

## Positive selection drives the evolution of endocrine regulatory bone morphogenetic protein system in mammals

Hafiz Ishfaq Ahmad<sup>1</sup>, Muhammad Jamil Ahmad<sup>1</sup>, Muhammad Muzammal Adeel<sup>3</sup>, Akhtar Rasool Asif<sup>2</sup> and Xiaoyong Du<sup>1,3</sup>

<sup>1</sup>Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of the Ministry of Education, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, P.R. China

<sup>2</sup>University of Veterinary and Animal Sciences, Lahore, Sub Campus Jhang, Pakistan

<sup>3</sup>Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, P.R. China

**Correspondence to:** Xiaoyong Du, **email:** duxiaoyong@mail.hzau.edu.cn

**Keywords:** selection; BMPs; nucleotide substitutions; maximum likelihood; evolution

**Received:** July 14, 2017 **Accepted:** December 06, 2017 **Epub:** January 13, 2018 **Published:** April 06, 2018

**Copyright:** Ahmad et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

The rapid evolution of reproductive proteins might be driven by positive Darwinian selection. The bone morphogenetic protein family is the largest within the transforming growth factor (TGF) superfamily. A little have been known about the molecular evolution of bone morphogenetic proteins exhibiting potential role in mammalian reproduction. In this study we investigated mammalian bone morphogenetic proteins using maximum likelihood approaches of codon substitutions to identify positive Darwinian selection in various species. The proportion of positively selected sites was tested by different likelihood models for individual codon, and M8 were found to be the best model. The percentage of positively elected sites under M8 are 2.20% with  $\omega = 1.089$  for BMP2, 1.6% with  $\omega = 1.61$  for BMP 4 0.53% for BMP15 with  $\omega = 1.56$  and 0.78% for GDF9 with  $\omega = 1.93$ . The percentage of estimated selection sites under M8 is strong statistical confirmation that divergence of bone morphogenetic proteins is driven by Darwinian selection. For the proteins, model M8 was found significant for all proteins with  $\omega > 1$ . To further test positive selection on particular amino acids, the evolutionary conservation of amino acid were measured based on phylogenetic linkage among sequences. For exploring the impact of these somatic substitution mutations in the selection region on human cancer, we identified one pathogenic mutation in human BMP4 and one in BMP15, possibly causing prostate cancer and six neutral mutations in BMPs. The comprehensive map of selection results allows the researchers to perform systematic approaches to detect the evolutionary footprints of selection on specific gene in specific species.

### INTRODUCTION

The bone morphogenetic protein family is the largest within the transforming growth factor (TGF) superfamily and the distinguished structural feature of TGF superfamily is the presence of seven conserved cysteine, which are involved in folding of molecule into distinct three dimensional structure called cysteine knot [1]. Recent studies revealed that BMP is an important component of regulatory system in endocrine tissues and

various BMP functions have been observed in ovaries, pituitary and adrenal glands [2] and also a role in bone formation or differentiation [3]. Therefore it is likely that recruitment, selection and atresia of developing follicles, ovulation and lutenization or luteolysis are escorted by spatial and temporal changes in BMP genes expression pattern [2]. BMP4 expressed in theca cells enhance follicle stimulating hormone (FSH) induced estradiol production and reduce production of progesterone [4]. BMP2 and 4 enhance FSH induced estradiol production by

stimulating FSH induced mitogen activated protein kinase (p38-MAPK) phosphorylation [5]. These findings have been extended to over expression studies in mouse [6], *Xenopus* [7], and chick [8], which indicates that BMP3 negatively regulates the BMPs and activin pathways, while the defined mechanism of inhibition is still unclear. BMP15 is solely expressed in oocytes [9, 10] and has been found to induce granulosa cell mitosis and reduce FSH action by inhibiting follicle stimulating hormone receptor expression [11]. FSH induced expression of 3 $\beta$ -hydroxysteroid dehydrogenase, luteinizing hormone receptor subunits, steroidogenic acute regulatory protein, and steroid side chain cleavage enzyme (p450scc) all are suppressed by BMP15 [11]. The findings revealed that FSH induced progesterone synthesis is inhibited by BMP15, like BMP4, BMP6, BMP7 and GDF9, BMP15 is also a part of luteinization inhibitors group [12]. BMP15 also stimulates the expression of kit ligand mRNA in granulosa cells [13]. BMP ligands and receptors are also expressed in adult pituitary glands, specifically BMP6, 7 and 15 mRNAs are expressed in mice pituitary [14], GDF9 mRNA in human pituitary, BMP15 and GDF9 mRNAs in brush tail opossum and BMP15 in sheep pituitaries [11]. Codon based likelihood models have been extremely used in recent development and have proven remarkably useful in selective pressure studies in various systems [15, 16]. An unequivocal evidence of positive selection in molecular evolution is remarkably higher in non-synonymous than synonymous substitution rate and the ratio dN/dS indicated here by  $\omega$ , quantifies the magnitude and direction of selection pressure on protein, with  $\omega = 1$ ,  $\omega = > 1$ , and  $\omega = < 1$  indicate neutral evolution, positive selection and purifying selection respectively [17]. These conditions have been used to study evolution of male reproductive protein in vertebrate and invertebrate species [18]. To expound the selection pressure underlying the rapid evolution of bone morphogenetic proteins, herein we perform an analysis revealing that positive Darwinian selection drives the evolution of bone morphogenetic proteins in several mammals. We revealed that positive selection has driven on bone morphogenetic proteins as evidenced by population genetic signals such as greater number of non-synonymous substitution rates, long range of haplotype homozygosity and lower genetic diversity. Our results support the hypothesis that there was a rapid evolutionary pressure on mammalian bone morphogenetic proteins genes during evolution.

## RESULTS

The  $\omega$  ratios for all bone morphogenetic proteins across the sites are  $< 1$  (Table 1). However, these proteins might have conserved amino acid and are showing purifying selection with  $\omega < 1$ . The conserved amino acids might mask the signals of positive selection and adaptable amino acids showed positive selection, that

were exposed or hidden residues according to the neural network algorithm for BMP2, BMP4, BMP15 and GDF9 (Supplementary Figures 1–4 respectively). The dN/dS ratio was an average of all positions, and so average dN/dS could not identify the positive selection precisely [24, 25]. Hence, the positive selection was tested particular amino acids using ML tests that measure selective pressures among sites stated by  $\omega$  values [19, 20].

