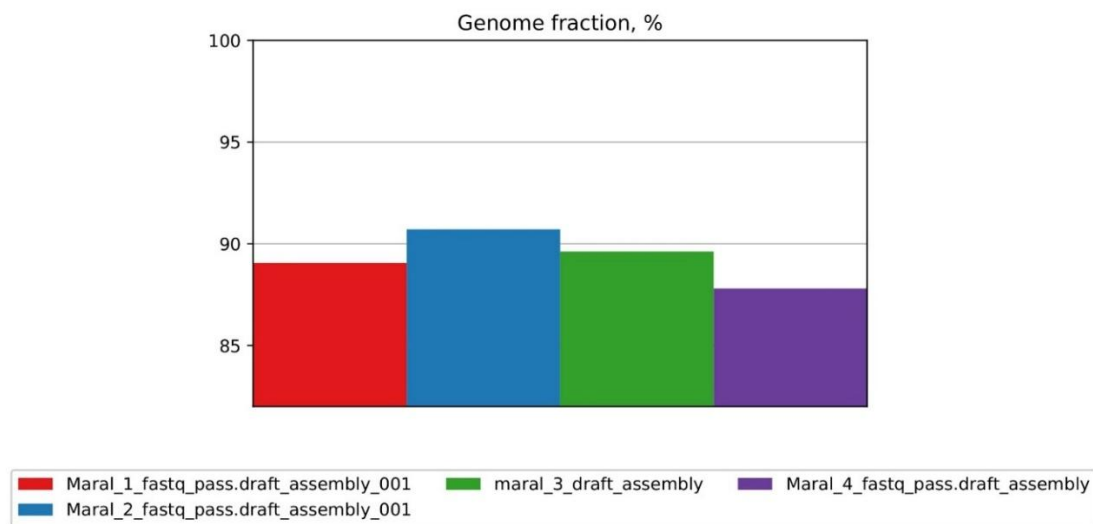


Figure 2. Genome fraction of the four assemblies relative to the reference genome CerEla 1.1

In addition, GC content distribution was highly consistent across all assemblies (42.1–42.7%) and showed no evidence of compositional bias or contamination (see Appendix A, Figure A3).

3.2 Gene-space completeness (BUSCO)

Although four draft assemblies (M1-M4) were evaluated using QUAST, only M1-M3 were selected for BUSCO analysis. Assembly M4 was excluded due to its higher fragmentation (largest number of contigs), lowest N50, and reduced genome fraction, indicating lower overall assembly quality.

The completeness of assemblies was assessed using BUSCO v.5.8.3 with the laurasiatherian_odb10 dataset, which contains 12,234 conserved orthologs, to quantify gene-space completeness of the three draft maral assemblies.

Table 2

BUSCO assessment of maral draft assemblies

	Maral_1	Maral_2	Maral_3
Complete (C)	91.9%	97.8%	93.9%
Complete single-copy (S)	78.7%	90.9%	80.8%
Complete duplicated (D)	13.2%	6.8%	1.2%
Fragmented (F)	4.5%	1.2%	3.4%
Missing (M)	3.6%	1.0%	2.7%
Internal stop codons	9.3%	9.5%	9.2%

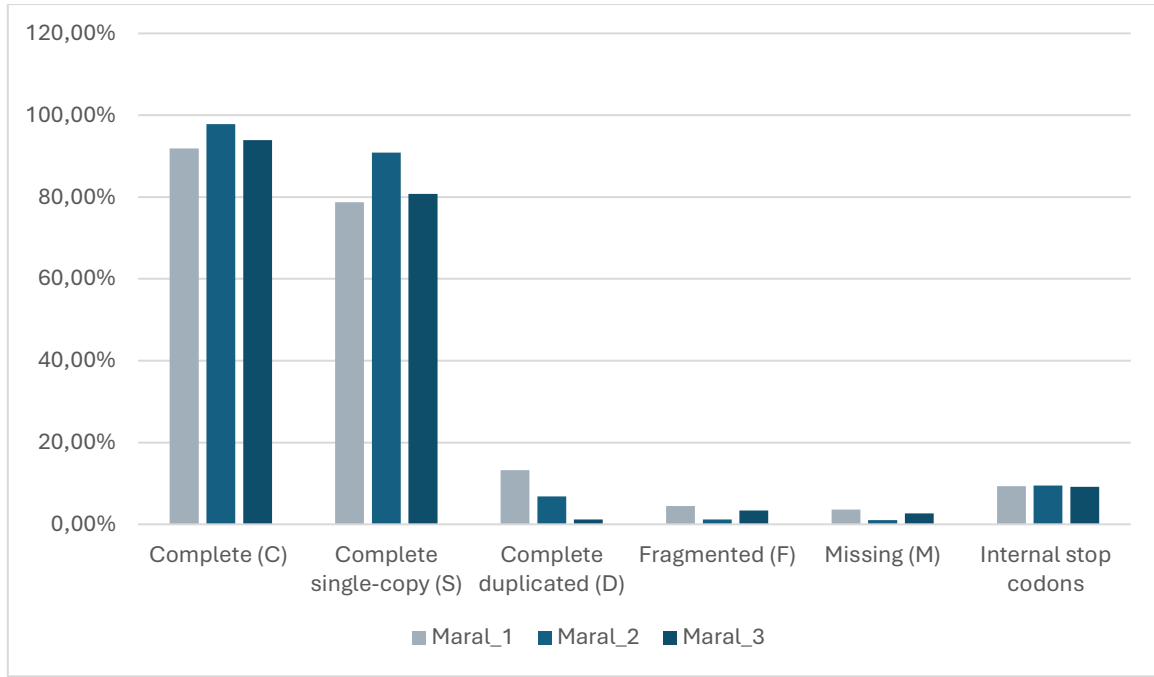


Figure 3. BUSCO completeness comparison across maral draft assemblies

Among the evaluated assemblies, Maral_2 showed the highest completeness, recovering 97,8% complete BUSCOs (90.9% single-copy, 6.8% duplicated), with only 1.2% fragmented and 1.0% missing BUSCOs. In comparison, maral_3 recovered 93.9% complete BUSCOs (80.8% single-copy, 13.0% duplicated) and Maral_1 recovered 91.9% complete BUSCOs (78.7% single-copy, 13.2% duplicated), both showing increased fragmentation and missingness relative to Maral_2.

Across all assemblies, ~9-10% of complete BUSCO predictions contained internal stop codons. In the context of a large mammalian genome, this level is expected and commonly observed, as it may arise from a combination of gene prediction and alignment artifacts, local indel- or frameshift-type assembly errors, incomplete exon reconstruction, or the presence of pseudogenes and disrupted gene copies. Importantly, the proportion of BUSCOs with internal stop codons was highly consistent across all three assemblies, indicating that this feature did not affect the relative ranking of assemblies (Rhie et al., 2021).

Based on its superior contiguity, highest genome fraction, and highest BUSCO completeness with minimal fragmentation and missingness, Maral_2 was selected as the reference assembly for all downstream annotation and comparative analyses.

3.3 Phylogenetic Placement of Maral within *Cervus*

To determine the evolutionary position of maral within the genus *Cervus*, a phylogenetic analysis was performed using mitochondrial DNA sequences. The resulting tree (Figure 4) resolved major *Cervid* lineages consistent with established mitochondrial phylogenies, including clear separation between *Cervus elaphus*, *Cervus canadensis*, and *Cervus nippon* groups.

The maral clusters within the *Cervus canadensis* clade and shows closest phylogenetic proximity to other *C. canadensis* subspecies. This clustering indicates a close evolutionary relationship within the wapiti lineage and supports its classification as a subspecies of *Cervus canadensis* rather than *Cervus elaphus*.

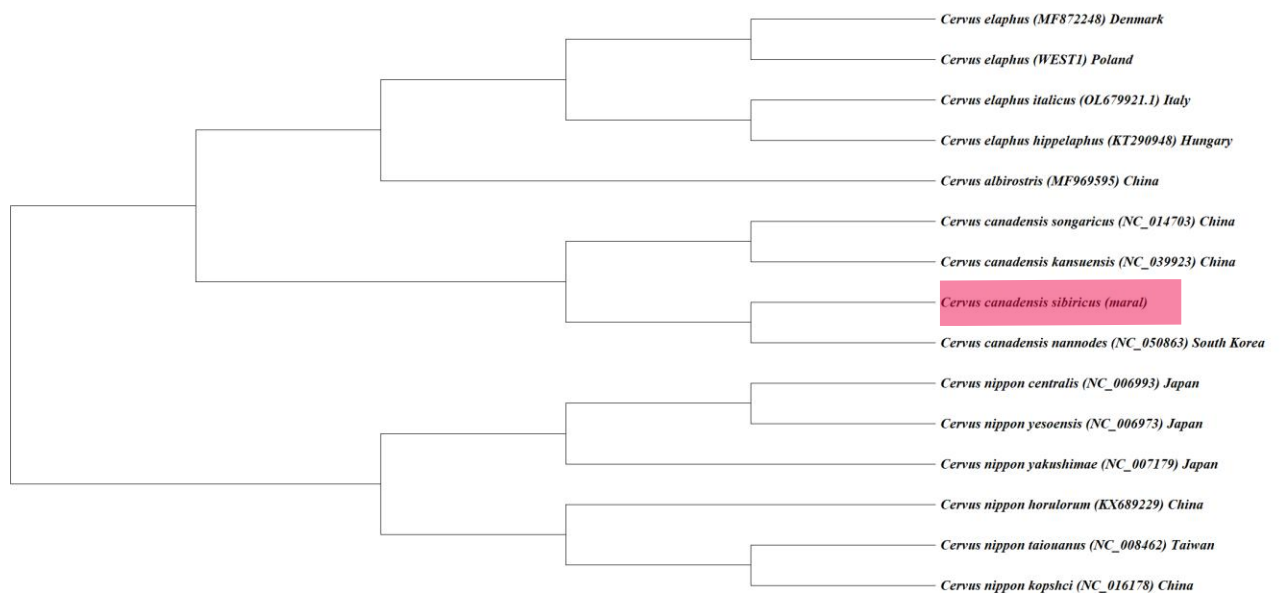


Figure 4. Mitochondrial DNA–based phylogenetic relationships of maral (*Cervus canadensis sibiricus*) within the genus *Cervus*

3.4 Functional annotation via InterPro

Using homology-based searches against the annotated Maral_2 genome, five BMP genes were identified: BMP2, BMP3, BMP4, BMP7, and BMP9. Protein lengths were 395 aa for BMP2, 475 aa for BMP3, 409 aa for BMP4, 430 aa for BMP7, and 429 aa for BMP9. These sequences were retained for downstream functional and evolutionary analyses.

