Journal of Animal Research: v.13 n.06, p. 881-898. December 2023

DOI: 10.30954/2277-940X.06.2023.3

Identification of Novel Single Nucleotide Polymorphisms in *Myf6* Gene Associated with Myogenesis in Large White Yorkshire *vis-à-vis* Non-descript Pigs of Punjab

Phelgye Soni[1], Simrinder Singh Sodhi[1*], C.S. Mukhopadhyay[2], Gurvinder Kaur[1] and Neeraj Kashyap[2]

[1]Department of Animal Biotechnology, College of Animal Biotechnology; Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, INDIA

[2]Department of Bioinformatics, College of Animal Biotechnology; Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, INDIA

**Corresponding author: SS Sodhi; E-mail: sodhisimrinder@gmail.com*

Received: 15 Sept., 2023 **Revised:** 28 Oct., 2023 **Accepted:** 01 Nov., 2023

ABSTRACT

The study focused on the *Myf6* gene associated with myogenesis, investigating its genetic variations and evolutionary dynamics in Large White Yorkshire and Non-descript pigs from Punjab. This study focused on the *Myf6* gene, a crucial regulator of myogenesis, to explore its genetic variations, evolutionary dynamics, and functional significance across species. Nucleotide sequence analysis revealed novel single nucleotide polymorphisms (SNPs) and indels, with a transversion mutation identified in Non-descript pigs in the intronic region of intron 2, potentially influencing gene expression through intron-mediated regulation. These findings offer genomic selection potential for breeding pigs with desired meat traits. The identified *Myf6* variations could serve as markers for efficient breeding, benefiting pig production. Additionally, genetic characterization aids Non-descript pig conservation and management. This study informs *Myf6* evolution, breeding strategies, and indigenous breed preservation.

HIGHLIGHTS

- Focused to identify SNPs in Large White Yorkshire and non-descript pigs.
- Understanding evolutionary perspective of *Myf6* gene in divergent species.

Keywords: Large White Yorkshire (LWY), non-descript pigs, *Myf6*, SNPs, pigs and genetic variations

Pork is one of the most consumed meats in the world with its highest per capita consumption in the Southeast Asian nations such as such as Vietnam, Philippines, Myanmar, Indonesia and Thailand. In most of SE Asian countries, pig is the most important livestock species and pork is the most preferred meat and therefore the global pork industry is shifting their production capacity to the Southeast Asian nations. However, the per capita consumption of pork in India is negligible compared to that of the rest of the pork consuming South-east Asian nations. Nonetheless, India has sizeable pig population of 9.06 million according to the recent livestock census which is at par with most of the major pork producing South-east Asian nations except for Vietnam. This huge gap that exists between the supply and demand of pork in India due to fact that most of the pigs reared in India are of the nondescript local breeds.

The majority of the pig population in India is of indigenous breeds (76%) though, population of cross-bred and exotic pigs is on the rise (Riedel *et al*., 2012). Furthermore, these indigenous pigs are well-preferred by the consumers due to quality and taste of pork over the exotic and crossbred pigs in North-Eastern region of the country where almost

How to cite this article: Soni, P., Sodhi, S.S., Mukhopadhyay, C.S., Kaur, G. and Kashyap, N. (2023). Identification of Novel Single Nucleotide Polymorphisms in *Myf6* Gene Associated with Myogenesis in Large White Yorkshire vis-a-vis Non-descript Pigs of Punjab. *J. Anim. Res.*, **13**(06): 881-898.

Source of Support: None; **Conflict of Interest:** None

40% of the country's total pig population are reared in traditional small-holder subsistence-oriented production systems which serves as a dual conduit, delivering both nutritional and financial benefits, while concurrently supporting the transformation of household waste into valuable manure for cropping (Bhadauria *et al*., 2007). However, these native varieties have lower growth rates compared to exotic breeds such as the Large White Yorkshire (LWY) which make them less financially viable for a profitable pig industry (Kadirvel *et al*., 2023) which is necessary to seal the gap that exists between the supply and demand of pork in India. Consequently, there's a need to characterize and enhance the growth and muscle development of indigenous pigs to elevate their suitability for large-scale industrial farming, particularly considering the changing global climate scenario.

In livestock production, the growth and development of skeletal muscle are closely related to the growth rate of animals. Therefore, it is essential to identify the genes and genetic variations that are directly related to muscle growth phenotype in order to improve the rate of growth and development of skeletal muscle (Sodhi *et al*., 2021). The myogenic regulatory factors (MRFs) are a group of genes that play a critical role in skeletal muscle development (Wyszyńska-Koko and Kurył, 2004; Te Pas *et al*., 2007). Among these the myogenic factor 6 (*Myf6*) operates as a transcriptional activator during the intricate progression of myogenesis. *Myf6* plays a critical role in facilitating the transcription of core regions within promoters of genes specific to muscle development. The expression level of the *Myf6* gene surges during the postnatal phase, approximately tenfold higher than its counterparts in the MRF gene family. This heightened expression of *Myf6* contributes to the maintenance of the skeletal muscle phenotype, underscoring the gene's vital role in the regulation of post-natal development of the skeletal muscle (Wyszyńska-Koko and Kurył, 2004).

The porcine *Myf6* gene (Gene ID: 397005) is situated on chromosome 5 within the genome and has a length of 1,854 bp on NC_010447.5 reference sequence and is composed of three exons and two introns with a CDS length of 729 bp and protein length of 242 aa. In swine, the *Myf6* gene and its polymorphisms can putatively impact meat quality and quantity. Any mutations in *Myf6* exons or the regulatory regions in its introns could potentially affect the development of muscle fibres. The detection of any single nucleotide polymorphisms (SNPs) in the *Myf6* gene in the non-descript pig breeds of India would be of immense value in enhancing the quality and quantity of pork production in India by genomic selection. There is however only a very limited information on the SNPs in *Myf6* of the indigenous pigs of India, and far fewer are any information on the comparative analyses of SNPs of the *Myf6* gene in exotic *vis-à-vis* the non-descript pig breeds of India. Therefore, the present study is an attempt to identify novel SNPs within the *Myf6* gene which could potentially shed light to the disproportionately lower myogenic progression in the non-descript pigs of Punjab compared to that of the Large White Yorkshire. The study also investigated the evolutionary dynamics of the *Myf6* gene in divergent species to identify conservations and divergence in the gene.

MATERIALS AND METHODS

Collection of tissue samples

In this work, tissue samples of *Longissimus dorsi* muscle of each group of pigs (n=10) spanning between the 12th and 13th rib, of adult Large White Yorkshire and non-descript pigs from Punjab were collected. All the pigs were neutered and were of similar age (one year). Samples of adult Large White Yorkshire were collected from an organized farm Indo Canadian Swine Breeders at Village Kotli, district Ludhiana. The breed of animals was characterized, based on their phenotypic traits and pedigree. Samples of adult Non-descript pigs from Punjab were collected from slaughter house of Municipal Corporation, Ludhiana. The fresh tissue samples procured from the slaughterhouse promptly after slaughter and were transported to the laboratory in a sampling box with ice packs. After that the samples were washed with DEPC and stored at -80°C in sealed zip lock pouch bags.

Extraction and quantification of DNA

DNA extraction was conducted using the Phenol-Chloroform-Isoamyl-alcohol method (Pearson and Stirling, 2003). The concentration of DNA was determined by NanoDrop spectrophotometer. The ratio of optical density (OD) readings obtained at 260 nm and 280 nm ranged between 1.7 and 1.9, indicated that the DNA was of good quality and without contamination. The DNA

samples were dissolved in TE buffer and stored at-20°C. The quantified DNA was diluted based on its concentration for further PCR reaction.

PCR amplification *Myf6* gene by PCR

DNA fragment of 574bp of *Myf6* gene was successfully amplified with master-mix reaction (B.R. Biochem) and custom synthesized primers (Eurofins) (Table 2). The PCR was carried out using the conditions in (Table 4) for amplification of a region of *Myf6* gene. Gel runs were performed along with 50bp marker for all the samples visualized under gel doc (G:BoxSyngene).