### Bone morphogenetic proteins

Positive selection was found in BMP2, BMP4, BMP15 and GDF9. We performed log likelihood test for all BMP proteins and the  $\omega$  was estimated for all sites. We compared various likelihood tests (M1 vs. M2, and M7 vs. M8 respectively) to determine positive selection. Parameter estimates under M1 and M2 were compared and there was positive selection in M2 for two of four proteins. The proportions of positive selection sites were 0.31% with  $\omega = 2.87$  for BMP4 and 2.23% with  $\omega = 2.12$  for BMP15 (Table 1). M8 was significant for all bone morphogenetic proteins. The percentage of positively selected sites under M8 are 2.20% for BMP2 with  $\omega = 1.089$ , 0.78% for GDF9 with  $\omega = 1.93$ , 1.6% for BMP4 with  $\omega = 1.61$  and 0.53% for BMP15 with  $\omega = 1.56$ .

### Positive selection on amino acids

To identify amino acid positions under selection pressure, we used the Bayes approaches to approximate the posterior probabilities for individual codon position. The codon with higher probabilities is likely to be under positive selection with  $\omega > 1$  [25]. Using Bayes Empirical Bayes (BEB) analysis for BMP2 had a total of 391 amino acids sites, and seven sites were detected under positive selection (Table 2; Figure 1). Only one of the seven sites has posterior probability  $> 95\%$  and the position of site is shown in protein structure (Figure 2). GDF9 has four hundred and fifty three amino acids, and only seven were found under positive selection and BMP4 had 401 amino acids, and eight were found under positive selection (Figure 2). Two of these eight sites are positively selected at posterior probability  $> 99\%$  and  $95\%$  respectively (Table 2; Figure 1). As well BMP15 has three hundred and ninety one amino acid of seventeen positive selection sites but no codon site could be recognized at  $99\%$  or  $95\%$  posterior probabilities (Table 2; Figure 1).

One neutral single nucleotide polymorphism p.S38S (score 0.15) was found in BMP2 and one pathogenic mutations causing prostate cancer, p.T214T was identified in BMP4 (Table 3). p.S38S is found in the functionally quiet pro-domain of BMP2 and is predicted by FATHMM and to be benign/tolerated. p.T214T is expected to be possibly detrimental or to disturb the protein structure. However, a basic amino acid residue at position 38 (S) is essential for recognition or cleavage of site. Serine is

**Table 1: Model parameter estimates, dN/dS ratios, log likelihood values and test statistics for PAML site models of positive selection in mammalian bone morphogenetic proteins**

Gene	<i>n</i>	<i>Lc</i>	<i>S</i>	<i>dN/dS</i>	Model	Parameter estimates	2ΔIM2 vs. M1	2ΔIM8 vs. M7	Positively selected sites
BMP2	39	398	3.5	0.08	M1	<i>P</i> 1 = 0.92674 <i>P</i> 2 = 0.07326 $\omega$ 1 = 0.04427 $\omega$ 2 = 1.00000	0	8.14*	<u>37</u> , 38, 120, <u>126</u> , <u>162</u> , 183, 190, 239
					M2	<i>P</i> 1 = 0.92674 <i>P</i> 2 = 0.05049 <i>P</i> 3 = 0.02277 $\omega$ 1 = 0.04427 $\omega$ 2 = 1.00000 $\omega$ 3 = 1.00000			
					M7	<i>p</i> = 0.16499 <i>q</i> = 1.54264			
					M8	<i>P</i> 0 = 0.97794 <i>p</i> = 0.21124 <i>q</i> = 2.75784 <i>P</i> 1 = 0.02206 $\omega$ 1 = <b>1.08959</b>			
GDF9	33	457	8.1	0.31	M1	<i>P</i> 1 = 0.59348, <i>P</i> 2 = 0.40652 $\omega$ 1 = 0.12471, $\omega$ 2 = 1.00000	0.48	2.59	30, 186, <b>245</b> , 254, <b>292</b> , <u>302</u> , <u>304</u>
					M2	<i>P</i> 1 = 0.59353, <i>P</i> 2 = 0.33608, <i>P</i> 3 = 0.07040 $\omega$ 1 = 0.12473, $\omega$ 2 = 1.00000, $\omega$ 3 = 1.00000			
					M7	<i>p</i> = 0.48463 <i>q</i> = 0.91798			
					M8	<i>P</i> 0 = 0.99219 <i>p</i> = 0.49979 <i>q</i> = 0.97580 <i>P</i> 1 = 0.00781, $\omega$ 1 = <b>1.93407</b>			
BMP4	29	570	4.3	0.09	M1	<i>P</i> 1 = 0.93285, <i>P</i> 2 = 0.06715 $\omega$ = 0.05872, $\omega$ 1 = 1.00000	3.41	26.91**	<b>255</b> , <b>256</b> , <u>259</u> , 333, <b>349</b> , <u>351</u> , 375, 425
					M2	<i>P</i> 1 = 0.93307, <i>P</i> 2 = 0.06381, <i>P</i> 3 = 0.00312 $\omega$ 1 = 0.05939, $\omega$ 2 = 1.00000, $\omega$ 3 = <b>2.87878</b>			
					M7	<i>p</i> = 0.23270 <i>q</i> = 1.77764			
					M8	<i>P</i> 0 = 0.98531 <i>p</i> = 0.31720 <i>q</i> = 3.28217 <i>P</i> 1 = 0.01469 $\omega$ 1 = <b>1.61688</b>			
BMP15	86	434	13.5	0.41	M1	<i>P</i> 1 = 0.54292 <i>P</i> 2 = 0.45708 $\omega$ 1 = 0.17179 $\omega$ 1 = 1.00000	8.4*	17.69**	31, 37, 89, <b>113</b> , 154, <b>169</b> , 177, 178, 229, 285, 335
					M2	<i>P</i> 1 = 0.53569 <i>P</i> 2 = 0.44200 <i>P</i> 3 = 0.02231 $\omega$ 1 = 0.17280 $\omega$ 2 = 1.00000 $\omega$ 3 = <b>2.12476</b>			
					M7	<i>p</i> = 0.60414 <i>q</i> = 0.76298			
					M8	<i>P</i> 0 = 0.94697 <i>p</i> = 0.66983 <i>q</i> = 0.97138 <i>P</i> 1 = 0.05303 $\omega$ 1 = <b>1.56019</b> <i>P</i> 1 = 0.21259 $\omega$ 1 = 1.00000			

The data have *n* sequences, each of *Lc* codons after alignment gaps are removed. *S* is the tree length, measured as the number of nucleotide substitutions per codon. The proportion of sites under positive selection (*p*1), or under selective constraint (*p*0), and parameters *p* and *q* for the beta distribution. Parameters indicating positive selection are in bold. *p*: significant at 5% level; *pp*: significant at 1% level. Sites potentially under positive selection identified under model M8 are listed according to the human sequence numbering. Positively selected sites with posterior probability 0.9 are underlined, 0.8–0.9 in bold, and 0.5–0.7 in plain text. The test statistic *2ΔI* is compared to a  $\chi^2$  distribution with 2 degrees of freedom, critical values 5.99, 9.21, and 13.82 at 5%, 1%, and 0.1% significance, respectively. \*\*: significant at 1% level; \*: significant at 5% level.

a polar amino acids and p.S38S is therefore suspected to have influence on post-translational cleavage and biosynthesis of protein. This mutation is anticipated to be benign or tolerated by FATHMM (Table 3). Only one mutation, p.T214T was identified genomic region of BMP4 encoding functional domain of TGF- $\beta$ .