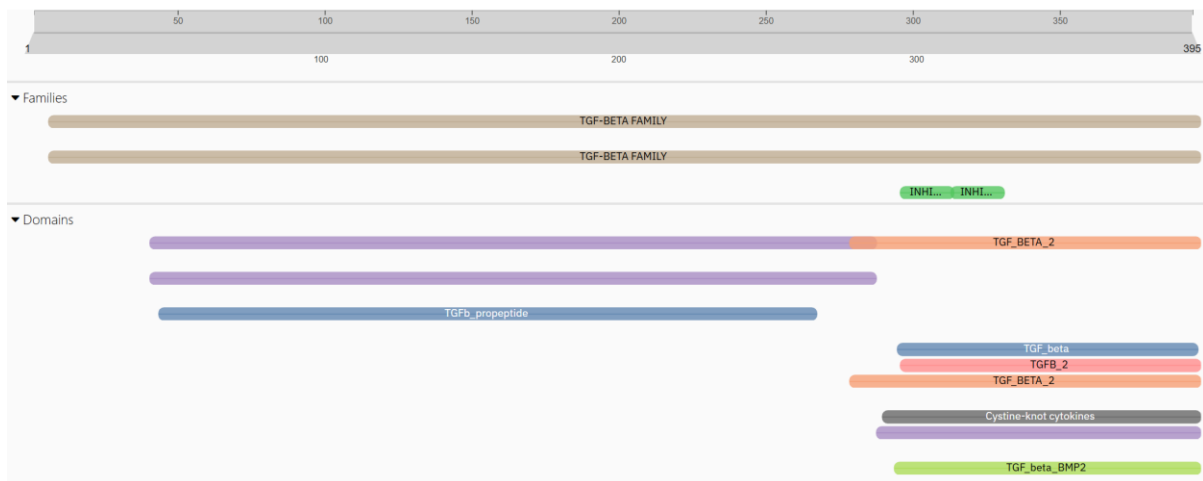


Figure 5. InterPro scan results of BMP2

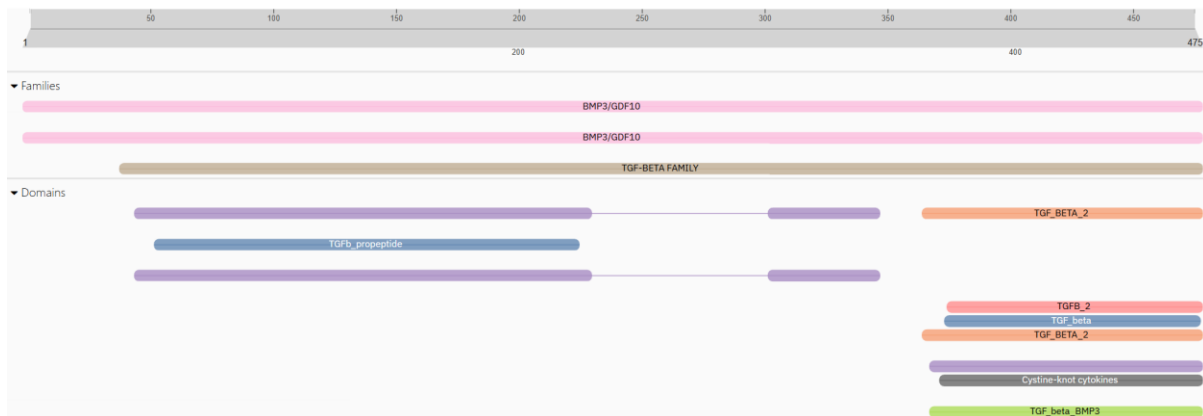


Figure 6. InterPro scan results of BMP3

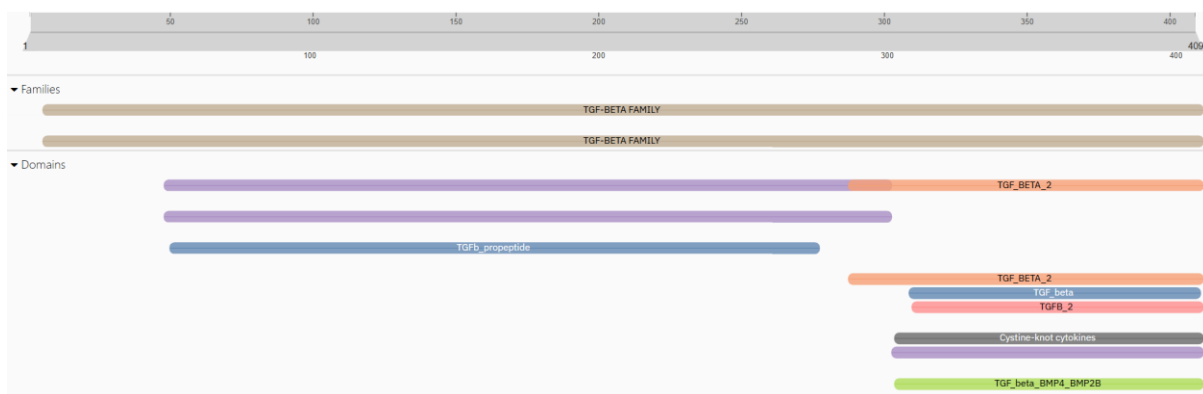


Figure 7. InterPro scan results of BMP4

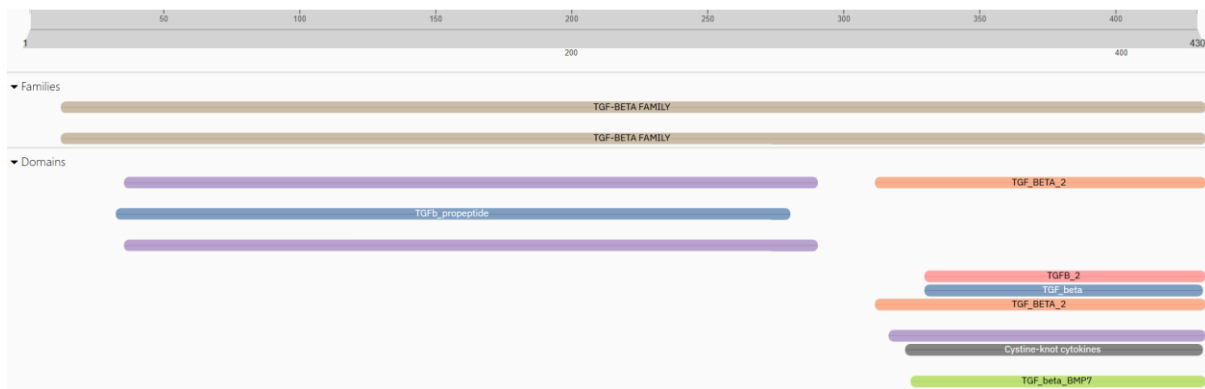


Figure 8. InterPro scan results of BMP7

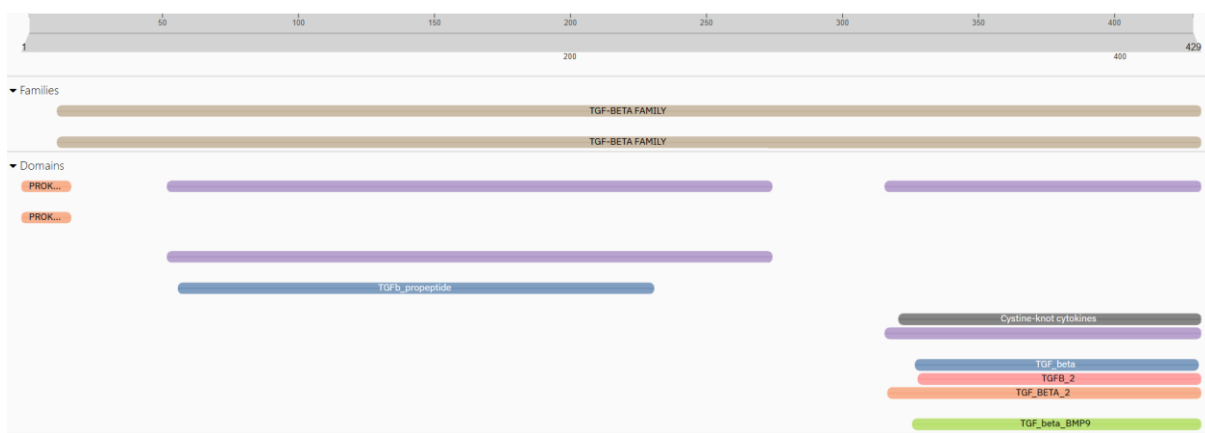


Figure 9. InterPro scan results of BMP9

InterProScan analysis indicated that each recovered maral BMP ligand displays the canonical BMP/TGF- β superfamily organization: an N-terminal signal peptide, an extended propeptide (pro-domain), and a C-terminal TGF- β superfamily mature ligand region containing the characteristic cystine-knot cytokine motif.

Across all five proteins (Figures 5-9), the TGF- β family signature was consistently identified at the protein family level, confirming their classification within the TGF- β superfamily. The cystine-knot cytokine motif, characteristic of active BMP ligands, was consistently detected within the C-terminal region of all proteins, overlapping with the mature ligand domain. This region also included BMP-specific annotations (TGF- β _BMP2, TGF- β _BMP3, TGF- β _BMP4/BMP2B, TGF- β _BMP7, and TGF- β _BMP9), further supporting correct protein identification. No lineage-specific domain gains, losses, or rearrangements were observed.

Notably, BMP3 showed a distinct InterPro annotation, distinguishing it from the other BMP ligands analyzed. This observation is consistent with the known functional role of BMP3

as a negative regulator of osteogenesis, in contrast to the osteoinductive activity of BMP2, BMP4, BMP7, and BMP9 (Daluiski et al., 2001).

3.5 Comparative Sequence Analysis of BMPs

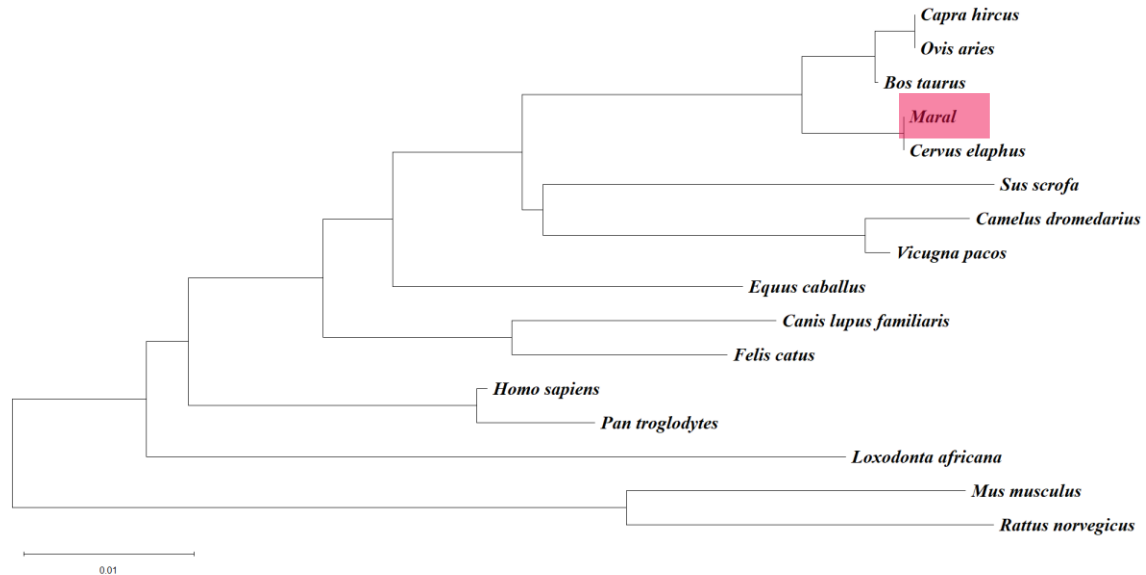


Figure 10. Phylogenetic relationships of BMP2 protein sequence across representative mammalian species