Table 1: Quality and concentration of extracted DNA

Sl. No.	Breed			
	LWY		Non-descript pigs	
	Concentration (ng/µL)	Quality (A260/280)	Concentration (ng/µL)	Quality (A260/280)
1	922.9	1.85	1220.0	1.92
2	1205.0	1.73	742.9	1.88
3	1267.9	1.85	639.9	1.90
4	753.0	1.92	1170.2	1.81
5	793.9	1.90	1961.9	1.74
6	1207.2	1.88	857.0	1.79
7	1238.0	1.92	1953.2	1.86
8	1258.3	1.58	987.2	1.87
9	746.2	1.91	954.3	1.88
10	190.2	1.72	2630.1	1.91

Table 2: Primer sequence for *Myf6* gene

Sl. No.	Site	Primer sequence	Product Size (bp)	Reference
1	Exon2 and Exon3	F-5' CGTAACGTGACC TCATCTACAGG-3' R-5' GCTTTTGCCTAA TCTGCTGCC-3'	584	Self-designed

Table 3: PCR reaction mixture

PCR components	Final concentration	Volume single reaction
Master-mix 10x	1x	10 µL
1000pmol Forward Primer	5 pmol/µL	2 µL
1000pmol Reverse Primer	5 pmol/µL	2 µL
Template DNA	100ng/ µL	4 µL
DNase free water		4 µL
Total		**20 µL**

Table 4: Amplification condition for the PCR reaction

Steps	Temperature	Time	Cycles
Predenaturation	95°C	5min	1
Denaturation	95°C	30s	34
Annealing	57.5°C	30s	
Extension	72°C	1min	
Final Extension	72°C	5min	1

Gel extraction and sequencing of amplicon for identification of SNPs

The bands were excised under a UV illuminator and gel extraction was performed using the QIAquick Gel Extraction Kit the manufacturer's protocol. All 10 samples for each breed were sent for by Sanger Sequencing at GeneSpec Pvt Ltd. The sequencing results of chromatogram files were viewed by Finch TV software for identification of SNPs, and the FASTA sequences were compared to the reference sequences at NCBI database using BLAST (https://blast.ncbi.nlm.nih.gov/). Identification of SNP was done using multiple sequence alignment using MUSCLE and visualization with JalView. Electropherogram analysis for identification and confirmation of SNP. The cds region of the sequences were used for *in silico* prediction of encoded protein using ExPasy Translate and BlastX.

Table 5: Concentration of gel extracted DNA

Sl. No.	Breed	
	LWY	Non-descript pigs
	Concentration (ng/µL)	Concentration (ng/µL)
1	22.6	13.2
2	16.2	17.8
3	16.2	24.4
4	22.3	17.3
5	16.5	18.2
6	17.2	17.4
7	25.2	15.0
8	14.7	31.0
9	13.5	21.6
10	24.2	15.2

Evolutionary Analysis

In order to examine the evolutionary dynamics of the *Myf6* gene across distinct species, 18 diverse *Myf6* sequences were procured from the NCBI database. Clustal Omega

and MUSCLE were employed to perform multiple sequence alignment (MSA), followed by visualization using Jalview. The process also involved identifying an appropriate evolutionary model and constructing a phylogram using MEGA 11. The pairwise distance analysis and the assessment of selection pressure were also conducted using, MEGA 11.

RESULTS AND DISCUSSION

DNA fragment of 574bp of *Myf6* gene was successfully amplified with the primer pair that runs along with 50bp marker showed single, intense, compact bands of 574bp on 2% agarose gel (85V for 45 min.) for all the samples.

PCR amplicon of the *Myf6* gene was sequenced at GeneSpec Pvt Ltd. following Sanger's dideoxy chain termination method (Sanger *et al*., 1977). The sequences were aligned with the reference sequence and variations were verified using FinchTV. The partial *Myf6* sequence covering regions of exon2, intron2 and exon3 of Large White Yorkshire and Non-descript pigs of Punjab was submitted to NCBI through the BankIt portal (accession no: OR540727 and OR540728).

Large White Yorkshire Partial *Myf6* Sequence

>Large White Yorkshire *Myf6* partial sequence (466bp)

GCGCACCTGCAGCTCCCAGTGGCCAAGTGTT TCGGATCATTCCAGGGGGCTCGTGATGACTG CCAAGGAAGGTAAAGCAAAAGGGCACTGG GCCCCGCG GGAG ACCA TCCCGGGA GGGG GCTG GGTGCTCG CTGG GGGC AGAA CGCC CCGG GCGC CTGC GGGA GCCG GTCC CGCCCCGC GCGA GCCG AGAG CGGAGGCGCCCT CGAG CCGG GCGC TCCC GCCGCGCG CACC GGCGCCCG CGCC CAGC GGCA CCCT TCCG TGGG GTTA CGTT TGTG TGCG TGTG TGTGTGTG TTTT TCTGC GCCCAGGA GGGA CAAA CATT GATT CGTC AGCC TCGA GCAG CCTT CGGT GCCT TTCT TCCAT CGTG GACA GCAT TTCC TCGG AGGA ACAC ACGC TCCC CTGC GTCG AGGA AGTG GGGG AGAA ATAA CTCG CGGC CGGA GACG GTCTCCACGCAGCAGCAAAAGCCCA

Non-descript pig of Punjab Partial *Myf6* Sequence

>Non-descript pig of Punjab *Myf6* partial sequence (481bp)

TGCG CACC TGCA GCTC CCAG TGGC CAAGT GTTT CGGA TCATT CCAGG GGGCT CGTG ATGA CTGCC AAGGA AGGT AAAGC AAAA GGG CACT GGGC CCCG CAGG AGAC CATC CCGG GAGG GGGC TGGG TGCT CGCT GGGG GCAG AACG CCCC GGGC GCCT GCGG GAGC CGGT CCCG CCCC GCGC GAGCCG AGAG CGGAG GCGCC CTCGA GCCGG GCGCT CCCGC CGCGC GCACC GGCG CCCG CGCCCAGCGGCA CCGC GGGC GGCT CCCT TCCG TGGG GTTA TGTT TGTG TGCG TGTG TGTG TGTT TTTC TGCG CCCA GGAG GGAC AAAC ATTG ATTC GTCA GCCT CGAG CAGC CTTC GGTG CCTT TCTT CCAT CGTG GACA GCAT TTCC TCGG AGGA ACAC ACGC TCCC CTGCGTCG AGGA AGTG GGGG AGAAATAACTCG CGGC CGGA GACG GTCTCCACGCAGCAGCAAAAGCCCACCCT

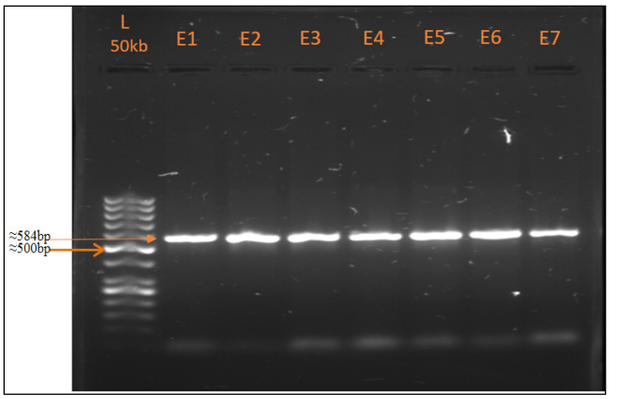


Fig. 1: Gel electrophoresis of PCR product of Large White Yorkshire pig samples amplified using the primer pair

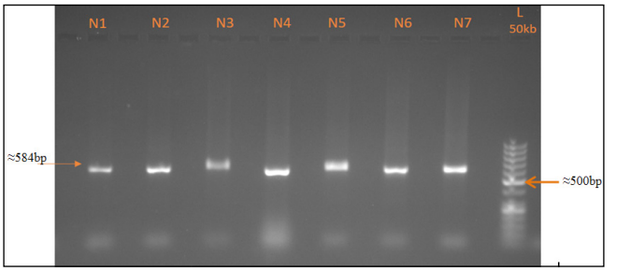


Fig. 2: Gel electrophoresis of PCR product of Non-descript pig samples amplified using the primer pair

The results for *Myf6* gene sequence in both Large White Yorkshire and Non-descript pig were compared with

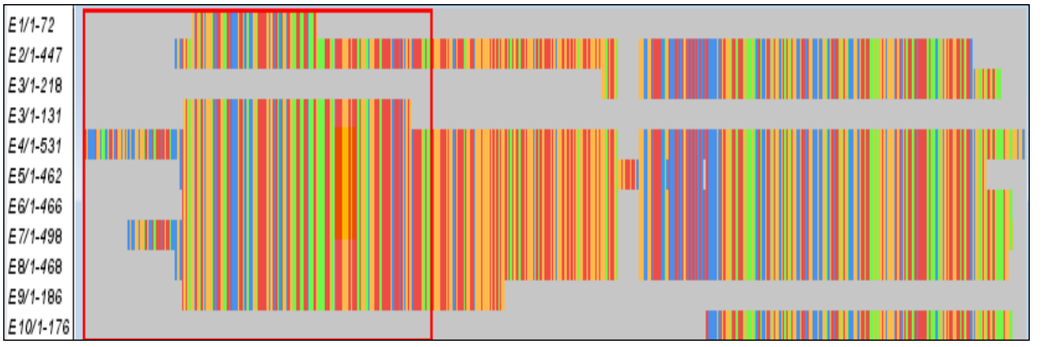


Fig. 3: Visualization of overview window of Large White Yorkshire sequences (10 sequences) using JalView