The codon sites that have undergone alteration of gene can direct to more chances of false positive results during analysis when using ML approach to

identify positive selection, predominantly in small data set which only a few sequences, though false positive rate is moderately increased [21]. To reduce the gene conversion influence on results of the study, each set of sequence was analyzed individually and the most similar sequences, which are the results of gene conversion were excluded from the analysis. Furthermore, a BEB analysis was used instead of NEB to recognize positively selected codon because NEB is less conservative and can be more

**Table 2: Positively selected sites under different PAML site models using bayes empirical bayes analysis**

Gene	Model	Codon	Amino Acid	Posterior Probability	Post mean $\pm$ SE for $\omega$
BMP-2	M8: selection, beta+ $\omega$	38	S	0.695	1.187 $\pm$ 0.532
		41	P	0.632	1.114 $\pm$ 0.554
		43	S	0.713	1.230 $\pm$ 0.472
		118	L	0.597	1.079 $\pm$ 0.555
		164	N	0.611	1.087 $\pm$ 0.569
GDF-9	M8: selection, beta+ $\omega$	236	K	0.607	1.115 $\pm$ 0.518
		186	S	0.585	1.225 $\pm$ 0.335
		253	L	0.696	1.300 $\pm$ 0.309
		290	G	0.832	1.395 $\pm$ 0.238
BMP-4	M8: selection, beta+ $\omega$	302	V	0.938*	1.463 $\pm$ 0.148
		99	I	0.823	1.368 $\pm$ 0.311
		100	H	0.827	1.370 $\pm$ 0.317
		102	T	0.998**	1.512 $\pm$ 0.123
		173	R	0.506	1.075 $\pm$ 0.449
		188	A	0.867	1.401 $\pm$ 0.309
		190	V	0.986*	1.503 $\pm$ 0.143
		214	T	0.536	1.071 $\pm$ 0.488
BMP-15	M8: selection, beta+ $\omega$	264	N	0.515	1.073 $\pm$ 0.461
		22	R	0.590	1.239 $\pm$ 0.368
		28	G	0.753	1.361 $\pm$ 0.332
		80	S	0.544	1.198 $\pm$ 0.392
		104	V	0.846	1.426 $\pm$ 0.285
		127	L	0.514	1.393 $\pm$ 0.236
		145	R	0.764	1.369 $\pm$ 0.322
		160	P	0.615	1.255 $\pm$ 0.376
		168	E	0.703	1.315 $\pm$ 0.291
		169	G	0.759	1.365 $\pm$ 0.329
		220	L	0.556	1.212 $\pm$ 0.373
273	S	0.547	1.198 $\pm$ 0.397		
		323	T	0.717	1.334 $\pm$ 0.339

Bayes Empirical likelihood ratio test statistic for model M8: selection, beta+  $\omega$ , indicate posterior probability  $P > 95\%$  (\*) and  $P > 99\%$  (\*\*). For codon position, the amino acid number is given followed by an abbreviation.

exposed to false positive results [22, 24]. While BEB is more conservative and yields a less chances of false positive results with codon sites that are influences by gene conversion [21].

## DISCUSSION

Recently, male derived molecules have been shown to be extraordinarily involved in reproduction between closely related species, which include proteins involved in signaling between males and females fertilization. Preceding studies using ovarian theca interna cell cultures revealed that various BMPs (BMP2, 4, and 6) potentially suppress LH induced androgen secretion [23]. BMPs along with their receptors and extracellular binding proteins (e.

g., chordin, noggin and gremlin) are extensively expressed in granulosa, theca cells and ovarian stroma [24, 25]. It is likely that BMPs exert autocrine as well as paracrine action to regulate steroidogenesis and other ovarian functions [26]. Previous studies directed us to hypothesize that evolution of BMPs has emerged implicating a variety ovarian factors having modulatory role in reproduction process. As recently in evolution, the mature domains of BMP ligands BMP2, 4, 6, and 7 shares 40% amino acid identity in phylogenetic clade [27]. For most of BMP sites under positive selection, a correlation was found among the sites and interaction with other molecules. To validate the correlation between BMPs and their functions, we detected positive selection in BMP2, 4, 15 and GDF9 based on  $\omega$  (dN/dS) ratio, is useful for estimation of selection pressure of genes [22, 23]. The

**Table 3: Prediction of pathogenic point amino acid substitutions mutation was estimated from the usage of functional analysis through hidden Markov model (FATHMM)**

Gene	Codon	SnP Id	Tissue Distribution	FATHMM prediction: (Functional Analysis through Hidden Markov Models)			
BMP-2	38	COSM1029277	Endometrium(1) /Prostate(1) p.S38S	Neutral (score 0.15)			
	41						
	43						
	118						
	164						
	236						
GDF-9	186						
	253						
	290						
	302						
BMP-4	99	XXX	Lung p.H100Y	Pathogenic (score 0.98)			
	100						
	102						
	173						
	188						
	190						
BMP-15	214	XXX	Hematopoietic and lymphoid tissue(1)/Prostate(1) p.T214T	Pathogenic (score 0.84)			
	264						
	22						
	28						
	80						
	104				COSM4649428	Large intestine(1) p.V104M	Neutral (score 0.03)
	104				COSM6187212	Lung(1) p.V104A	Neutral (score 0.03)
	127						
	145						
	160						
	168				COSM385794	Lung(1) p.E168K	Neutral (score 0.10)
169	COSM3562207	Skin(1) p.G169R	Neutral (score 0.01)				
220							
273							
323	COSM309487	Lung(2) p.T323T	Neutral (score 0.01)				