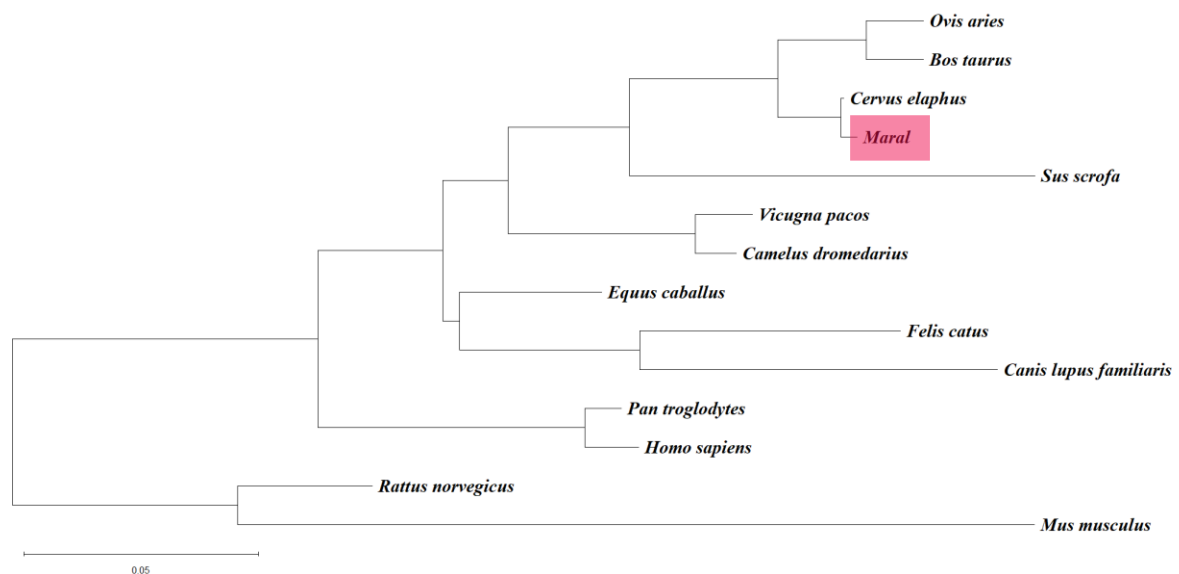


Figure 11. Phylogenetic relationships of BMP3 protein sequence across representative mammalian species

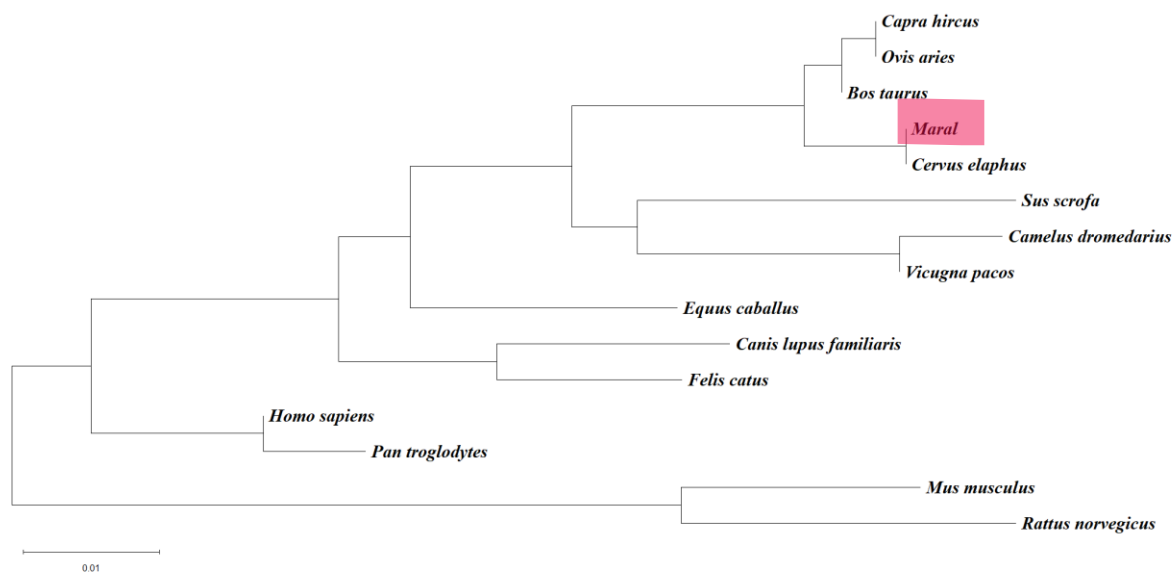


Figure 12. Phylogenetic relationships of BMP4 protein sequence across representative mammalian species

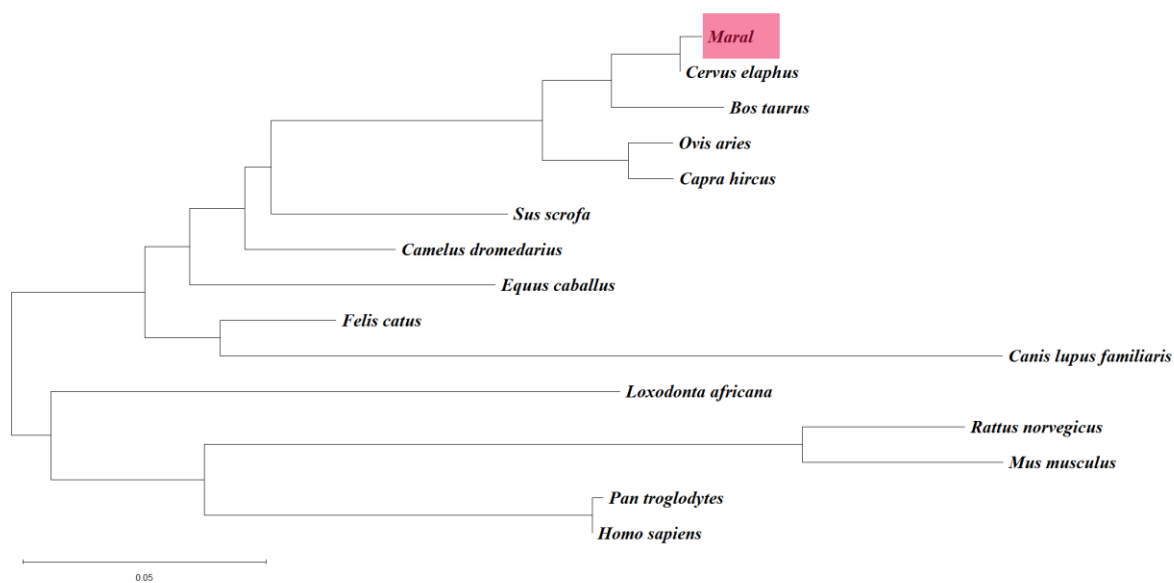


Figure 13. Phylogenetic relationships of BMP7 protein sequence across representative mammalian species

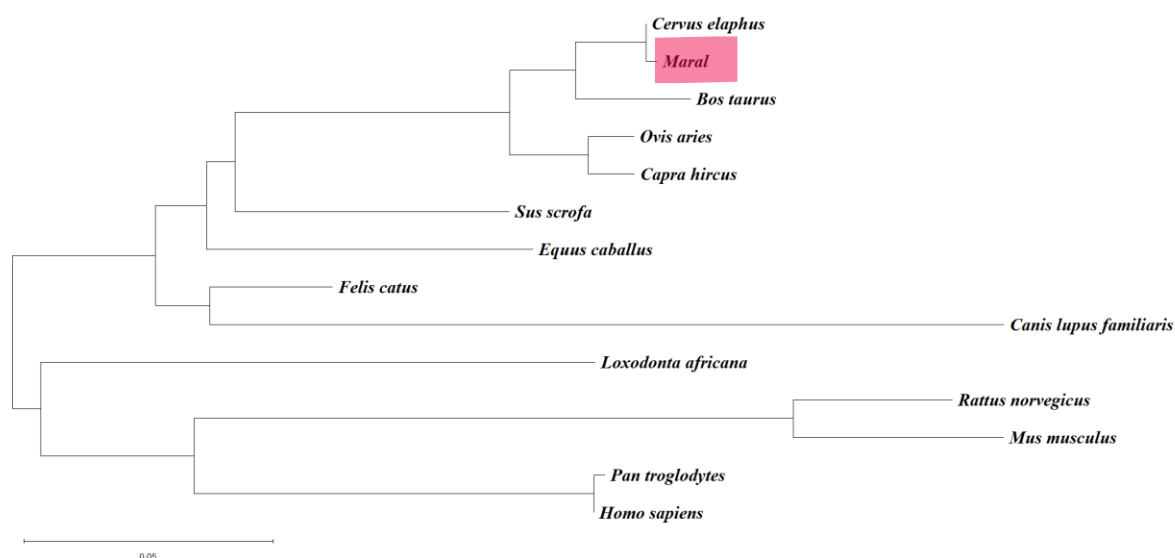


Figure 14. Phylogenetic relationships of BMP9 protein sequence across representative mammalian species

Phylogenetic analysis of BMP sequences was performed across representative mammalian species. The resulting trees (Figure 10-14) reflect the known evolutionary relationships among mammals. It should be noted that for some BMPs, certain species were excluded from the analysis due to the absence of their high-quality protein sequences. Specifically, *Capra hircus* was excluded from the BMP3, BMP7, and BMP9 trees; *Loxodonta africana* was excluded from the BMP3 and BMP4 trees; and *Camelus dromedarius* and *Vicugna pacos* were excluded from the BMP9 tree.

Across all five trees, primates (*Homo sapiens* and *Pan troglodytes*) consistently clustered together, and artiodactyls (*Bos taurus*, *Ovis aries*, *Capra hircus*, *Cervus elaphus*, *Maral*, and *Sus scrofa*) formed a coherent clade, consistent with established mammalian taxonomy. Rodents (*Mus musculus* and *Rattus norvegicus*) consistently formed a distinct outgroup, reflecting their characteristically faster evolutionary rates relative to other mammals. *Loxodonta africana* (African elephant) branched early across most trees, in accordance with its classification as an Afrotherian mammal distantly related to the other species included in the dataset.

BMP2 and BMP4 exhibited the highest degree of sequence conservation across species, as indicated by a scale bar of 0.01 substitutions per site, meaning that on average only one amino acid difference exists per 100 positions across the included species. This high conservation is consistent with the critical and non-redundant roles of BMP2 and BMP4 in

embryonic development, where strong purifying selection is expected to limit sequence divergence.

In contrast, BMP3, BMP7, and BMP9 displayed comparatively greater sequence divergence, with a scale bar of 0.05 substitutions per site, indicating five times more amino acid variation across species than observed for BMP2 and BMP4. This suggests that these BMPs tolerate greater sequence variability while retaining their biological function, possibly reflecting lineage-specific functional adaptations.

Table 3

Pairwise sequence alignment of maral BMPs with human orthologs (EMBOSS Needle)

<i>Protein</i>	<i>Length (aa)</i>	<i>Identity (%)</i>	<i>Similarity (%)</i>	<i>Gaps (%)</i>	<i>Score</i>
<i>BMP2</i>	396	93.9	96.5	0.3	1966.0
<i>BMP3</i>	475	85.9	89.9	0.6	2107.5
<i>BMP4</i>	409	98.0	98.5	0.2	2136.0
<i>BMP7</i>	431	97.7	98.1	0.2	2228.0
<i>BMP9</i>	433	80.8	89.1	1.8	1834

To assess the degree of sequence conservation between human and maral BMP proteins, pairwise sequence alignment was performed using the EMBOSS Needle tool. The results are summarized in Table 3. Four parameters were recorded for each alignment: identity, similarity, gaps, and alignment score. Identity refers to the percentage of positions in the alignment where both sequences share the exact same amino acid. Similarity includes both identical positions and positions where the amino acids are biochemically similar. Gaps represent the percentage of positions where an insertion or deletion was introduced to optimize the alignment. The alignment score reflects the overall quality of the alignment, with higher score indicating greater similarity.

BMP4 showed the highest sequence identity between human and maral at 98% followed closely by BMP7 at 97.7%, indicating near-identical protein sequences for these two BMPs. BMP2 also demonstrated high conservation at 93.9% identity. In contrast, BMP3 and BMP9 showed comparatively lower identity values of 85.9% and 80.8% respectively, suggesting these proteins have diverged more.

Gap percentages were low across all BMPs, ranging from 0.2% to 1.8%, indicating that the protein sequences are well-aligned with minimal insertions or deletions. The highest percentage was observed in BMP9 (1.8%), consistent with its lower identity score.