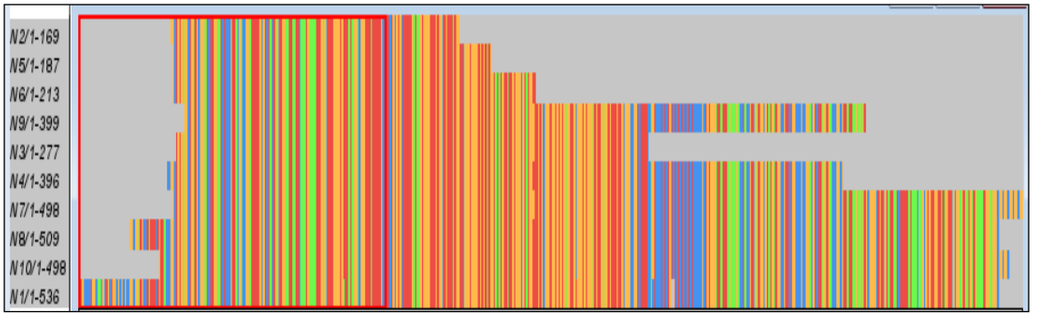


Fig. 4: Visualization using of overview window of Non-descript pig sequences (10 sequences) using Jalview

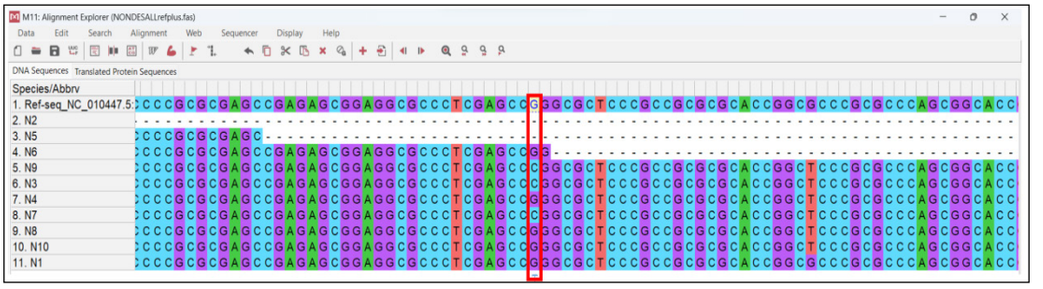


Fig. 5: Identification of mutations in Non-descript pig sequences with respect to the reference sequence

the reference sequence by Clustal Omega software and analysed with the help of Finch TV software. In Non-descript pig of Punjab in the intronic region of intron 2, a mutation was found at position 100763660 with reference to the reference sequence NC_010447.5:100762918-100764771 Sus scrofa isolate TJ Tabasco breed Duroc chromosome 5, Sscrofa 11.1, whole genome shotgun sequence (Fig. 5). This is a transversion change from G to C which was observed in three of the sample sequences which does not lead to a change in amino acid sequence as it is an intronic mutation. In the Large White Yorkshire pig no such mutations were observed in the sequence.

Indels were also detected in both pig populations. In Non-descript pigs, a single nucleotide insersion of nucleotide T at position between 100763566 and 100763567 with respect to NC_010447.5:100762918-100764771 Sus scrofa isolate TJ Tabasco breed Duroc chromosome 5, Sscrofa 11.1, whole genome shotgun sequence was

observed in the intron 2 region (N1, N4, N7, N8, N9, N10) (Table 7).

Similarly, (E2, E3, E4, E6, E7, E8) Large White Yorkshire pigs exhibited insertion of nucleotide T at position between 100763566 and 100763567 as well as a 12 nucleotide deletion CCGCGGGCGGCT/---------- between positions 100763566 and 100763567 with respect to NC_010447.5:100762918-100764771 Sus scrofa isolate TJ Tabasco breed Duroc chromosome 5, Sscrofa 11.1, whole genome shotgun sequence was observed in the intron 2 region (Table 6).

In silico prediction of the protein coded by the partial sequence of *Myf6* was done for both Large White Yorkshire and Non-descript pig sequences. The ExPasy Translate tool was employed to translate the coding sequences into their corresponding amino acid sequences for both Large White Yorkshire and Non-descript pig sequences. Following the translation, BlastX analysis was conducted to compare the translated protein sequences against a comprehensive protein database. Remarkably, the results of *in silico* analysis indicated that there were no substantial differences observed in the protein sequences derived from the coding regions for both Large White Yorkshire and Non-descript pig partial *Myf6* sequences. As the mutations identified in this study are present in the intronic region, these mutations in the analyzed sequences doesn't have a direct impact on the final protein product. As the mutations are present in the intronic region, it may have an effect through intron-mediated regulation of the gene expression. Further experimental validations and studies are warranted to corroborate these findings and explore the functional significance of intronic regions in a broader genomic context.

Determination of best model for evolutionary analysis

The 18 divergent *Myf6* nucleotide sequences were subjected to evolutionary analysis using MEGA11. The best evolutionary model based on factors like model accuracy, computational efficiency and compatibility are assessed for their model fit through statistical tests, such as Alkaline Information Criterion corrected (AICc) and Bayesian Information Criterion (BIC). These are statistical metrics used for model fitting and hypothesis testing

BIC and AICc values are calculated for various models and the model with the lowest value is considered the best fit to describe the substitution pattern. A lower BIC and AICc indicates better trade-off between goodness of fit and model complexity. The K2+G (Kimura 2-Parameter with gamma distribution) was the best evolutionary model for the *Myf6* gene as indicated in the Fig. 2.

Table 6: Indels in Large White Yorkshire Pig *Myf6* Sequence

Sl. No.	Sample	INDEL	POSITION wrt >NC_010447.5:100762918-100764771 Sus scrofa isolate TJ Tabasco breed Duroc chromosome 5, Sscrofa11.1, whole genome shotgun sequence	
1	E2	CCGCGGGCGGCT/----------	100763607 to 100763618	Deletion
2	E3	CCGCGGGCGGCT/----------	100763607 to 100763618	Deletion
3	E4	CCGCGGGCGGCT/----------	100763607 to 100763618	Deletion
4	E6	CCGCGGGCGGCT/----------	100763607 to 100763618	Deletion
5	E7	CCGCGGGCGGCT/----------	100763607 to 100763618	Deletion
6	E8	CCGCGGGCGGCT/----------	100763607 to 100763618	Deletion
7	E2	-/T	Between 100763566 and 100763567	Insertion
8	E4	-/T	Between 100763566 and 100763567	Insertion
9	E6	-/T	Between 100763566 and 100763567	Insertion
10	E7	-/T	Between 100763566 and 100763567	Insertion
11	E8	-/T	Between 100763566 and 100763567	Insertion

Table 7: Indels in Non-descript Pig *Myf6* Sequence

Sl. No.	Sample	INDEL	POSITION wrt >NC_010447.5:100762918-100764771 Sus scrofa isolate TJ Tabasco breed Duroc chromosome 5, Sscrofa 11.1, whole genome shotgun sequence	
1	N1	-/T	Between 100763566 and 100763567	Insertion
2	N4	-/T	Between 100763566 and 100763567	Insertion
3	N7	-/T	Between 100763566 and 100763567	Insertion
4	N8	-/T	Between 100763566 and 100763567	Insertion
5	N9	-/T	Between 100763566 and 100763567	Insertion
6	N10	-/T	Between 100763566 and 100763567	Insertion

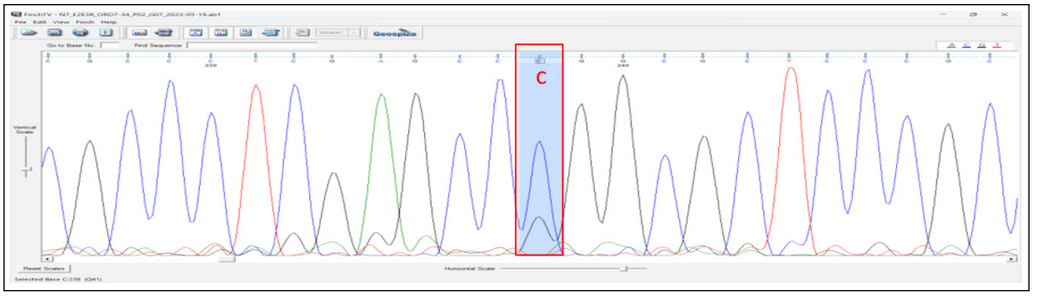


Fig. 6: Electropherogram of *Myf6* sequence of sample N7 depicting C allele viewed in FinchTV