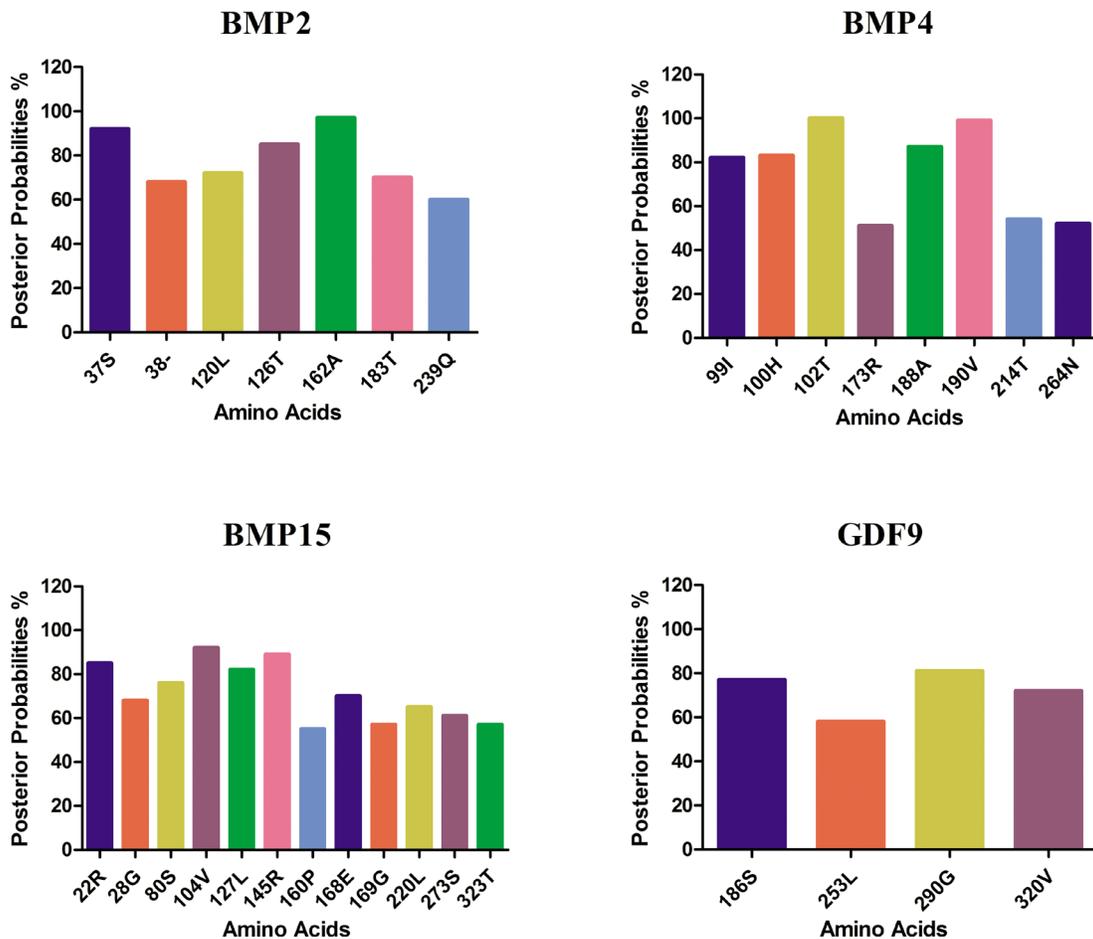
$\omega > 1$  indicates positive selection [28]. In our study positive selection was found with  $\omega > 1$  in BMP2, 4, 15 and GDF9 (Table 1). This shows that non-synonymous (dN) positions evolved faster than those of synonymous positions and the Darwinian selection effect purifying or balancing selection preferred new variations and higher allelic polymorphism [29] that might insert new variations in protein structure

confirmation, thus disturbing the signaling pathways [30]. The amino acid replacements among species might be a result of distinct deviation from their shared ancestries which agrees with previous studies [31] that as orthologs vary from their most recent common ancestors, their different evolutionary routes direct to deviation in the discerning restraints on homologous sites [32].

In this study we predicted the linkage between germ line mutation in BMPs and risk of prostate cancer. We identified one pathogenic mutation, p.T214T in BMP4, causing prostate cancer and one neutral mutation p.S38S in BMP2. p.T214T localizes at N-terminal of TGF- $\beta$ 1 pro-domain pruning the protein prior to active domain. BMP4 play a role in the osteogenesis of the PCa-118b xenograft. The confirmation for this declaration is provided by the fact that deactivating mutations in SMAD 4 and BMPR1A [33], that are members of the TGF- $\beta$  superfamily and cause prostate cancer. Although the bone forming phenotype in prostate cancer, bioinformatics analysis of the mutations identified as pathogenic in BMP4 and exposed that both osteoblastic and osteolytic lesions are present in the same loci. These facts provide strong indication that haplo-insufficiency of BMP4 will elevates PCa risk. About 70% of missense mutations have confirmed the presence 1% of population frequencies [34]. While we identified one neutral and one pathogenic mutation in BMP2 and BMP4 correspondingly mapped on pro-domain of the expressed protein. The cumulative risk of PCa is a result of variants causing rare disorder. The usage of various tools for

next generation sequencing approaches, to screen PCA causative variants will need expression of variants through bioinformatics techniques.