3.6 Structural Annotation

The three-dimensional structural models of selected maral BMPs were generated using ColabFold. Model confidence was assessed using pLDDT, where per-residue values > 90 indicate very high confidence, 70-90 indicate confident backbone prediction, 50-70 indicate low confidence, and <50 indicates very low confidence often associated with flexible regions.

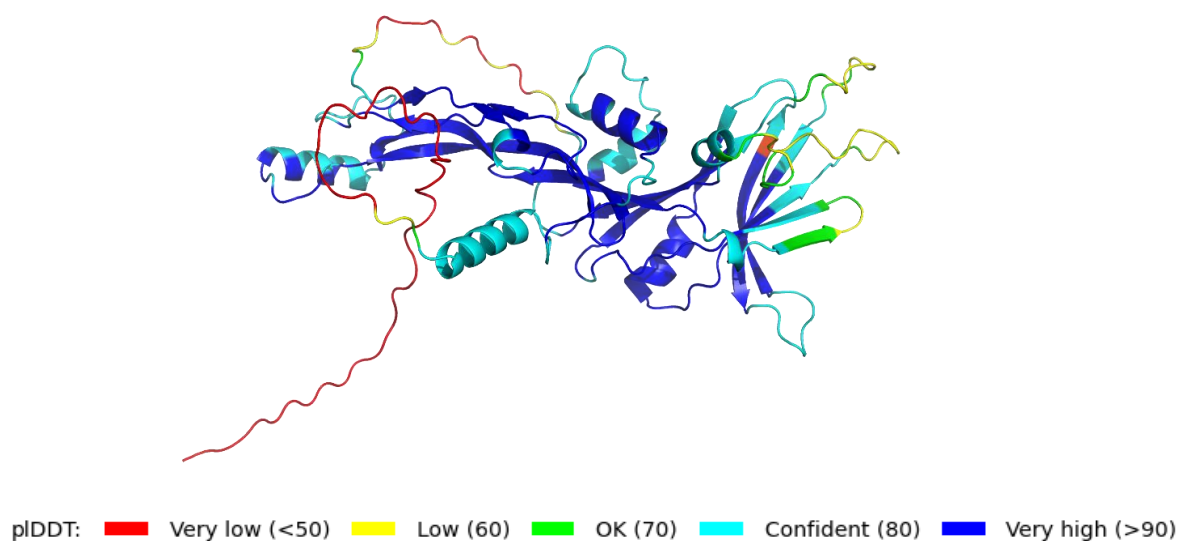


Figure 15. Predicted three-dimensional structure of maral BMP2 generated using ColabFold

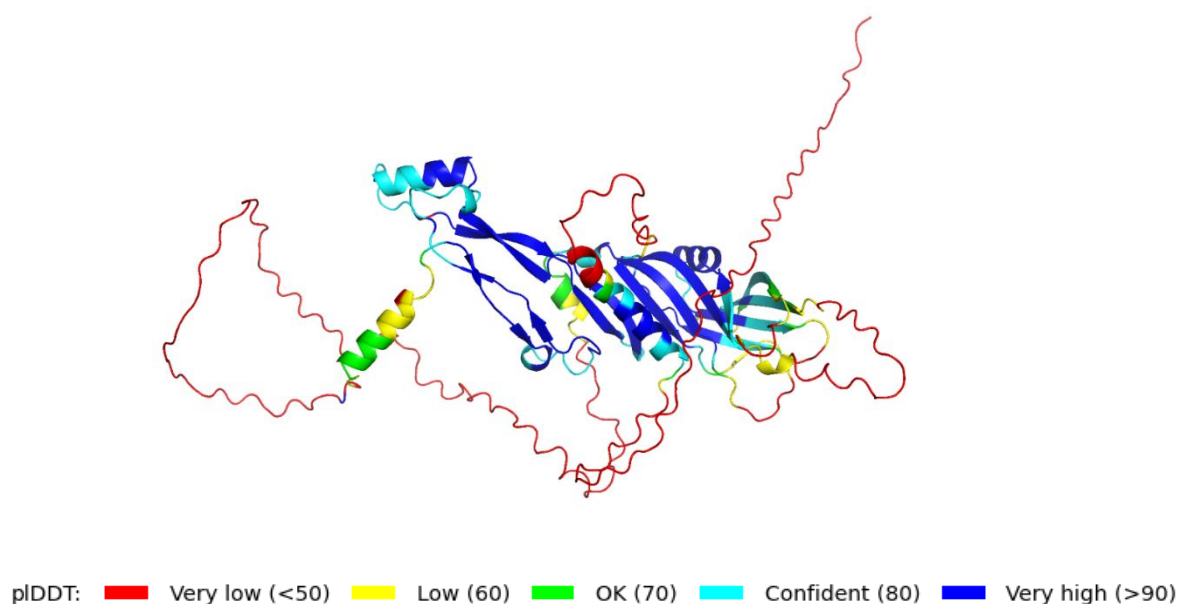


Figure 16. Predicted three-dimensional structure of maral BMP3 generated using ColabFold

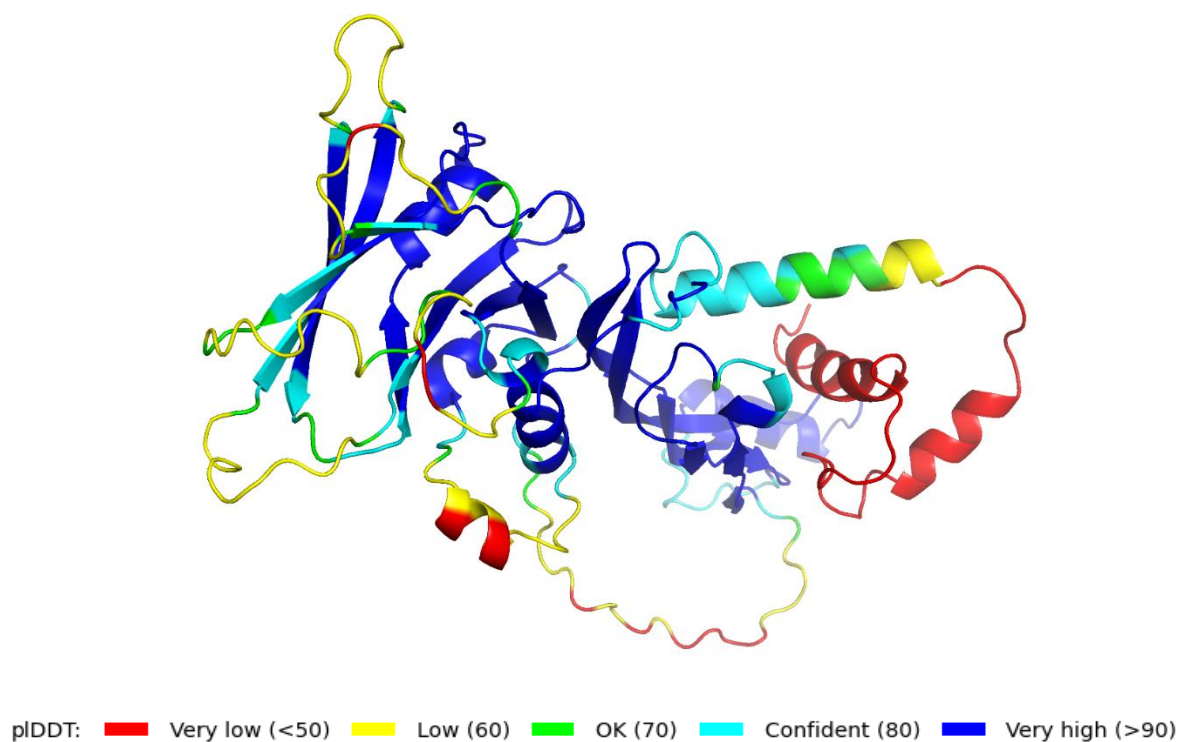


Figure 17. Predicted three-dimensional structure of maral BMP4 generated using ColabFold

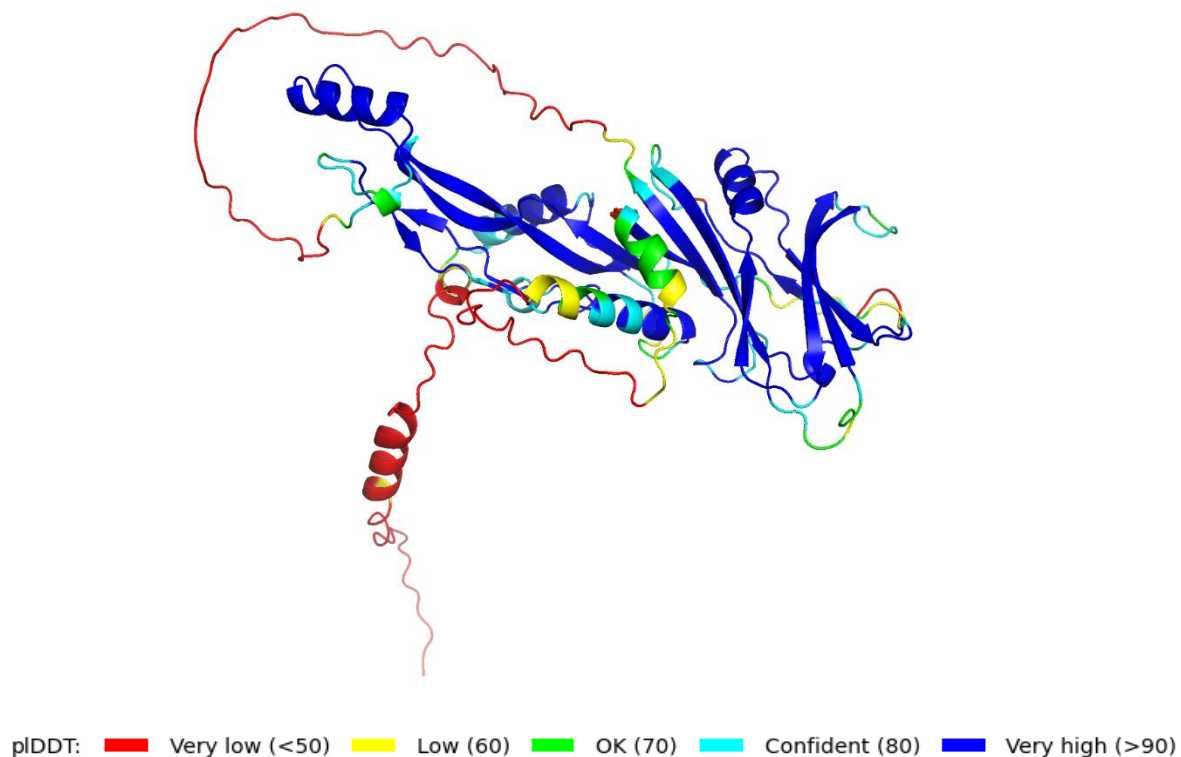


Figure 18. Predicted three-dimensional structure of maral BMP7 generated using ColabFold