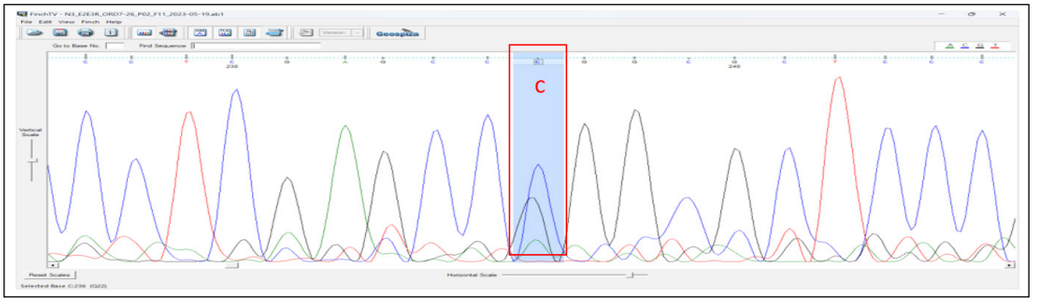


Fig. 7: Electropherogram of *Myf6* sequence of sample N3 depicting C allele viewed in FinchTV

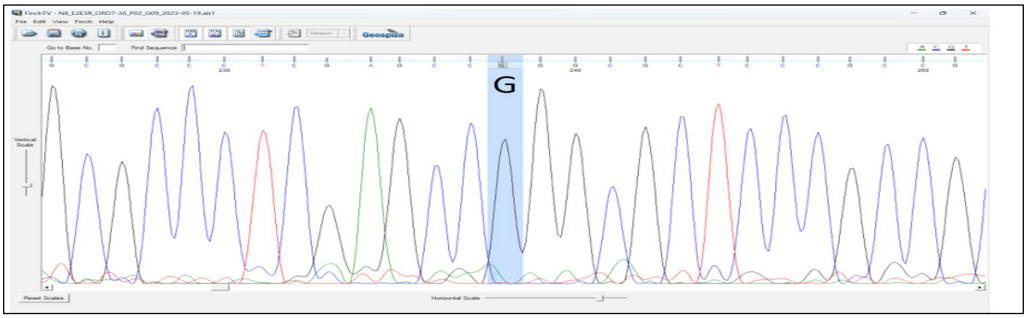


Fig. 8: Electropherogram of *Myf6* sequence of sample N8 depicting C allele viewed in FinchTV

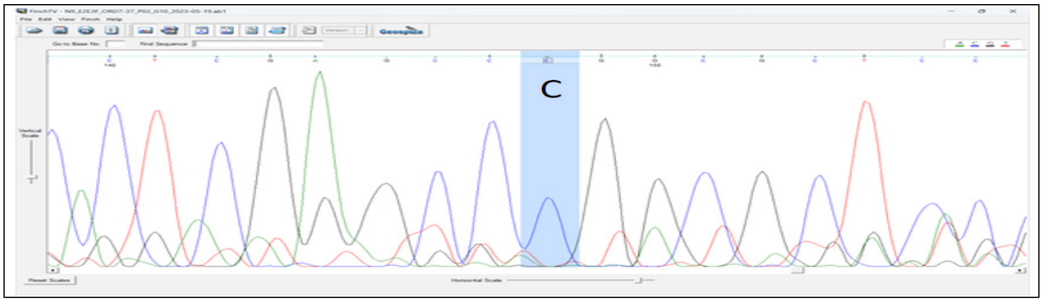


Fig. 9: Electropherogram of *Myf6* sequence of sample N9 depicting C allele viewed in FinchTV

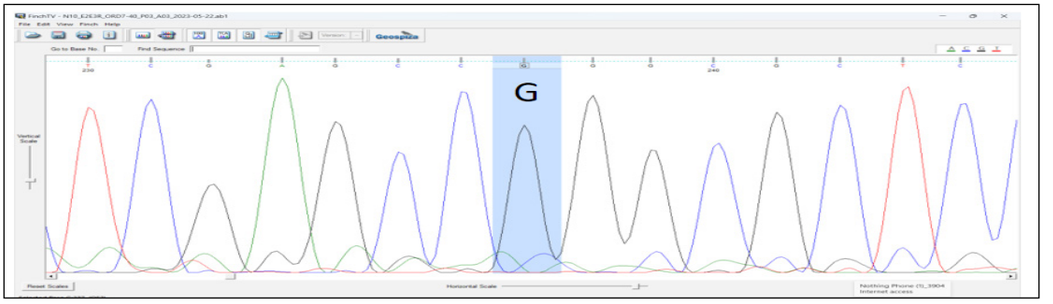


Fig. 10: Electropherogram of *Myf6* sequence of sample N10 depicting C allele viewed in FinchTV

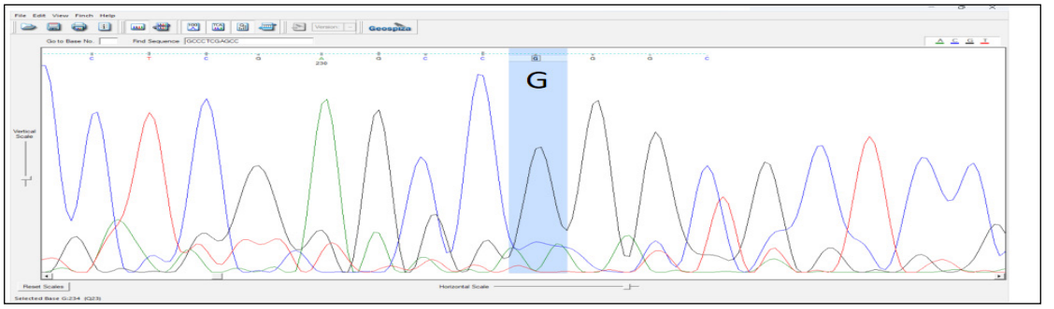


Fig. 11: Electropherogram of *Myf6* sequence of sample N6 depicting C allele viewed in FinchTV

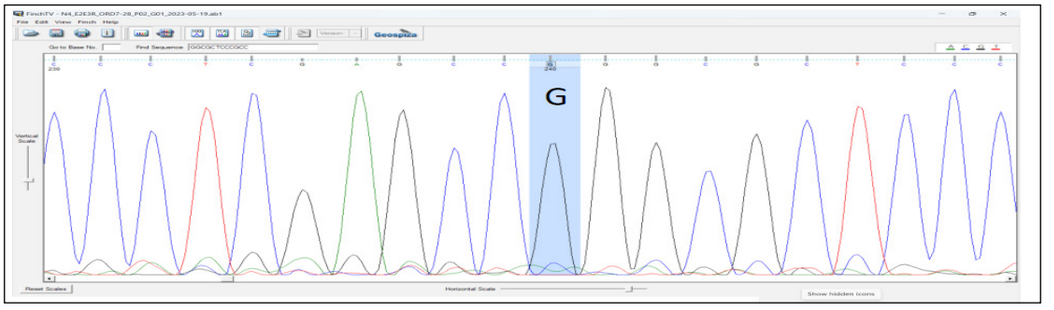


Fig. 12: Electropherogram of *Myf6* sequence of sample N4 depicting C allele viewed in FinchTV

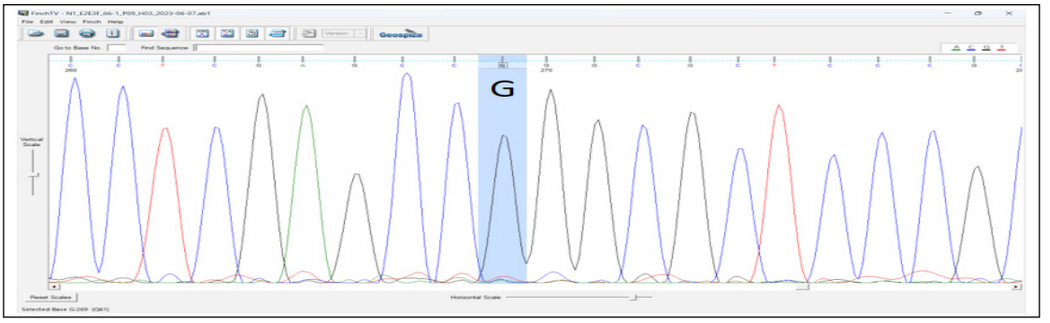


Fig. 13: Electropherogram of *Myf6* sequence of sample N1 depicting C allele viewed in FinchTV

The selected Kimura 2-Parameter model with gamma distribution (K2+G) emerged as the best evolutionary model for the *Myf6* gene. This model strikingly balances the goodness of fit with model complexity, thereby ensuring a robust representation of the genetic variations observed. The choice of the K2+G model serves as a pivotal foundation for subsequent evolutionary analyses, providing a reliable framework for exploring selection pressures, nucleotide substitutions, and genetic relationships within the *Myf6* gene.