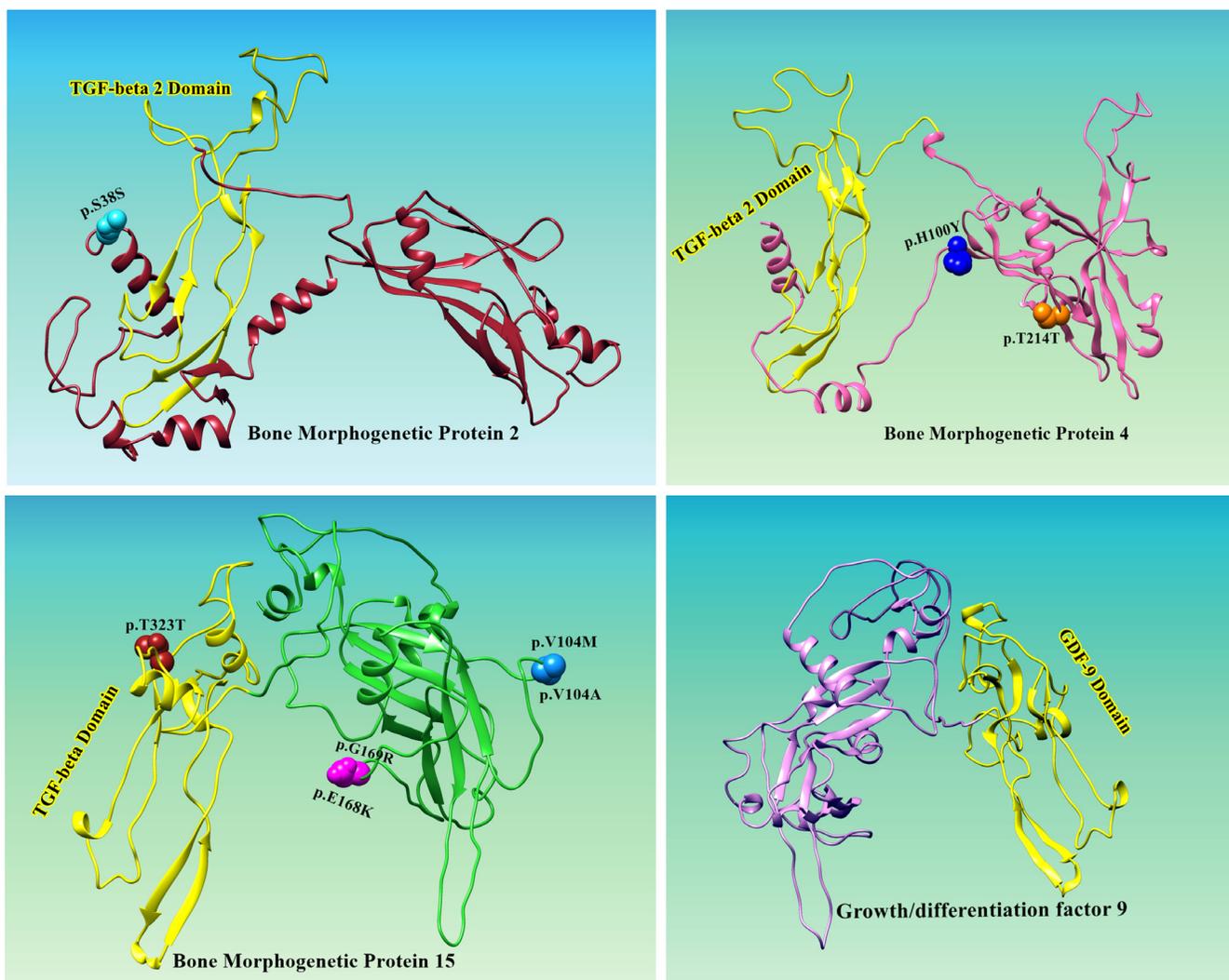
Three sites were found under positive selection using the maximum likelihood model. The conventional models M1 vs. M2 comparison did not show significance for BMP2 and GDF9 but it was statistically significant for BMP4 and BMP15 identified 0.31% and 2.23% positive selection with  $\omega$  values  $\omega_3 = 2.87$  and 2.12 respectively (Table1). The results achieved under different sets of models (M1 vs. M2 and M7 vs. M8) vary in some aspects. While the M7 vs. M8 comparison show significant difference from former models, allowing positive selection for BMP2, 4, 15 and GDF9. For BMP4 and BMP15, both M1 vs. M2 and M7 vs. M8 were significant, the comparison detected same eight sites for BMP4 and seventeen codon sites for BMP15 had been identified under M1 vs. M2 comparison. Additionally, the positive selection signals were tested in birds and reptiles [35]. The observed positive selection in birds was a distinctive signal and not a pervasive trend, since only ~22% of genes exhibited sign of positive selection in reptiles. Likewise,



**Figure 1: Amino acid residues identified likely to be under positive selection by bayes empirical bayes.** The amino acid sites of  $\omega > 1$  under M8 model. The posterior probability of each site was calculated by BEB. Sites show positive selection at different posterior probabilities.

the presence of positive selection shown different targets bone-associated genes in different clades. Of the 89 genes, ~18% were detected under positive selection in both mammals and birds, only 12.4% were identified in only mammals and 34.8% sites in genes encoding proteins were found under positively selection in birds and involved in bone resorption [36]. The variations in the results found using various models revealed that M1 vs. M2 comparison is more conservative test which may be unable to identify positively selected sites detected by less conservative models M7 vs. M8 comparisons [19, 20]. It is remarkable to find that for BMP2 and BMP4, the amino acid sites that were detected having experience of positive selection are located mature extracellular domains of

BMPs receptors which are involved in oocyte maturation and early embryonic development [37, 38]. BMP receptors composed of extra cellular domains, membrane bound domains and intracellular domains with active serine and threonine regions [39]. BMPs initiate signal transduction cascade by forming heterodimer complex through binding cell surface receptors and this complex consists of serine and threonine kinase receptors [40]. Three of four types of receptors present in TGF- $\beta$  family interact with BMPs for example; BMP2 and BMP4 bind type I receptors and recruit type II receptors, while BMP6 and BMP7 interact with type II and recruit type I receptors [37, 38]. Moreover the distinct expression patterns of BMP4 and BMP6 mRNA in different systems propose that BMPs are



**Figure 2: Location of positively selected amino acid sites identified BMP2, BMP4, BMP15 and GDF9 genes.** Three dimensional structure prediction of BMPs and GDF9 was carried out by using Ab-initio modeling approach. Primary sequences of human BMP2 (ACV32596.1), BMP4 (AAH20546.1), BMP15 (AAI17265.1) and GDF9 (AAH96229.1) were subjected to I-TESSAR to predict suitable structures. Structure validation of all predicted models was done by MolProbity server. To test the steric hindrance of amino acid residues Ramachandran values were calculated by using Ramachandran Plot2.0 tool. UCSF Chimera was applied for visualization and geometry optimization of predicted proteins. All the residues identified as under selection fall in the domain containing the ligand binding site. The sites which fall in the region identified as the ligand binding site and another cluster in a region immediately following the signal sequence.

involved in events that control embryonic development pattern [41]. Although the signaling pathways for BMP ligands have been studied, there are alternative pathways for BMPs that mediate biological activity in various cell types [42], particularly the MAPK signaling molecules family, including p38, ERK1/2 and N-terminal kinases have been shown to exhibit intracellular transduction of BMP signal pathways which regulate granulosa cells function [43].

Selection pressures occurring indifferent lineages may result in parallel or convergent alterations at amino acid site refer to amino acids changes from different ancestral to the same descendant amino acid along independent evolutionary lineages [44, 45]. Among the positively detected sites we observed that all sites fall in extracellular ligands binding domains for a prodomain folding and C-terminal mature peptide except for five sites of which one for BMP2 (37S) and four for BMP15 (40V, 49I, 75Q and 90R) were at N-terminus (Figure 2; BMP2, BMP15) and C-terminal proteins are cleaved proteolytically upon dimerization at an Arginine sequence by serine endoprotease from prodomain [46]. Previous studies revealed that BMPs are consist of 50 to 100 amino acids with seven cysteine, of which six form cysteine knot and the seventh cysteine is used for dimerization, thus developing the biologically active signaling molecules [47]. All the BMPs; BMP2, 4, 6 and 7 have seventh cysteine and form homodimer or heterodimer except BMP3 and BMP15 lack seventh cysteine but are biologically active as monomers [48, 49]. Among molecular level positive selection reported in previous studies favors more abundant non-synonymous substitutions and here we detected positive selection having all non-synonymous substitutions that occurred as results of duplication of genes during evolution in mammals [50, 51].