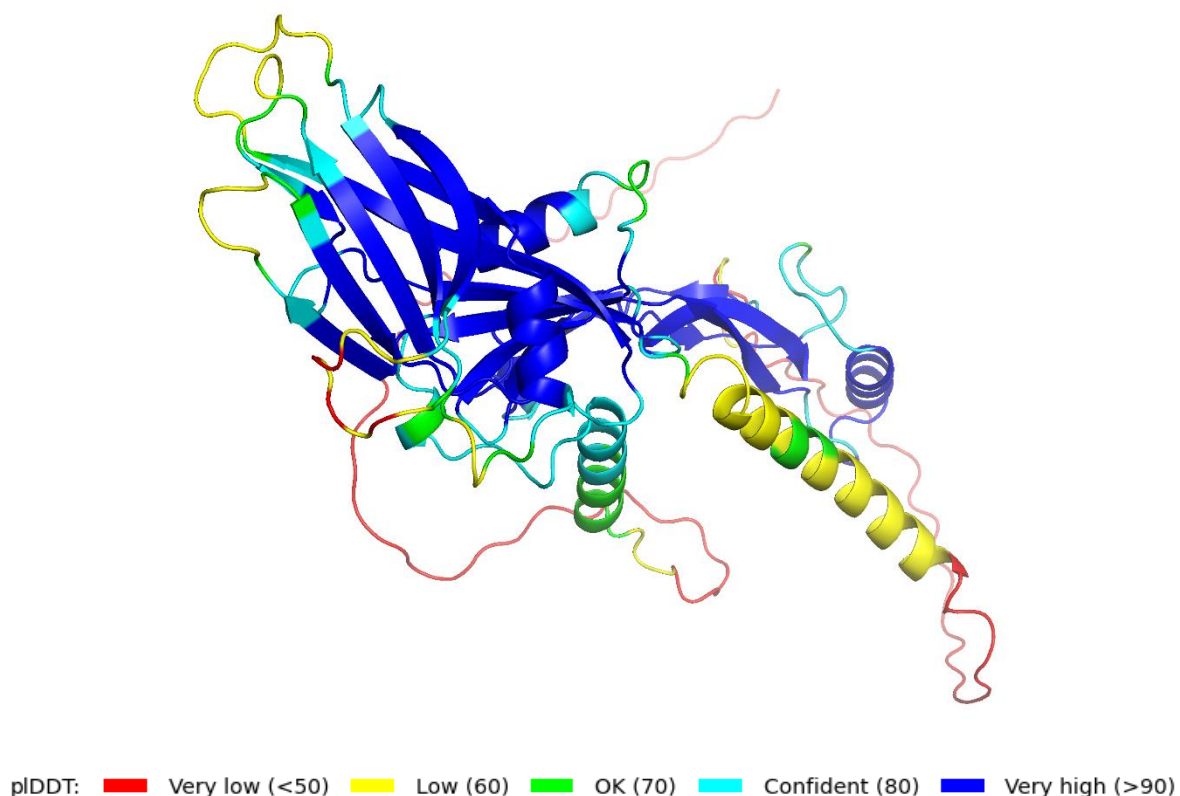


Figure 19. Predicted three-dimensional structure of maral BMP9 generated using ColabFold.

All predicted BMP structures exhibited moderate to high confidence across the majority of their sequences, indicating reliable structural predictions. High-confidence regions were distributed throughout the proteins, with well-defined β -sheets and α -helices.

In BMP2 (Figure 15), the central structured region showed high confidence, particularly across the β -sheet rich core, while lower-confidence regions were mainly in extended terminal and loop segments.

BMP3 (Figure 16) showed the most extensive low-confidence regions among the proteins analyzed. Large portions of the terminal and peripheral regions were predicted with lower confidence, producing a more fragmented appearance compared with the other BMPs. This suggests that BMP3 contains a greater proportion of flexible or disordered segments in the predicted model.

BMP4 (Figure 17) displayed a well-resolved structure, with clearly defined β -sheets and α -helices.

Similarly, BMP7 and BMP9 (Figure 18, 19) showed a conserved high-confidence structural core, although several terminal and loop regions exhibited lower confidence, indicating localized structural flexibility.

Across all five proteins, the low-confidence segments correspond to loop regions or extended segments lacking stable-secondary structure. These regions are likely to represent flexible or less-ordered parts of the protein, which are inherently more difficult to predict with high confidence.

3.7 Conservation Analysis

Residue-wise conservation analysis across aligned mammalian BMP sequences, calculated using Jalview conservation scores (Figures 20-24), revealed a consistent regional variation across all BMP proteins.

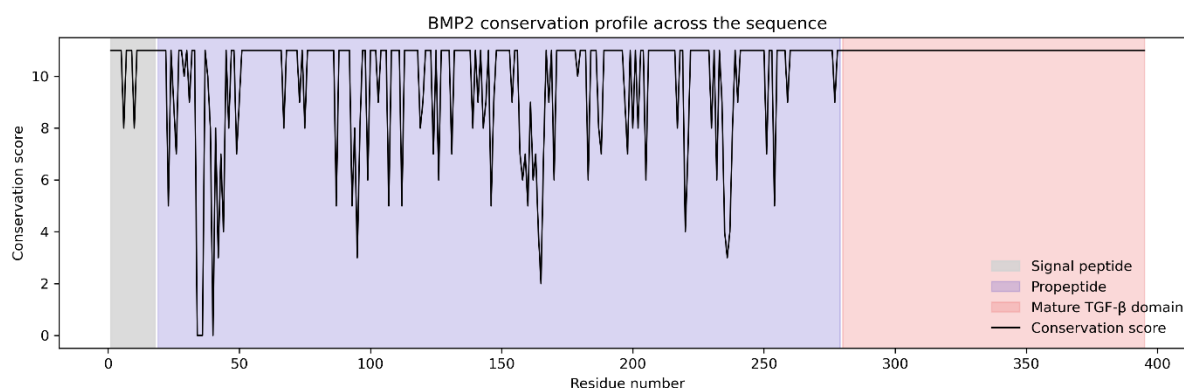


Figure 20. BMP2 Residue-wise Jalview conservation scores for BMP across mammalian species.

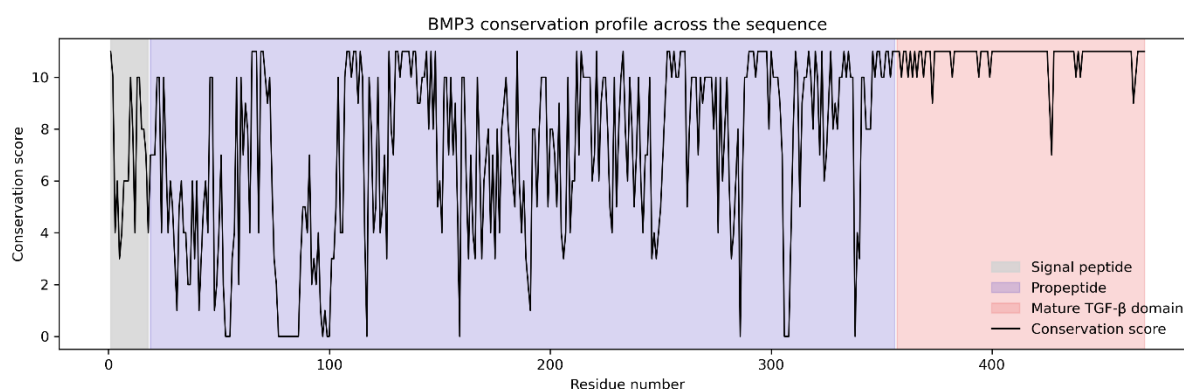


Figure 21. BMP3 Residue-wise Jalview conservation scores for BMP across mammalian species.

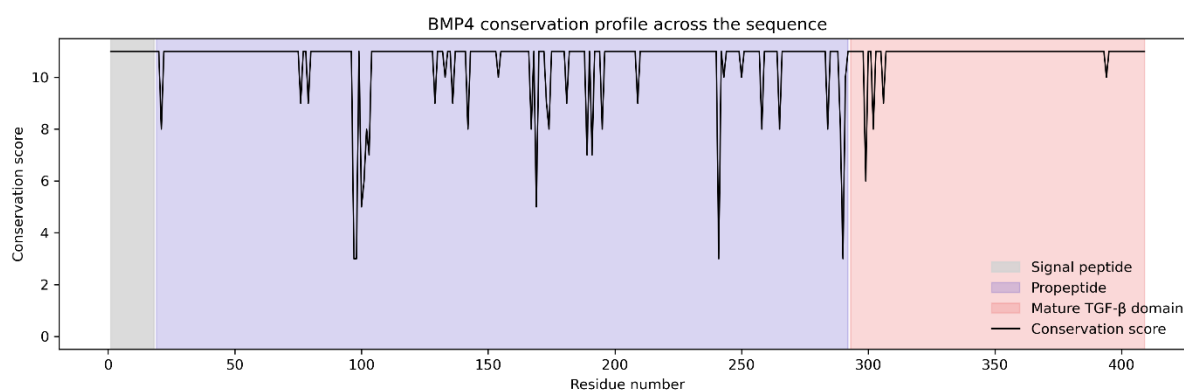


Figure 22. BMP4 Residue-wise Jalview conservation scores for BMP across mammalian species.

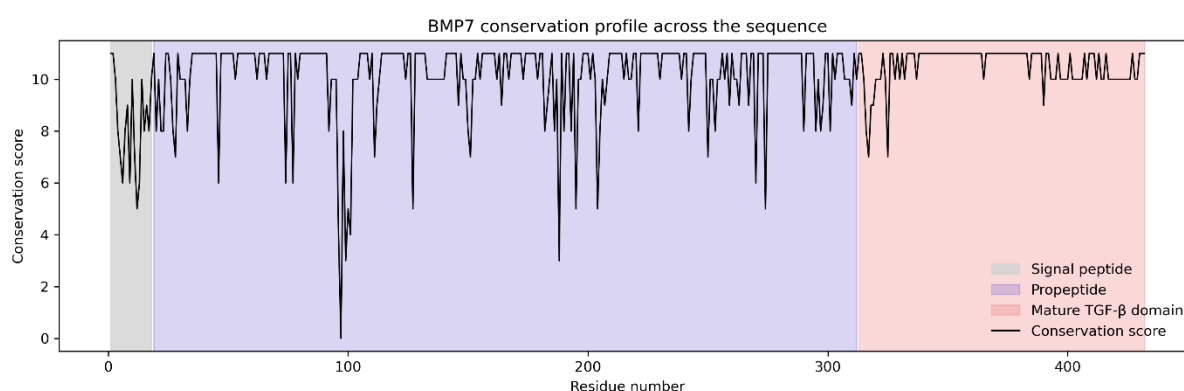


Figure 23. BMP7 Residue-wise Jalview conservation scores for BMP across mammalian species.

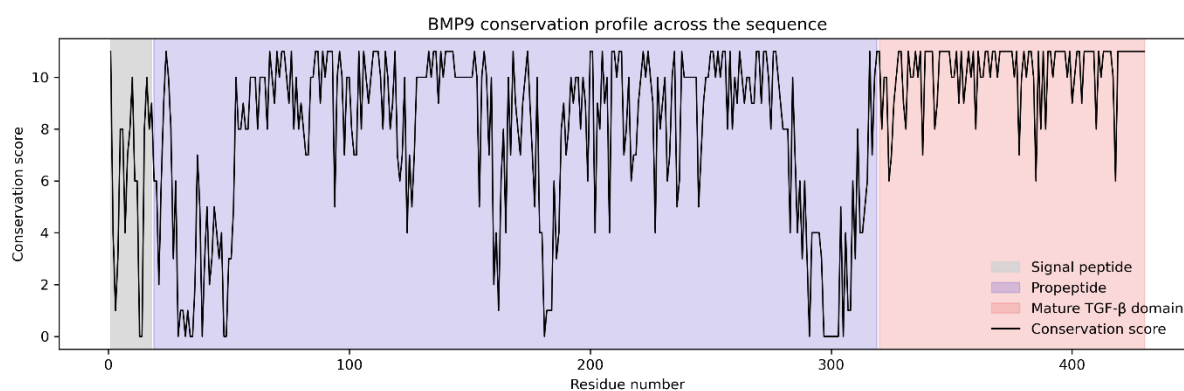


Figure 24. BMP9 Residue-wise Jalview conservation scores for BMP across mammalian species.

In all analyzed sequences, the C-terminal domain exhibited uniformly high conservation scores, indicating strong sequence preservation across species.

This highly conserved region corresponds to the cystine-knot cytokine domain. The cystine-knot motif is formed by a conserved arrangement of cysteine residues that establish multiple disulfide bonds, creating a stable three-dimensional framework essential for protein folding, dimerization, and receptor binding (Gipson et al., 2020).