Determining positive selection

The Z-test analysis of positive selection provides crucial insights into the evolutionary dynamics of the *Myf6* gene across diverse species. By examining the statistical significance of positive selection and identifying species with adaptive variations, this analysis sheds light on the molecular evolution of the *Myf6* gene and the potential functional implications of selective pressures.

The Z-test is a powerful tool to assess the presence of positive selection, which favors the fixation of advantageous mutations in a gene. The analysis compares nonsynonymous (dN) and synonymous (dS) substitution rates, enabling the identification of sites under selective pressure (Nielsen *et al*., 1998). A positive Z-score indicates that nonsynonymous substitutions are more prevalent than expected, hinting at potential adaptive changes. The probability of rejecting the null hypothesis of strict-neutrality (dN = dS) in favor of the alternative hypothesis (dN > dS) (below diagonal) is shown (Fig. 22). Values of P less than 0.05 are considered significant at the 5% level and are highlighted. The test statistic (dN - dS) is shown above the diagonal. dS and dN are the numbers of synonymous and nonsynonymous substitutions per site, respectively. The variance of the difference was computed using the analytical method. Analyses were conducted using the Nei-Gojobori method. This analysis involved 18 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 2709 positions in the final dataset. Evolutionary analyses were conducted in MEGA11. Z-score with absolute value greater than 1.96 are considered statistically significant at 5% level.

The Z-test analysis reveals intriguing patterns of positive selection across species. Notably, species pairs such as sheep-cattle, brown bear-zebra fish, and frog-brown bear exhibit statistically significant positive selection signatures. These findings suggest that specific sites within the *Myf6* gene have experienced adaptive changes, potentially linked to functional divergence or adaptation to ecological niches (Yang, 2007). Positive selection in the *Myf6* gene holds significant implications for muscle development and function. Adaptive changes in key regions of the gene may contribute to variations in muscle development strategies among species. For instance, species like brown bear and frog might have evolved unique muscle development mechanisms to suit their distinct habitats and lifestyles (Carroll, 2008).

The observed positive selection signatures reflect the delicate balance between conservation and adaptation in the *Myf6* gene. While certain regions remain conserved across species, adaptive changes in other regions enable organisms to thrive in diverse environments. These findings emphasize the dynamic nature of molecular evolution and the intricate interplay between genetic

Fig. 14: Translation of Large White Yorkshire pig partial cds sequence using ExPasy Translate

Fig. 15: Translation of Non-descript pig partial cds sequence using ExPasy Translate

stability and innovation (Hughes, 2008).

The positive selection signatures uncovered by the Z-test analysis provide evolutionary storylines for the *Myf6* gene. They highlight specific species pairs that have undergone adaptive changes in response to unique selective pressures. The identified sites of positive selection offer a starting point for further functional analyses to elucidate the precise role of these adaptations in muscle development and related physiological processes (Bustamante *et al*., 2005).

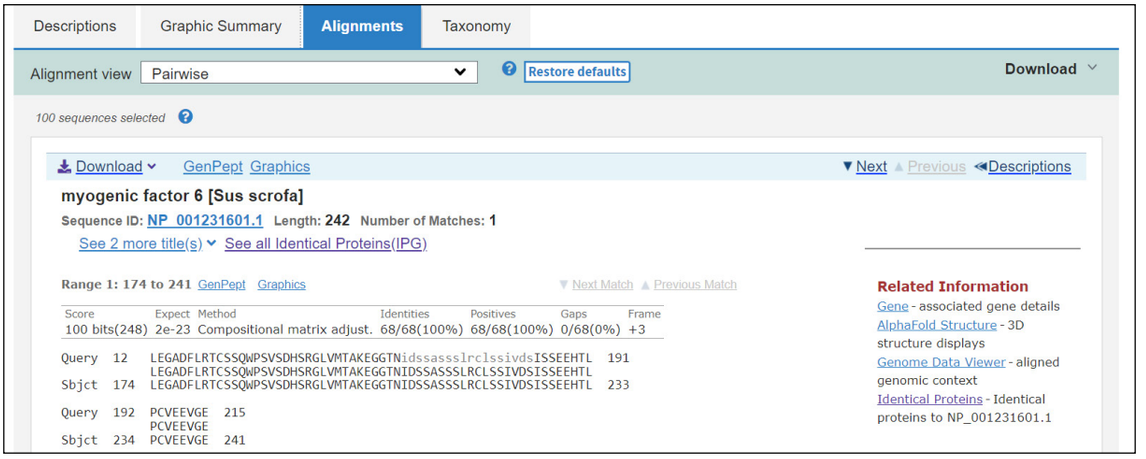


Fig. 16: BlastX result of Large White Yorkshire pig *Myf6* partial cds sequence

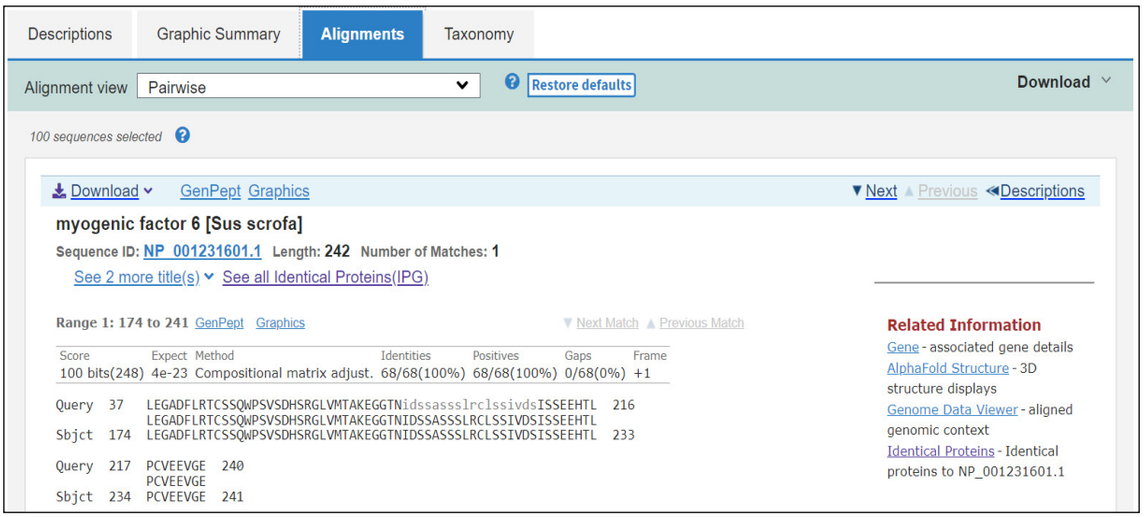


Fig. 17: BlastX result of Non-descript pig *Myf6* partial cds sequence

Fig. 18: Alignment of translated protein sequences *Myf6* of Large White Yorkshire pig and Non-descript pig

Determining negative selection

Purifying selection acts to preserve the functional integrity of genes by removing deleterious mutations. The Z-test measures the deviation of observed dN/dS ratios from the neutral expectation, enabling the identification of sites that are conserved due to their functional significance (Yang, 2007). Negative Z-scores reflect purifying selection, signifying the removal of nonsynonymous mutations. The probability of rejecting the null hypothesis of strict-neutrality (dN = dS) in favor of the alternative hypothesis (dN < dS) (below diagonal) is shown (Fig. 23). Values of P less than 0.05 are considered significant at the 5% level and are highlighted. The test statistic (dS - dN) is shown above the diagonal. dS and dN are the numbers of synonymous and nonsynonymous substitutions per site, respectively. The variance of the difference was computed using the analytical method. Analyses were conducted using the Nei-Gojobori method. This analysis involved 18 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 2709 positions in the final dataset. Evolutionary analyses were conducted in MEGA11. Z-score with absolute value greater than 1.96 are considered statistically significant at 5% level.

The Z-test analysis reveals key insights into the evolution of the *Myf6* gene across species. Notably, species pairs such as sheep-cattle, chimpanzee-human, and horse-cattle exhibit negative Z-scores, indicating purifying selection acting on specific regions. These conserved sites likely play critical roles in muscle development and related physiological processes across species (Kosiol *et al*., 2008).

The presence of purifying selection in the *Myf6* gene underscores its importance in muscle development and function. Conserved regions are likely essential for the proper folding and interaction of *Myf6* protein, contributing to the formation of functional muscle fibers. Mutations in these regions could lead to impaired muscle development and associated health issues (Cvijović *et al*., 2018).

The observed purifying selection signatures highlight the delicate balance between evolutionary conservation and adaptation. While certain regions of the *Myf6* gene are highly conserved across species, other regions may have undergone adaptive changes to suit specific ecological niches or physiological demands. This interplay between conservation and innovation shapes the evolutionary trajectory of the gene (Carroll, 2008).