## MATERIALS AND METHODS

### Sequence analysis and dataset preparation

The nucleotide and amino acid sequences of BMP genes retrieved from GenBank ([www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank)) (Supplementary Table 1) and accomplished sequences of proteins were aligned using the MUSCLE [52], implemented in MEGA6.0 program using the amino acid sequences and back translated to nucleotide for selection analysis [53].

### Positive selection analysis

In order to identify codons under positive selection, only BMPs and GDF9 that were represented in at least 20 species were assessed, as conferred by poon and collaborators [54]. Hence, BMP6 and BMP7 were excluded from analyses as the multiple sequence alignment generated were not reliable and prone to affect the recognition of selection, leading false positive results

[55]. Phylogenetic analysis was performed on accepted mammalian phylogeny [56] by generating un-rooted tree of aligned species. Branch lengths were calculated using tree topology using the codon model in PAML package. The different  $\omega$  ratios (dN/dS) were compared to identify selection pressure in particular codons using maximum-likelihood methods implemented in the MEGA6 [54] and PAML version 4 [57].

We compared different likelihood ratio tests. The M7 (null model) assumes  $\beta$  distribution with  $\omega$  in limited (0 and 1) interval. The M8 is an alternative model that includes two parameters ( $\omega$  and beta), so  $\omega$  value achieved from the data were greater than one. Additionally, to find out amino acid exposed to selection were inferred using Bayes theorem by estimating posterior probabilities for each site [57, 58]. Three dimensional structure prediction of BMPs and GDF9 was carried out by using Ab-initio modeling approach [59]. Primary sequences of BMP2 (ACV32596.1), BMP4 (AAH20546.1), BMP15 (AAI17265.1) and GDF9 (AAH96229.1) were subjected to I-TESSAR [60] to predict suitable structures. Structure validation of all predicted models was done by MolProbity server [61]. To test the steric hindrance of amino acid residues Ramachandran values were calculated by using Ramachandran Plot2.0 tool [62]. UCSF Chimera [63] was applied for visualization and geometry optimization of predicted proteins. The ConSurf server was used to predict the level of evolutionary conservation amino acid sites in protein based on phylogenetic linkage among sequences [64]. For a more traditional approach, and as used previously [65], positive selection sites detected in more than one maximum likelihood approach were considered. We found that the statistical approaches used in this study are able to determine positive selection, but cannot deliver information about positive selection mechanism. Therefore, to printout the location of positively selected amino acid residues might be helpful for additional laboratory examination.

For these coding sites subjected to positive selection, we used the COSMIC (Catalogue of Somatic Mutations in Cancer) v82 (released 03-AUG-17) database for exploring the impact of these somatic substitution mutations in human cancer [66]. The COSMIC database includes hundreds of thousands of human cancer-associated somatic mutations that are classified by tumor type and disease. The prediction of pathogenic point amino acid substitutions mutation was estimated from the usage of Functional Analysis through Hidden Markov Model (FATHMM) [67].

## CONCLUSIONS

Our study reveals the advantages in combining various approaches to explore selection pressure and molecular evolution in biological systems, predominantly in genes intensely knotted to ecology, and highlights

the significance of studies integrating natural sequence variation in organisms from various environments. Selection studies of bone morphogenetic proteins could expedite the improvement of distinctive approaches. These methodologies could possibly play a vital role for selection of higher breeding values and that their genetic enrichment to produce next generation.

### Author contributions

HIA wrote original draft of manuscript, MJA, ARA, MMA, and XD helped in data analysis and XD supervised the study.

### ACKNOWLEDGMENTS AND FUNDING

This work was supported by the National Nature Scientific Foundation of China (Grant No. 31402040 and Grant No. 31572375), the Fundamental Research Funds for the Central Universities (2662017JC027, 2662016PY006 and 2662015BQ024). The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

### CONFLICTS OF INTEREST

None declared.

### REFERENCES

1. Vitt UA, Hsu SY, Hsueh AJ. Evolution and classification of cystine knot-containing hormones and related extracellular signaling molecules. *Mol Endocrinol*. 2001; 15:681–94. <https://doi.org/10.1210/mend.15.5.0639>.
2. Otsuka F. Multiple endocrine regulation by bone morphogenetic protein system. *Endocr J*. 2010; 57:3–14. <https://doi.org/10.1507/endocrj.K09E-310>.
3. Takano M, Otsuka F, Matsumoto Y, Inagaki K, Takeda M, Nakamura E, Tsukamoto N, Miyoshi T, Sada KE, Makino H. Peroxisome proliferator-activated receptor activity is involved in the osteoblastic differentiation regulated by bone morphogenetic proteins and tumor necrosis factor- $\alpha$ . *Mol Cell Endocrinol*. 2012; 348:224–32. <https://doi.org/10.1016/j.mce.2011.08.027>.
4. Shimasaki S, Zachow RJ, Li D, Kim H, Iemura S, Ueno N, Sampath K, Chang RJ, Erickson GF. A functional bone morphogenetic protein system in the ovary. *Proc Natl Acad Sci U S A*. 1999; 96:7282–7. <https://doi.org/10.1073/pnas.96.13.7282>.
5. Inagaki K, Otsuka F, Miyoshi T, Yamashita M, Takahashi M, Goto J, Suzuki J, Makino H. p38-Mitogen-activated protein kinase stimulated steroidogenesis in granulosa cell-oocyte cocultures: role of bone morphogenetic proteins 2 and 4. *Endocrinology*. 2009; 150:1921–30. <https://doi.org/10.1210/en.2008-0851>.
6. Gamer LW, Cox K, Carlo JM, Rosen V. Overexpression of BMP3 in the developing skeleton alters endochondral bone formation resulting in spontaneous rib fractures. *Dev Dyn*. 2009; 238:2374–81. <https://doi.org/10.1002/dvdy.22048>.
7. Gamer LW, Nove J, Levin M, Rosen V. BMP-3 is a novel inhibitor of both activin and BMP-4 signaling in *Xenopus* embryos. *Dev Biol*. 2005; 285:156–68. <https://doi.org/10.1016/j.ydbio.2005.06.012>.
8. Gamer LW, Ho V, Cox K, Rosen V. Expression and function of BMP3 during chick limb development. *Dev Dyn*. 2008; 237:1691–8. <https://doi.org/10.1002/dvdy.21561>.
9. Otsuka F, Yao Z, Lee T, Yamamoto S, Erickson GF, Shimasaki S. Bone morphogenetic protein-15. Identification of target cells and biological functions. *J Biol Chem*. 2000; 275:39523–8. <https://doi.org/10.1074/jbc.M007428200>.
10. Liang S, Guo J, Choi JW, Shin KT, Wang HY, Jo YJ, Kim NH, Cui XS. Protein phosphatase 2A regulatory subunit B55 $\alpha$  functions in mouse oocyte maturation and early embryonic development. *Oncotarget*. 2017; 8:26979–91. <https://doi.org/10.18632/oncotarget.15927>.
11. Otsuka F, Yamamoto S, Erickson GF, Shimasaki S. Bone morphogenetic protein-15 inhibits follicle-stimulating hormone (FSH) action by suppressing FSH receptor expression. *J Biol Chem*. 2001; 276:11387–92. <https://doi.org/10.1074/jbc.M010043200>.
12. Shimasaki S, Moore RK, Erickson GF, Otsuka F. The role of bone morphogenetic proteins in ovarian function. *Reprod Suppl*. 2003; 61:323–37.
13. Otsuka F, Shimasaki S. A novel function of bone morphogenetic protein-15 in the pituitary: selective synthesis and secretion of FSH by gonadotropes. *Endocrinology*. 2002; 143:4938–41. <https://doi.org/10.1210/en.2002-220929>.
14. Miyoshi T, Otsuka F, Inagaki K, Otani H, Takeda M, Suzuki J, Goto J, Ogura T, Makino H. Differential regulation of steroidogenesis by bone morphogenetic proteins in granulosa cells: involvement of extracellularly regulated kinase signaling and oocyte actions in follicle-stimulating hormone-induced estrogen production. *Endocrinology*. 2007; 148:337–45. <https://doi.org/10.1210/en.2006-0966>.
15. Rennison DJ, Owens GL, Taylor JS. Opsin gene duplication and divergence in ray-finned fish. *Mol Phylogenet Evol*. 2012; 62:986–1008. <https://doi.org/10.1016/j.ympev.2011.11.030>.
16. Yang Z, dos Reis M. Statistical properties of the branch-site test of positive selection. *Mol Biol Evol*. 2011; 28:1217–28. <https://doi.org/10.1093/molbev/msq303>.
17. Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci U S A*. 2001; 98:2509–14. <https://doi.org/10.1073/pnas.051605998>.
18. Rooney AP, Zhang J, Nei M. An unusual form of purifying selection in a sperm protein. *Mol Biol Evol*. 2000; 17:278–83. <https://doi.org/10.1093/molbev/msm088>.