In contrast, the N-terminal regions, corresponding to the signal peptide and pro-domain, showed greater variability, with multiple local minima in conservation scores. The regions are less structurally constrained and are known to participate in protein processing, folding assistance, and regulatory interactions, allowing for greater sequence divergence across species.

3.8 Surface Hydrophobicity analysis

Hydrophobicity mapping provides a qualitative structural lens on how predicted BMP ligand surfaces may present interaction-ready patches to receptors, antagonists, or extracellular matrix components. In this study, surface hydrophobicity was computed from predicted BMP structures by coloring residues in PyMOL.

Across maral BMPs (Figures 25-29), surfaces displayed heterogeneous hydrophobic/hydrophilic patch distributions across the predicted folds. This provides a qualitative view of potential surface property differences among BMP ligands for subsequent hypothesis generation regarding interactions with receptors, antagonists, or extracellular matrix components.

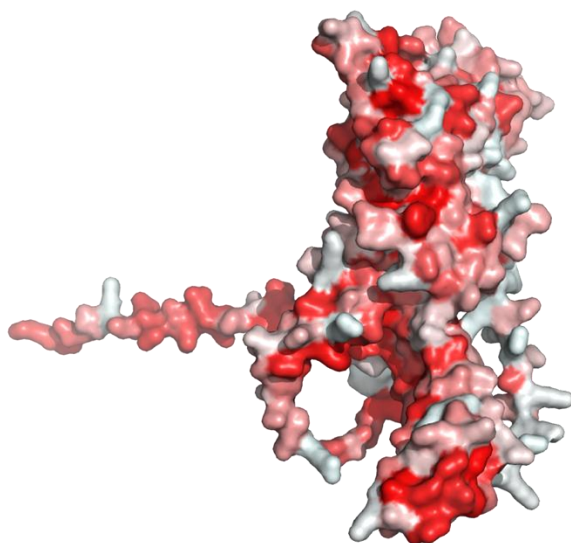


Figure 25. Hydrophobicity mapping of the predicted maral BMP2 structure

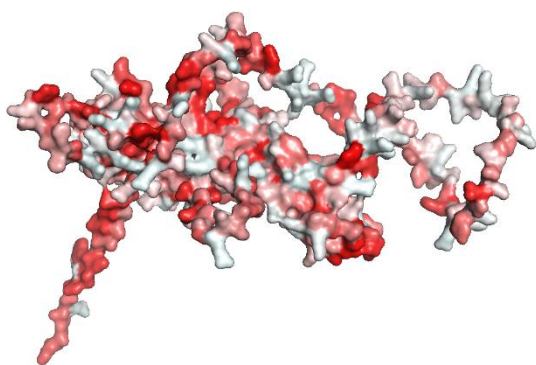


Figure 26. Hydrophobicity mapping of the predicted maral BMP3 structure

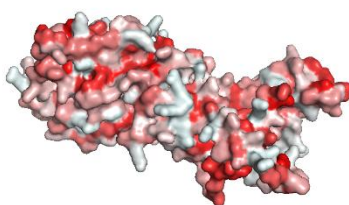


Figure 27. Hydrophobicity mapping of the predicted maral BMP4 structure

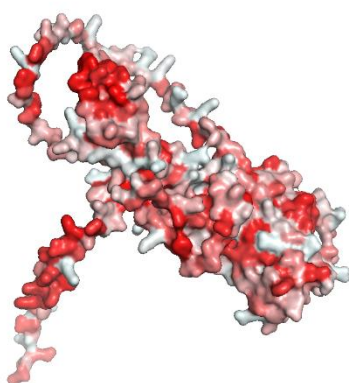


Figure 28. Hydrophobicity mapping of the predicted maral BMP7 structure

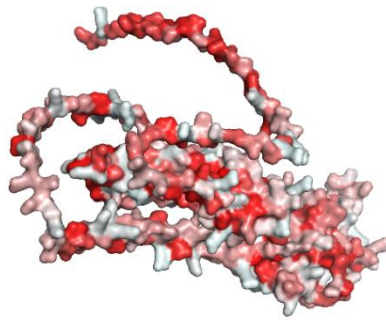


Figure 29. Hydrophobicity mapping of the predicted maral BMP9 structure

BMP2 displayed the most pronounced overall hydrophobicity among the five proteins. The compact globular core of the mature domain was dominated by deep red coloring, indicating dense hydrophobic character across the receptor-binding face.

In BMP3 profile the overall hydrophobicity was lower, with a greater proportion of white and light pink surface area, indicating a more hydrophilic surface. This is consistent with BMP3's distinct functional role as a negative regulator of osteogenesis. Its divergent surface properties may underlie its inability to productively engage the canonical osteogenic BMP receptor complexes.

BMP4 presented a well-defined bilobal structure with hydrophobic patches concentrated on both lobes of the mature domain. BMP 7 showed strong hydrophobicity concentrated in the compact structured core of the mature domain, similar in pattern to BMP2.

BMP9 displayed a hydrophobic core within the compact region of the mature domain, with the extended loop regions showing more mixed hydrophobic/hydrophilic character.

4. DISCUSSION

Among the four draft assemblies, Maral_2 demonstrated superior contiguity, genome fraction, and gene-space completeness. The high N50 value and BUSCO completeness indicate that the assembly captures the majority of conserved mammalian genes, providing a reliable foundation for downstream gene annotation.

The presence of approximately 9–10% internal stop codons across assemblies is consistent with typical long-read sequencing artifacts, particularly indel-related errors inherent to ONT sequencing. Importantly, this pattern was consistent across assemblies and did not compromise relative assembly ranking. The selected reference assembly therefore represents a robust genomic resource for further cervid research.

The maral sequence (*Cervus canadensis sibiricus*) clustered within the *Cervus canadensis* clade, forming a well-supported group with other Asian wapiti subspecies, while being distinct from Western deer lineage (*Cervus elaphus*). These findings are consistent with (Mackiewicz et al., 2022), who also reported a clear separation between Asian wapiti and European red deer based on Mitochondrial genome data.

This thesis' primary aim was to connect BMP ligand sequences to a unique regenerative capability of deer antlers. The conservation analyses support two conclusions.

First, the high conservation of the mature TGF- β /BMP ligand domain is biologically expected. C-terminal mature peptide binds to type I and type II serine/threonine kinase receptors and activates canonical receptor-regulated SMADs (Wang et al., 2014). Under such constraint, extensive amino-acid divergence in the folded mature ligand domain would disrupt the receptor engagement or downstream pathway activation, which in turn would compromise BMPs' essential roles in skeletal development and tissue homeostasis.

Second, the greater variability in pro-domain should be interpreted as a plausible location where functional differences could arise without disrupting the mature ligand fold. BMP pro-domains regulate folding, dimerization, intracellular trafficking, and proteolytic processing that releases the active growth factor (Wang et al., 2014). This supports an idea that regeneration-relevant differences may be regulatory and processing-based rather than large changes in mature-ligand residues.

The high sequence identity between maral and human BMPs, combined with the conserved phylogenetic placement of maral within mammalian BMP trees, suggests that any

functional specialization underlying antler regeneration is unlikely to arise from large-scale changes in primary sequence. Instead, as discussed above, regulatory and processing-level differences in the pro-domain remain the more plausible candidate mechanism.

Furthermore, the phylogenetic trees revealed a notable five-fold difference in evolutionary rate between BMP groups. BMP2 and BMP4 showed markedly lower sequence divergence across mammalian species (scale bar: 0.01 substitutions per site), whereas BMP3, BMP7, and BMP9 exhibited comparatively greater divergence (scale bar: 0.05 substitutions per site). This difference likely reflects stronger purifying selection acting on BMP2 and BMP4, consistent with their well-established, non-redundant roles in osteogenesis and skeletal patterning, where amino acid changes would be poorly tolerated. In contrast, the relatively higher divergence of BMP3, BMP7, and BMP9 may indicate greater functional flexibility or redundancy within these subfamilies. In the context of antler regeneration, this distinction is biologically meaningful: if BMP2 and BMP4 are evolutionarily constrained across all mammals, their contribution to the deer-specific regenerative phenotype is more likely mediated through differences in expression level, timing, or upstream regulation rather than through changes in protein sequence itself.

The major limitation of this study is that it is primarily sequence/structure comparative and does not yet incorporate antler tissue expression. Without expression data, it is not possible to test whether maral BMPs are uniquely deployed (timing, location, isoforms, or dosage) in antler regeneration compared with other cervids or mammals.

To address this limitation the next step would be to integrate RNA-seq (bulk) or single-cell data from antlerogenic periosteum/pedicle periosteum/growth center layers and test whether BMP ligands, receptors, and antagonists are differentially expressed during growth vs casting.

5. CONCLUSION

This study provides a comprehensive genomic and comparative characterization of BMPs in the maral, integrating whole-genome assembly, gene identification, evolutionary analysis, and structural bioinformatics. Maral_2 provided the most reliable foundation for downstream annotation and BMP retrieval. Phylogenetically, the maral mitochondrial sequence clustered within the wapiti lineage and was distinct from red deer species, supporting correct taxonomic placement for comparative analyses.

Across the characterized BMPs, the C-terminal mature domain was highly conserved, whereas the pro-domain was more variable, indicating a regulatory and processing divergence without disrupting the conserved mature fold. Collectively, these results support the idea that maral's exceptional antler regeneration is more likely driven by regulation, precursor processing, and pathway-level modulations than by novel mature-ligand sequences.

6. REFERENCE LIST

- Ba, H., Chen, M., & Li, C. (2022). Cross-Species Analysis Reveals Co-Expressed Genes Regulating Antler Development in Cervidae. *Frontiers in Genetics*, 13, 878078.
<https://doi.org/10.3389/fgene.2022.878078>
- Carreira, A. C. O., Zambuzzi, W. F., Rossi, M. C., Filho, R. A., Sogayar, M. C., & Granjeiro, J. M. (2015). Bone Morphogenetic Proteins. In *Vitamins & Hormones* (Vol. 99, pp. 293–322). Elsevier. <https://doi.org/10.1016/bs.vh.2015.06.002>
- Color h—PyMOL Wiki. (n.d.). Retrieved 29 March 2026, from https://pymolwiki.org/Color_h
- Daluiski, A., Engstrand, T., Bahamonde, M. E., Gamer, L. W., Agius, E., Stevenson, S. L., Cox, K., Rosen, V., & Lyons, K. M. (2001). Bone morphogenetic protein-3 is a negative regulator of bone density. *Nature Genetics*, 27(1), 84–88. <https://doi.org/10.1038/83810>
- Feleke, M., Bennett, S., Chen, J., Hu, X., Williams, D., & Xu, J. (2021a). New physiological insights into the phenomena of deer antler: A unique model for skeletal tissue regeneration. *Journal of Orthopaedic Translation*, 27, 57–66. <https://doi.org/10.1016/j.jot.2020.10.012>
- Gipson, G. R., Goebel, E. J., Hart, K. N., Kappes, E. C., Kattamuri, C., McCoy, J. C., & Thompson, T. B. (2020). Structural perspective of BMP ligands and signaling. *Bone*, 140, 115549. <https://doi.org/10.1016/j.bone.2020.115549>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Kierdorf, U., & Kierdorf, H. (2011). Deer Antlers – A Model of Mammalian Appendage Regeneration: An Extensive Review. *Gerontology*, 57(1), 53–65. <https://doi.org/10.1159/000300565>
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540–546. <https://doi.org/10.1038/s41587-019-0072-8>