The regions under purifying selection provide valuable insights into the functional domains of the *Myf6* gene. These conserved sites likely contribute to key protein interactions, DNA binding, or other essential functions. Further research into the specific roles of these conserved regions could illuminate the precise mechanisms by which *Myf6* contributes to muscle development (Mayrose *et al*., 2013).

Pair Wise Distance (PWD)

The PWD matrix serves as a quantitative tool for gauging the genetic disparities between species. By examining the values within the matrix, we gain insights into the varying degrees of genetic relatedness and divergence among the examined species. This analysis is pivotal in reconstructing evolutionary histories and discerning patterns of adaptation across different lineages (Nei and Kumar, 2000). The PWD matrix also informs phylogenetic relationships by quantifying genetic distances. The PWD values can be employed to construct phylogenetic trees, visually representing the evolutionary branching patterns among species. The branching patterns can be validated through statistical techniques, offering a robust framework for elucidating the evolutionary relationships among diverse organisms (Felsenstein, 1985).

Certain notable PWD values stand out, underlining the genetic uniqueness of specific species. For instance, the PWD value of 1.19355 between *Ursus arctos* (Brown bear) and *Danio rerio* (Zebrafish) is indicative of a substantial genetic divergence. This stark difference could be attributed to the disparate ecological niches occupied by these species – the aquatic habitat of zebrafish and the terrestrial realm of the brown bear – resulting in distinct evolutionary trajectories (Benton *et al*., 2015).

Conversely, species exhibiting lower PWD values suggest a closer genetic affinity. The relatively low PWD value of 0.00605 between *Homo sapiens* (Humans) and *Pan troglodytes* (Chimpanzee) underscores their shared evolutionary heritage. This proximity is consistent with our understanding of the genetic relatedness between humans and our primate counterparts, further reinforcing the concept of common ancestry (King *et al*., 1975).

Construction of Phylogenetic tree

The constructed phylogenetic tree, utilizing the maximum likelihood method, offers a window into the evolutionary connections between species based on their *Myf6* gene sequences. By assessing the branch lengths, we can glean insights into the genetic distances and evolutionary divergence among these diverse species, shedding light on the dynamics of the *Myf6* gene across the animal kingdom.

MEGA11 (Tamura *et al*., 2021) software was used for the construction of phylogenetic tree and estimation of evolutionary divergence. The nucleotide sequences in FASTA Format were subjected to analysis for phylogenetic tree construction using the maximum likelihood method. The best model for analysing the nucleotide sequence data selected was K2+G for divergent *Myf6* gene. The reliability of the branching of the tree was checked by 1000 bootstrap re-sampling and also calculated Pair wise distance (PWD). The utilization of the K2+G model, supported by 1000 bootstrap resampling and Pair Wise Distance (PWD) calculations, ensures the reliability of the tree's branching patterns.

Notably, the clustering observed among ruminants and primates highlights shared genetic features and evolutionary histories related to muscle development. The closer lineage of *Mus musculus* with primates implies potential functional similarities in *Myf6* gene regulation and myogenesis processes. However, the significant genetic distance exhibited by species like Piscean, avian, and amphibians underscores their distinct evolutionary trajectories. Particularly striking is the substantial divergence of the brown bear from other mammalians, suggesting unique evolutionary pressures that have shaped its *Myf6* gene.

***Danio rerio* (Zebrafish) and *Xenopus tropicalis* (Tropical clawed frog)**

The branch length between Danio rerio and Xenopus tropicalis is 0.567, indicating a moderate genetic distance. This suggests that these two species diverged relatively recently in evolutionary history. This observation is consistent with their placement in distinct orders, *Danio rerio* belonging to Cypriniformes and *Xenopus tropicalis* to Anura. Danio rerio is a teleost fish, while *Xenopus tropicalis* is an amphibian. These two species belong to different vertebrate classes, Actinopterygii and Amphibia, respectively. The genetic divergence underscores the evolutionary trajectory of fish and amphibian lineages.

***Gallus gallus* (Red junglefowl) and *Ursus arctos* (Brown bear)**

The branch length between *Gallus gallus* and Ursus arctos is 0.215 and 0.680, respectively. These values suggests a more substantial genetic divergence distance between these two species. This is expected as they belong to different classes and orders within the animal kingdom with *Gallus gallus* classified as Aves and *Ursus arctos* as Mammalia. The genetic separation underscores the vast evolutionary gap between avian and mammalian lineages.

***Pan troglodytes* (Chimpanzee) and *Homo sapiens* (Humans)**

The branch length between *Pan troglodytes* and *Homo sapiens* is very small (0.0014 and 0.0045, respectively). This exceptionally close genetic relationship is consistent with the well-established genetic affinity between chimpanzees and humans, reflecting their shared evolutionary ancestry. Both species are classified in the family Hominidae, and the minuscule genetic divergence aligns with their recent common evolutionary history. The close genetic affinity between chimpanzees and humans has been extensively studied and highlights the intricacies of primate evolution.

***Macaca mulatta* (Rhesus monkey) and *Canis lupus familiaris* (Dog)**

The branch length between *Macaca mulatta* and *Canis lupus familiaris* is relatively small (0.021 and 0.0098, respectively). This observation hints at a shared evolutionary heritage, potentially influenced by common selective pressures. *Macaca mulatta* falls within the Cercopithecidae family, while *Canis lupus familiaris* belongs to the Canidae family. Macaca mulatta is a primate, while *Canis lupus familiaris* is a domesticated mammal. This proximity in genetic divergence is likely due to shared mammalian ancestry.

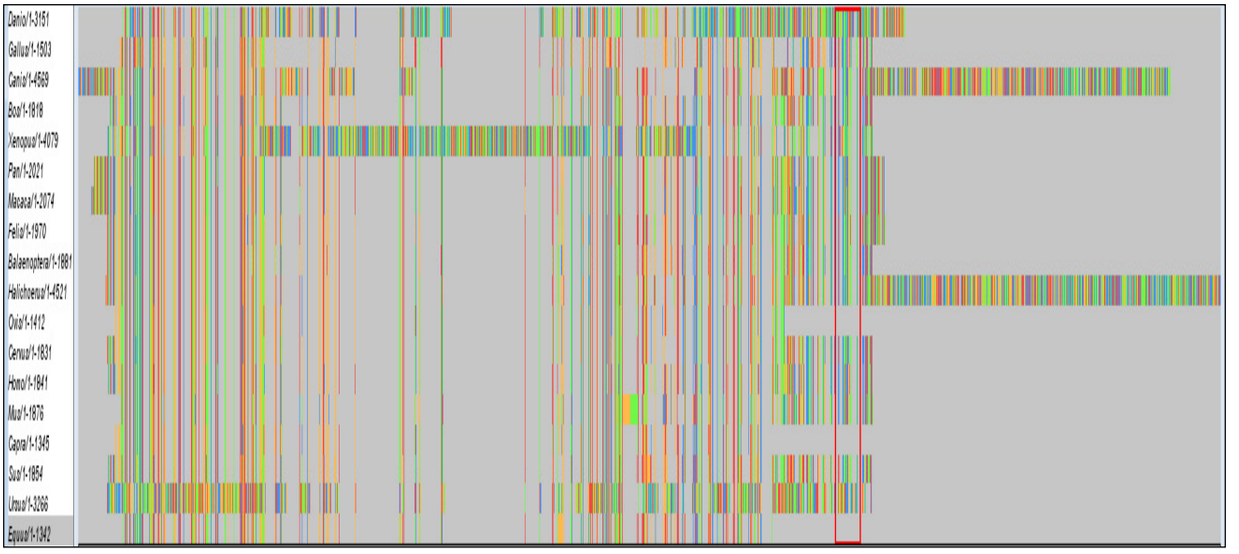
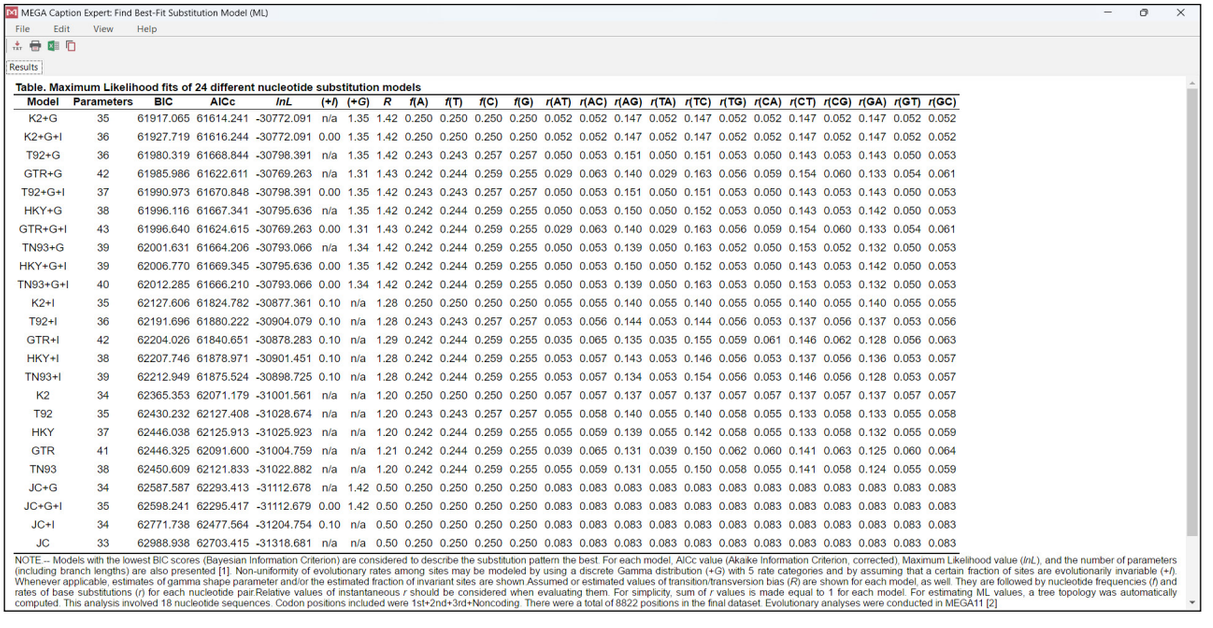