19. Ahmad HI, Liu G, Jiang X, Liu C, Chong Y, Huarong H. Adaptive molecular evolution of MC1R gene reveals the evidence for positive diversifying selection in indigenous goat populations. *Ecol Evol.* 2017; 7:5170–80. <https://doi.org/10.1002/ece3.2919>.
20. Ahmad HI, Liu G, Jiang X, Liu C, Fangzheng X, Chong Y, Ijaz N, Huarong H. Adaptive selection at agouti gene inferred breed specific selection signature within the indigenous goat populations. *Asian-Australas J Anim Sci.* 2017. <https://doi.org/10.5713/ajas.16.0994>. [Epub ahead of print].
21. Ahmad HI, Liu G, Jiang X, Edallew SG, Wassie T, Tesema B, Yun Y, Pan L, Liu C, Chong Y, Yu ZJ, Jilong H. Maximum-likelihood approaches reveal signatures of positive selection in BMP15 and GDF9 genes modulating ovarian function in mammalian female fertility. *Ecol Evol.* 2017; 7:8895–902. <https://doi.org/10.1002/ece3.3336>.
22. Casola C, Hahn MW. Gene conversion among paralogs results in moderate false detection of positive selection using likelihood methods. *J Mol Evol.* 2009; 68:679–87. <https://doi.org/10.1007/s00239-009-9241-6>.
23. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005; 22:1107–18. <https://doi.org/10.1093/molbev/msi097>.
24. Glister C, Richards SL, Knight PG. Bone morphogenetic proteins (BMP) -4, -6, and -7 potently suppress basal and luteinizing hormone-induced androgen production by bovine theca interna cells in primary culture: could ovarian hyperandrogenic dysfunction be caused by a defect in thecal BMP signaling? *Endocrinology.* 2005; 146:1883–92. <https://doi.org/10.1210/en.2004-1303>.
25. Glister C, Satchell L, Knight PG. Granulosa and thecal expression of bone morphogenetic protein- and activin-binding protein mRNA transcripts during bovine follicle development and factors modulating their expression *in vitro*. *Reproduction.* 2011; 142:581–91. <https://doi.org/10.1530/REP-11-0150>.
26. Fenwick MA, Mansour YT, Franks S, Hardy K. Identification and regulation of bone morphogenetic protein antagonists associated with preantral follicle development in the ovary. *Endocrinology.* 2011; 152:3515–26. <https://doi.org/10.1210/en.2011-0229>.
27. Knight PG, Glister C. TGF- $\beta$  superfamily members and ovarian follicle development. *Reproduction.* 2006; 132:191–206.
28. Lowery JW, Lavigne AW, Kokabu S, Rosen V. Comparative genomics identifies the mouse *Bmp3* promoter and an upstream evolutionary conserved region (ECR) in mammals. *PLoS One.* 2013; 8:e57840. <https://doi.org/10.1371/journal.pone.0057840>.
29. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 2005; 22:2472–9. <https://doi.org/10.1093/molbev/msi237>.
30. Bergstrom T, Gyllensten U. Evolution of Mhc class II polymorphism: the rise and fall of class II gene function in primates. *Immunol Rev.* 1995; 143:13–31.
31. Cui HX, Zhao SM, Cheng ML, Guo L, Ye RQ, Liu WQ, Gao SZ. Cloning and expression levels of genes relating to the ovulation rate of the Yunling black goat. *Biol Reprod.* 2009; 80:219–26. <https://doi.org/10.1095/biolreprod.108.069021>.
32. Marini NJ, Thomas PD, Rine J. The use of orthologous sequences to predict the impact of amino acid substitutions on protein function. *PLoS Genet.* 2010; 6:e1000968. <https://doi.org/10.1371/journal.pgen.1000968>.
33. Howe JR, Bair JL, Sayed MG, Anderson ME, Mitros FA, Petersen GM, Velculescu VE, Traverso G, Vogelstein B. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat Genet.* 2001; 28:184–7. <https://doi.org/10.1038/88919>.
34. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet.* 2007; 80:727–39. <https://doi.org/10.1086/513473>.
35. Machado JP, Johnson WE, O'Brien SJ, Vasconcelos V, Antunes A. Adaptive evolution of the matrix extracellular phosphoglycoprotein in mammals. *BMC Evol Biol.* 2011; 11:342. <https://doi.org/10.1186/1471-2148-11-342>.
36. Machado JP, Johnson WE, Gilbert MT, Zhang G, Jarvis ED, O'Brien SJ, Antunes A. Bone-associated gene evolution and the origin of flight in birds. *BMC Genomics.* 2016; 17:371. <https://doi.org/10.1186/s12864-016-2681-7>.
37. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006; 4:e72. <https://doi.org/10.1371/journal.pbio.0040072>.
38. Moore RK, Shimasaki S. Molecular biology and physiological role of the oocyte factor, BMP-15. *Molecular and cellular endocrinology.* 2005; 234:67–73. <https://doi.org/10.1016/j.mce.2004.10.012>.
39. Lan ZJ, Gu P, Xu X, Jackson KJ, DeMayo FJ, O'Malley BW, Cooney AJ. GDNF-dependent repression of BMP-15 and GDF-9 mediates gamete regulation of female fertility. *The EMBO journal.* 2003; 22:4070–81.
40. Lin HY, Wang XF, Ng-Eaton E, Weinberg RA, Lodish HF. Expression cloning of the TGF- $\beta$  type II receptor, a functional transmembrane serine/threonine kinase. *Cell.* 1992; 68:775–85. [https://doi.org/10.1016/0092-8674\(92\)90152-3](https://doi.org/10.1016/0092-8674(92)90152-3).
41. Heldin CH, Miyazono K, Ten Dijke P. TGF- $\beta$  signalling from cell membrane to nucleus through SMAD proteins. *Nature.* 1997; 390:465–71.
42. Nohe A, Hassel S, Ehrlich M, Neubauer F, Sebald W, Henis YI, Knaus P. The mode of bone morphogenetic protein (BMP) receptor oligomerization determines different BMP-2 signaling pathways. *Journal of Biological Chemistry.* 2002; 277:5330–8.
43. De Caestecker M. The transforming growth factor- $\beta$  superfamily of receptors. *Cytokine & growth factor reviews.* 2004; 15:1–11. <https://doi.org/10.1016/j.cytogfr.2003.10.004>.