- Lowery, J. W., & Rosen, V. (2018). The BMP Pathway and Its Inhibitors in the Skeleton. *Physiological Reviews*, 98(4), 2431–2452. <https://doi.org/10.1152/physrev.00028.2017>
- Mackiewicz, P., Matosiuk, M., Świsłocka, M., Zachos, F. E., Hajji, G. M., Saveljev, A. P., Seryodkin, I. V., Farahvash, T., Rezaei, H. R., Torshizi, R. V., Mattioli, S., & Ratkiewicz, M. (2022). Phylogeny and evolution of the genus *Cervus* (Cervidae, Mammalia) as revealed by complete mitochondrial genomes. *Scientific Reports*, 12(1), 16381. <https://doi.org/10.1038/s41598-022-20763-x>
- McKay, W. F., Peckham, S. M., & Badura, J. M. (2007). A comprehensive clinical review of recombinant human bone morphogenetic protein-2 (INFUSE® Bone Graft). *International Orthopaedics*, 31(6), 729–734. <https://doi.org/10.1007/s00264-007-0418-6>
- Nasoori, A. (2020). Formation, structure, and function of extra-skeletal bones in mammals. *Biological Reviews*, 95(4), 986–1019. <https://doi.org/10.1111/brv.12597>
- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., ... Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), 737–746. <https://doi.org/10.1038/s41586-021-03451-0>
- Salazar, V. S., Gamer, L. W., & Rosen, V. (2016). BMP signalling in skeletal development, disease and repair. *Nature Reviews Endocrinology*, 12(4), 203–221. <https://doi.org/10.1038/nrendo.2016.12>
- Sampath, T. K., & Reddi, A. H. (2020). Discovery of bone morphogenetic proteins – A historical perspective. *Bone*, 140, 115548. <https://doi.org/10.1016/j.bone.2020.115548>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Sui, Z., Zhang, L., Huo, Y., & Zhang, Y. (2014). Bioactive components of velvet antlers and their pharmacological properties. *Journal of Pharmaceutical and Biomedical Analysis*, 87, 229–240. <https://doi.org/10.1016/j.jpba.2013.07.044>

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis

Version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027.

<https://doi.org/10.1093/molbev/msab120>

Wang, R. N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z.,

Denduluri, S., Idowu, O., Li, M., Shen, C., Hu, A., Haydon, R. C., Kang, R., Mok, J., Lee, M.

J., Luu, H. L., & Shi, L. L. (2014). Bone Morphogenetic Protein (BMP) signaling in

development and human diseases. *Genes & Diseases*, 1(1), 87–105.

<https://doi.org/10.1016/j.gendis.2014.07.005>

7. APPENDICES

Appendix A

Figure A1

NGx plot comparing contiguity of maral genome assemblies relative to estimated genome size

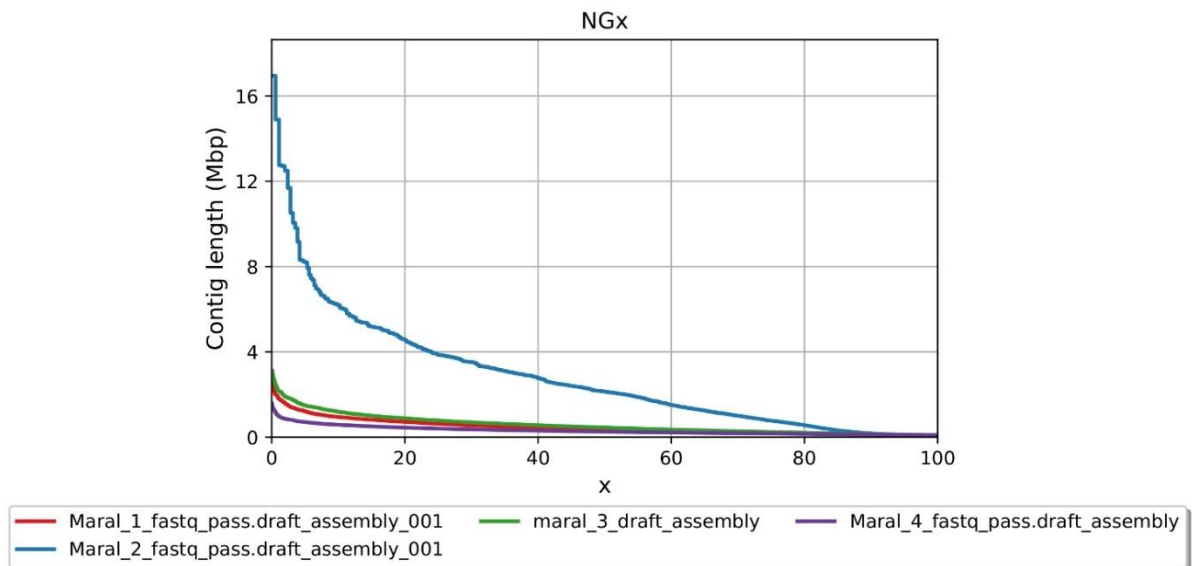


Figure A2

NGAx plot showing alignment-based contiguity of maral genome assemblies relative to the reference genome size

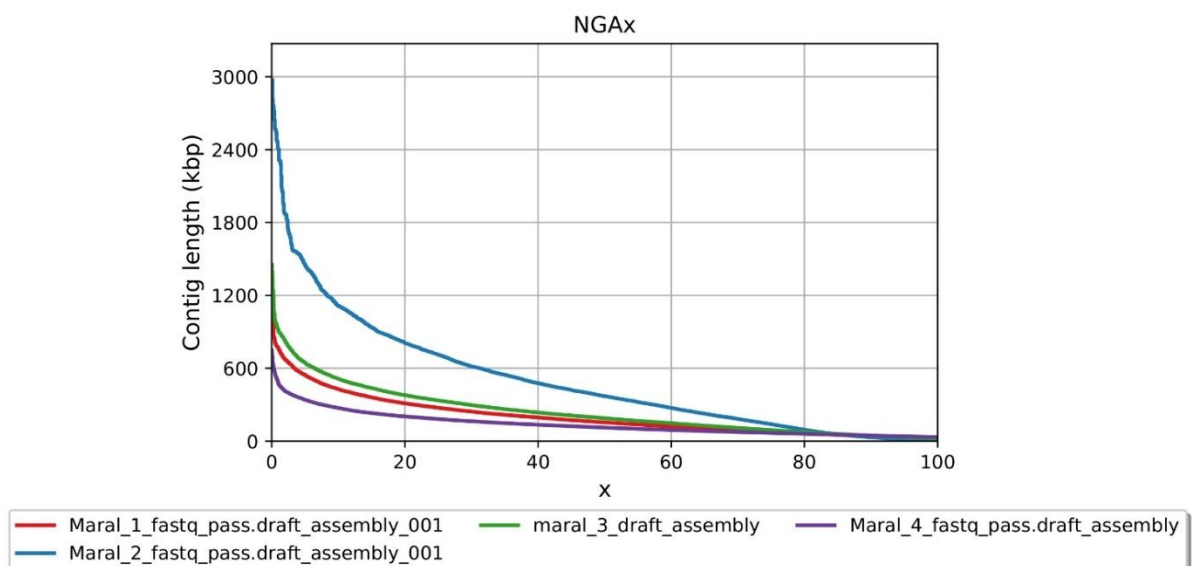
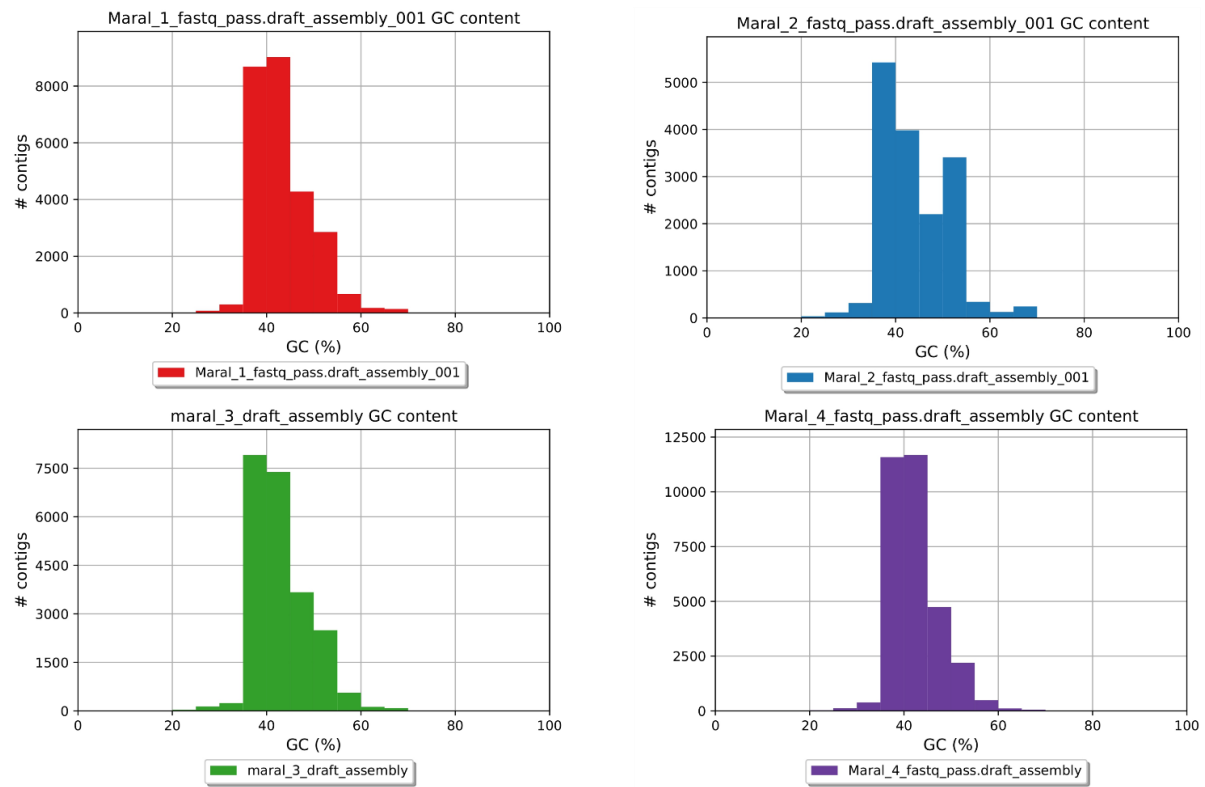


Figure A3

GC content of four maral assemblies



Appendix B

Multiple Sequence Alignments of BMP Proteins

Figure B1

Multiple sequence alignment of BMP2 across mammalian species generated in Jalview