Fig. 19: An Overview of Multiple sequence alignment (using JALVIEW software) of divergent *Myf6* gene



MEGA Caption Expert: Find Best-Fit Substitution Model (ML)

File Edit View Help

Results

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

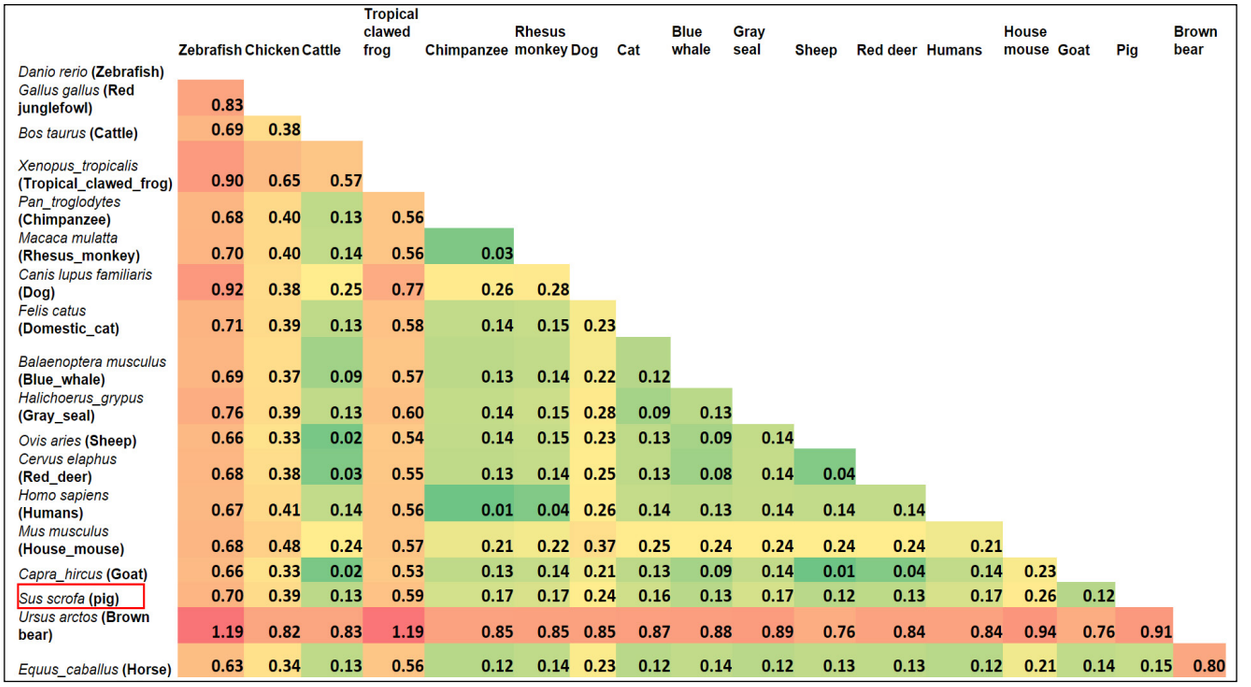
Model	Parameters	BIC	AICc	*lnL*	*(+I)*	*(+G)*	*R*	*f*(A)	*f*(T)	*f*(C)	*f*(G)	*r*(AT)	*r*(AC)	*r*(AG)	*r*(TA)	*r*(TC)	*r*(TG)	*r*(CA)	*r*(CT)	*r*(CG)	*r*(GA)	*r*(GT)	*r*(GC)
K2+G	35	61917.065	61614.241	-30772.091	n/a	1.35	1.42	0.250	0.250	0.250	0.250	0.052	0.052	0.147	0.052	0.147	0.052	0.052	0.147	0.052	0.147	0.052	0.052
K2+G+I	36	61927.719	61616.244	-30772.091	0.00	1.35	1.42	0.250	0.250	0.250	0.250	0.052	0.052	0.147	0.052	0.147	0.052	0.052	0.147	0.052	0.147	0.052	0.052
T92+G	36	61980.319	61668.844	-30798.391	n/a	1.35	1.42	0.243	0.243	0.257	0.257	0.050	0.053	0.151	0.050	0.151	0.053	0.050	0.143	0.053	0.143	0.050	0.053
GTR+G	42	61985.986	61622.611	-30769.263	n/a	1.31	1.43	0.242	0.244	0.259	0.255	0.029	0.063	0.140	0.029	0.163	0.056	0.059	0.154	0.060	0.133	0.054	0.061
T92+G+I	37	61990.973	61670.848	-30798.391	0.00	1.35	1.42	0.243	0.243	0.257	0.257	0.050	0.053	0.151	0.050	0.151	0.053	0.050	0.143	0.053	0.143	0.050	0.053
HKY+G	38	61996.116	61667.341	-30795.636	n/a	1.35	1.42	0.242	0.244	0.259	0.255	0.050	0.053	0.150	0.050	0.152	0.053	0.050	0.143	0.053	0.142	0.050	0.053
GTR+G+I	43	61996.640	61624.615	-30769.263	0.00	1.31	1.43	0.242	0.244	0.259	0.255	0.029	0.063	0.140	0.029	0.163	0.056	0.059	0.154	0.060	0.133	0.054	0.061
TN93+G	39	62001.631	61664.206	-30793.066	n/a	1.34	1.42	0.242	0.244	0.259	0.255	0.050	0.053	0.139	0.050	0.163	0.052	0.050	0.153	0.052	0.132	0.050	0.053
HKY+G+I	39	62006.770	61669.345	-30795.636	0.00	1.35	1.42	0.242	0.244	0.259	0.255	0.050	0.053	0.150	0.050	0.152	0.053	0.050	0.143	0.053	0.142	0.050	0.053
TN93+G+I	40	62012.285	61666.210	-30793.066	0.00	1.34	1.42	0.242	0.244	0.259	0.255	0.050	0.053	0.139	0.050	0.163	0.053	0.050	0.153	0.053	0.132	0.050	0.053
K2+I	35	62127.606	61824.782	-30877.361	0.10	n/a	1.28	0.250	0.250	0.250	0.250	0.055	0.055	0.140	0.055	0.140	0.055	0.055	0.140	0.055	0.140	0.055	0.055
T92+I	36	62191.696	61880.222	-30904.079	0.10	n/a	1.28	0.243	0.243	0.257	0.257	0.053	0.056	0.144	0.053	0.144	0.056	0.053	0.137	0.056	0.137	0.053	0.056
GTR+I	42	62204.026	61840.651	-30878.283	0.10	n/a	1.29	0.242	0.244	0.259	0.255	0.035	0.065	0.135	0.035	0.155	0.059	0.061	0.146	0.062	0.128	0.056	0.063
HKY+I	38	62207.746	61878.971	-30901.451	0.10	n/a	1.28	0.242	0.244	0.259	0.255	0.053	0.057	0.143	0.053	0.146	0.056	0.053	0.137	0.056	0.136	0.053	0.057
TN93+I	39	62212.949	61875.524	-30898.725	0.10	n/a	1.28	0.242	0.244	0.259	0.255	0.053	0.057	0.134	0.053	0.154	0.056	0.053	0.146	0.056	0.128	0.053	0.057
K2	34	62365.353	62071.179	-31001.561	n/a	n/a	1.20	0.250	0.250	0.250	0.250	0.057	0.057	0.137	0.057	0.137	0.057	0.057	0.137	0.057	0.137	0.057	0.057
T92	35	62430.232	62127.408	-31028.674	n/a	n/a	1.20	0.243	0.243	0.257	0.257	0.055	0.058	0.140	0.055	0.140	0.058	0.055	0.133	0.058	0.133	0.055	0.058
HKY	37	62446.038	62125.913	-31025.923	n/a	n/a	1.20	0.242	0.244	0.259	0.255	0.055	0.059	0.139	0.055	0.142	0.058	0.055	0.133	0.058	0.132	0.055	0.059
GTR	41	62446.325	62091.600	-31004.759	n/a	n/a	1.21	0.242	0.244	0.259	0.255	0.039	0.065	0.131	0.039	0.150	0.062	0.060	0.141	0.063	0.125	0.060	0.064
TN93	38	62450.609	62121.833	-31022.882	n/a	n/a	1.20	0.242	0.244	0.259	0.255	0.055	0.059	0.131	0.055	0.150	0.058	0.055	0.141	0.058	0.124	0.055	0.059
JC+G	34	62587.587	62293.413	-31112.678	n/a	1.42	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G+I	35	62598.241	62295.417	-31112.679	0.00	1.42	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	34	62771.738	62477.564	-31204.754	0.10	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC	33	62988.938	62703.415	-31318.681	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