44. Jones CM, Lyons KM, Hogan BL. Involvement of Bone Morphogenetic Protein-4 (BMP-4) and Vgr-1 in morphogenesis and neurogenesis in the mouse. *Development*. 1991; 111:531–42.
45. Piek E, Heldin CH, Ten Dijke P. Specificity, diversity, and regulation in TGF- $\beta$  superfamily signaling. *The FASEB Journal*. 1999; 13:2105–24. <https://doi.org/10.1096/fasebj.13.15.2105>.
46. Seger R, Hanoch T, Rosenberg R, Dantes A, Merz WE, Strauss JF 3rd, Amsterdam A. The ERK signaling cascade inhibits gonadotropin-stimulated steroidogenesis. *J Biol Chem*. 2001; 276:13957–64. <https://doi.org/10.1074/jbc.M006852200>.
47. Zhang J. Positive Darwinian selection in gene evolution. *Darwin's Heritage Today: Proceedings of the Darwin 200 Beijing International Conference: 24–26 October 2009; Beijing*. Citeseer. 2010; pp. 288–309.
48. Khan FA, Liu H, Zhou H, Wang K, Ul QM, Pandupuspitasari NS, Shujun Z. Analysis of *Bos taurus* and *Sus scrofa* X and Y chromosome transcriptome highlights reproductive driver genes. *Oncotarget*. 2017; 8:54416–54433. <https://doi.org/10.18632/oncotarget.17081>.
49. Goodwin ZA, de Guzman Strong C. Recent positive selection in genes of the mammalian epidermal differentiation complex locus. *Frontiers in genetics*. 2016; 7:227. <https://doi.org/10.3389/fgene.2016.00227>.
50. Nelsen SM, Christian JL. Site-specific cleavage of BMP4 by furin, PC6, and PC7. *J Biol Chem*. 2009; 284:27157–66. <https://doi.org/10.1074/jbc.M109.028506>.
51. Nohe A, Keating E, Knaus P, Petersen NO. Signal transduction of bone morphogenetic protein receptors. *Cellular signalling*. 2004; 16:291–9. <https://doi.org/10.1016/j.cellsig.2003.08.011>.
52. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. 2004; 32:1792–7. <https://doi.org/10.1093/nar/gkh340>.
53. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*. 2013; 30:2725–9. <https://doi.org/10.1093/molbev/mst197>.
54. Poon AF, Frost SD, Pond SL. Detecting signatures of selection from DNA sequences using Datamonkey. *Methods Mol Biol*. 2009; 537:163–83. [https://doi.org/10.1007/978-1-59745-251-9\\_8](https://doi.org/10.1007/978-1-59745-251-9_8).
55. Majewski J, Ott J. Amino acid substitutions in the human genome: evolutionary implications of single nucleotide polymorphisms. *Gene*. 2003; 305:167–73. [https://doi.org/10.1016/S0378-1119\(03\)00379-2](https://doi.org/10.1016/S0378-1119(03)00379-2).
56. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. Molecular phylogenetics and the origins of placental mammals. *Nature*. 2001; 409:614–8.
57. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007; 24:1586–91. <https://doi.org/10.1093/molbev/msm088>.
58. Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. 2000; 155:431–49.
59. Hardin C, Pogorelov TV, Luthey-Schulten Z. Ab initio protein structure prediction. *Curr Opin Struct Biol*. 2002; 12:176–81. [https://doi.org/10.1016/S0959-440X\(02\)00306-8](https://doi.org/10.1016/S0959-440X(02)00306-8).
60. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015; 12:7–8. <https://doi.org/10.1038/nmeth.3213>.
61. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:12–21. <https://doi.org/10.1107/S0907444909042073>.
62. Gopalakrishnan K, Sowmiya G, Sheik SS, Sekar K. Ramachandran plot on the web (2.0). *Protein Pept Lett*. 2007; 14:669–71. <https://doi.org/10.2174/092986607781483912>.
63. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25:1605–12. <https://doi.org/10.1002/jcc.20084>.
64. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*. 2003; 19:163–4. <https://doi.org/10.1093/bioinformatics/19.1.163>.
65. Areal H, Abrantes J, Esteves PJ. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol Biol*. 2011; 11:368. <https://doi.org/10.1186/1471-2148-11-368>.
66. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017; 45:D777–D83. <https://doi.org/10.1093/nar/gkw1121>.
67. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN, Gaunt TR. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 2013; 34:57–65. <https://doi.org/10.1002/humu.22225>.