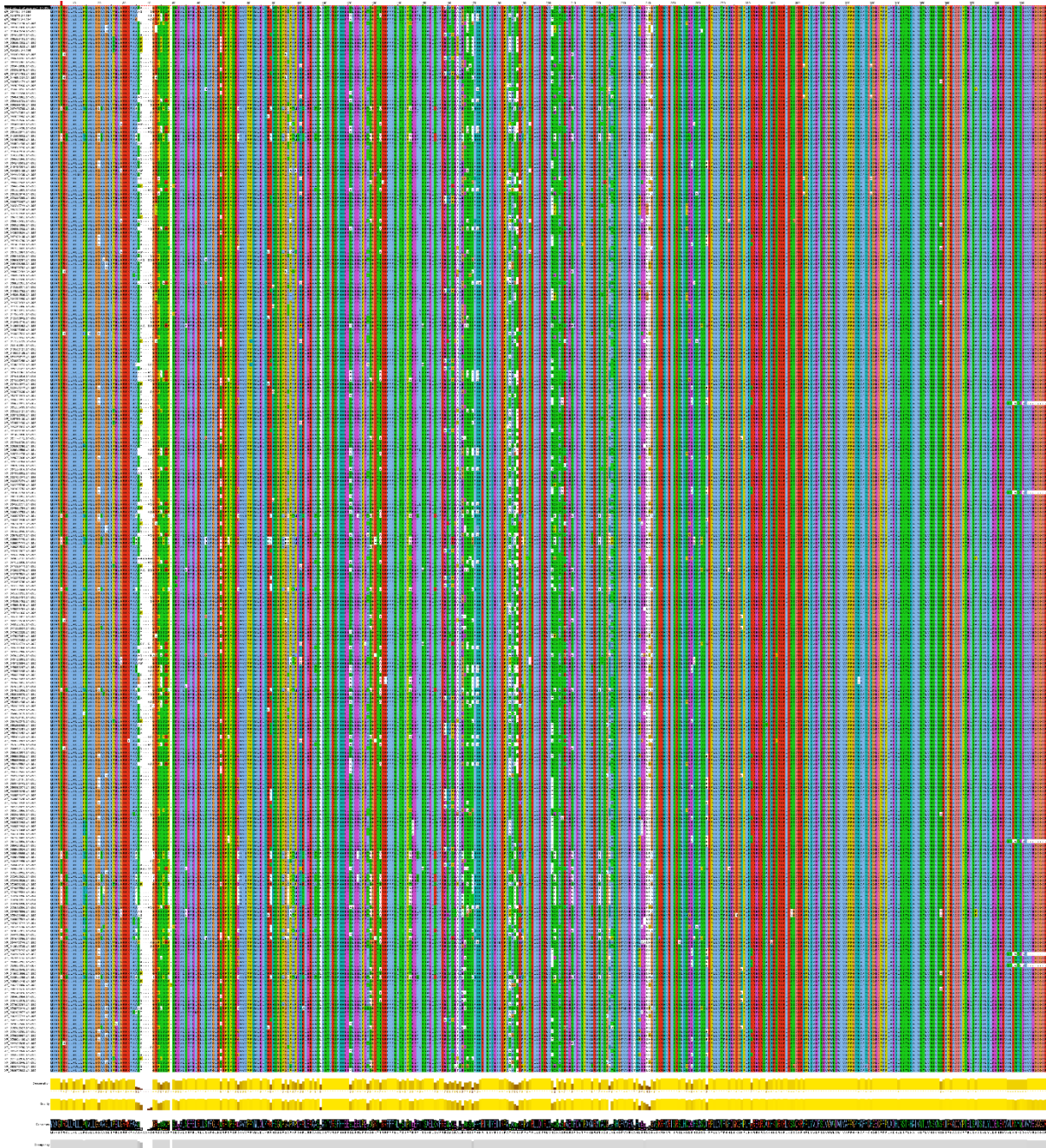


Figure B2

Multiple sequence alignment of BMP3 across mammalian species generated in Jalview

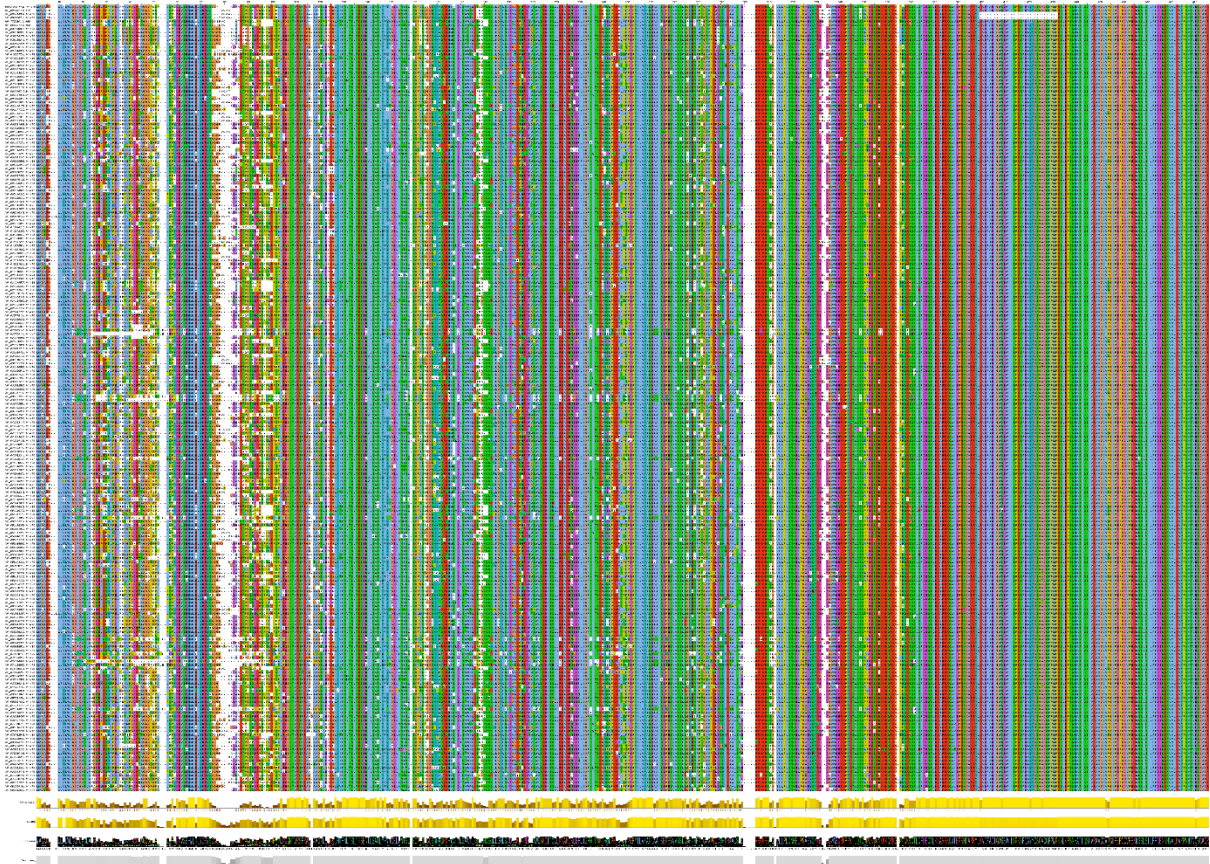


Figure B3

Multiple sequence alignment of BMP4 across mammalian species generated in Jalview

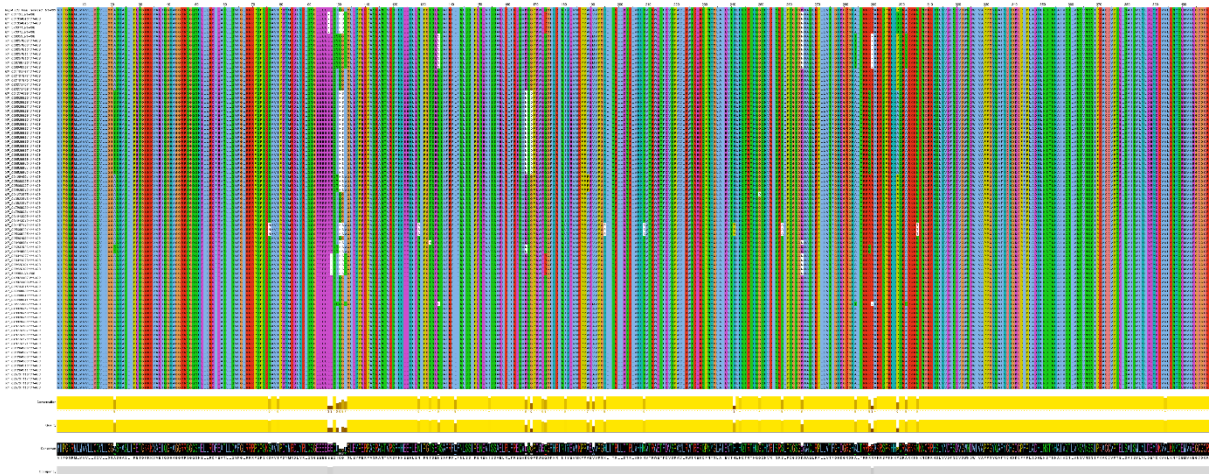


Figure B4

Multiple sequence alignment of BMP7 across mammalian species generated in Jalview

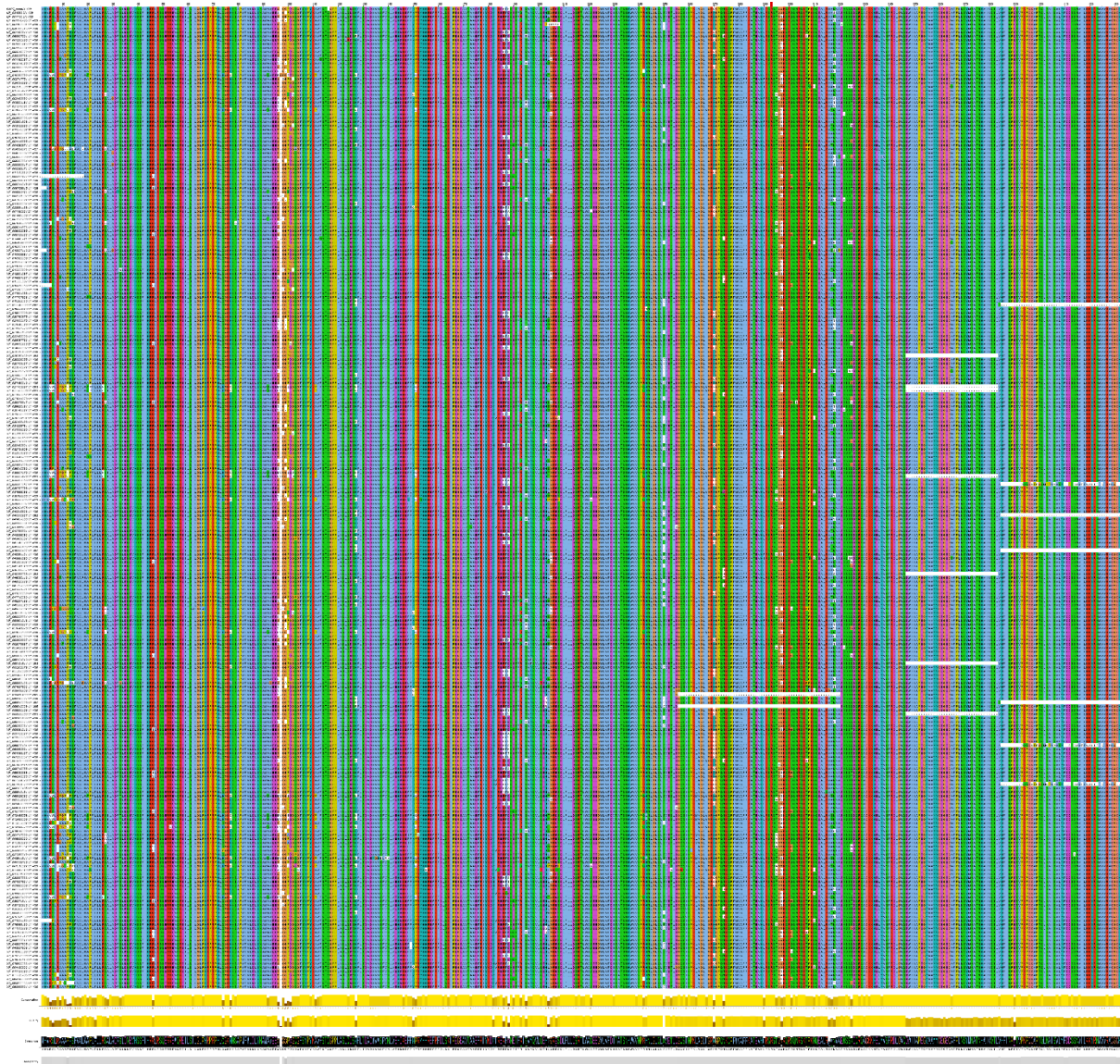


Figure B5

Multiple sequence alignment of BMP9 across mammalian species generated in Jalview