NOTE -- Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (*lnL*), and the number of parameters (including branch lengths) are also presented [1]. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+*G*) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+*I*). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown Assumed or estimated values of transition/transversion bias (*R*) are shown for each model, as well. They are followed by nucleotide frequencies (*f*) and rates of base substitutions (*r*) for each nucleotide pair.Relative values of instantaneous *r* should be considered when evaluating them. For simplicity, sum of *r* values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. This analysis involved 18 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 8822 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2]

Fig. 20: Selection of best model for evolutionary analysis on the basis of *Myf6* gene in MEGA 11

Ovis aries (Sheep) and *Capra hircus* (Goat)

The branch length between *Ovis aries* and *Capra hircus* is 0.0026 and 0.0041, respectively. This indicates a relatively small genetic distance. Both species are classified under the subfamily Caprinae, within the Bovidae family and their genetic similarity may reflect a shared evolutionary path. The relatively recent common ancestry has likely contributed to the preservation of genetic resemblances. This points to their shared evolutionary history within the same taxonomic group.



	Zebrafish	Chicken	Cattle	Tropical clawed frog	Chimpanzee	Rhesus monkey	Dog	Cat	Blue whale	Gray seal	Sheep	Red deer	Humans	House mouse	Goat	Pig	Brown bear
Danio rerio **(Zebrafish)**																	
Gallus gallus **(Red junglefowl)**	0.83																
Bos taurus **(Cattle)**	0.69	0.38															
Xenopus_tropicalis **(Tropical_clawed_frog)**	0.90	0.65	0.57														
Pan_troglodytes **(Chimpanzee)**	0.68	0.40	0.13	0.56													
Macaca mulatta **(Rhesus_monkey)**	0.70	0.40	0.14	0.56	0.03												
Canis lupus familiaris **(Dog)**	0.92	0.38	0.25	0.77	0.26	0.28											
Felis catus **(Domestic_cat)**	0.71	0.39	0.13	0.58	0.14	0.15	0.23										
Balaenoptera musculus **(Blue_whale)**	0.69	0.37	0.09	0.57	0.13	0.14	0.22	0.12									
Halichoerus_grypus **(Gray_seal)**	0.76	0.39	0.13	0.60	0.14	0.15	0.28	0.09	0.13								
Ovis aries **(Sheep)**	0.66	0.33	0.02	0.54	0.14	0.15	0.23	0.13	0.09	0.14							
Cervus elaphus **(Red_deer)**	0.68	0.38	0.03	0.55	0.13	0.14	0.25	0.13	0.08	0.14	0.04						
Homo sapiens **(Humans)**	0.67	0.41	0.14	0.56	0.01	0.04	0.26	0.14	0.13	0.14	0.14	0.14					
Mus musculus **(House_mouse)**	0.68	0.48	0.24	0.57	0.21	0.22	0.37	0.25	0.24	0.24	0.24	0.24	0.21				
Capra_hircus **(Goat)**	0.66	0.33	0.02	0.53	0.13	0.14	0.21	0.13	0.09	0.14	0.01	0.04	0.14	0.23			
Sus scrofa **(pig)**	0.70	0.39	0.13	0.59	0.17	0.17	0.24	0.16	0.13	0.17	0.12	0.13	0.17	0.26	0.12		
Ursus arctos **(Brown bear)**	1.19	0.82	0.83	1.19	0.85	0.85	0.85	0.87	0.88	0.89	0.76	0.84	0.84	0.94	0.76	0.91	
Equus_caballus **(Horse)**	0.63	0.34	0.13	0.56	0.12	0.14	0.23	0.12	0.14	0.12	0.13	0.13	0.12	0.21	0.14	0.15	0.80

Fig. 21: Pair wise distance (PWD) of nucleotide sequences of divergent Myf6 gene MEGA 11 to find the distance between particular taxa

	Zebrafish	Chicken	Dog	Cattle	Tropical clawed frog	Chimpanzee	Rhesus monkey	Cat	Blue whale	Gray seal	Sheep	Red deer	Humans	House mouse	Goat	Pig	Brown bear	Horse
Danio rerio **(Zebrafish)**		-1.49	-0.15	0.66	-0.46	0.74	0.44	0.81	0.12	0.47	0.59	1.18	0.88	0.20	0.45	0.17	2.23	1.31
Gallus gallus **(Red junglefowl)**	1.00		0.43	-0.15	-0.10	-0.03	0.58	0.26	-0.30	-0.03	-0.18	-0.52	0.18	-0.71	0.07	-0.57	0.85	-0.10
Canis lupus familiaris **(Dog)**	1.00	0.33		-0.08	-0.09	0.80	0.53	0.21	-1.08	0.26	-0.51	-0.14	0.09	-1.34	-0.26	-1.37	0.64	-1.77
Bos Taurus **(Cattle)**	0.25	1.00	1.00		1.48	0.72	-0.05	0.17	1.03	-0.16	2.26	1.30	0.95	0.09	3.31	-0.58	0.89	-0.05
Xenopus tropicalis **(Tropical clawed frog)**	1.00	1.00	1.00	0.07		1.15	1.22	0.39	0.56	1.04	1.14	0.77	1.05	0.40	1.03	0.42	2.02	0.10
Pan troglodytes **(Chimpanzee)**	0.23	1.00	0.21	0.23	0.13		-2.00	0.53	0.79	0.65	-0.51	1.04	-0.33	0.07	-0.30	-0.58	1.58	-1.51
Macaca mulatta **(Rhesus monkey)**	0.33	0.28	0.30	1.00	0.11	1.00		0.14	0.54	-0.12	-0.71	0.30	-1.29	-0.66	-0.46	-0.80	1.35	-2.00
Felis catus **(Domestic cat)**	0.21	0.40	0.42	0.43	0.35	0.30	0.44		-0.31	-1.62	-0.54	0.48	0.53	-0.62	-0.66	-0.64	1.40	-1.98
Balaenoptera musculus **(Blue whale)**	0.45	1.00	1.00	0.15	0.29	0.22	0.30	1.00		-0.56	0.12	0.59	0.97	-0.03	0.23	-0.77	1.15	-1.14
Halichoerus grypus **(Gray seal)**	0.32	1.00	0.40	1.00	0.15	0.26	1.00	1.00	1.00		-0.44	0.01	0.06	-0.22	-0.23	-0.30	1.34	-1.87
Ovis aries **(Sheep)**	0.28	1.00	1.00	0.01	0.13	1.00	1.00	1.00	0.45	1.00		-0.15	-0.23	-0.53	1.11	-0.81	1.07	-0.80
Cervus elaphus **(Red deer)**	0.12	1.00	1.00	0.10	0.22	0.15	0.38	0.32	0.28	0.50	1.00		1.02	-0.10	0.48	-0.02	0.84	-0.53
Homo sapiens **(Humans)**	0.19	0.43	0.46	0.17	0.15	1.00	1.00	0.30	0.17	0.48	1.00	0.16		0.17	-0.12	-0.34	1.61	-1.39
Mus musculus **(House mouse)**	0.42	1.00	1.00	0.46	0.34	0.47	1.00	1.00	1.00	1.00	1.00	1.00	0.43		-0.10	-0.32	1.43	-0.74
Capra hircus **(Goat)**	0.33	0.47	1.00	0.00	0.15	1.00	1.00	1.00	0.41	1.00	0.14	0.32	1.00	1.00		-0.39	0.95	-0.38
Sus scrofa **(pig)**	0.43	1.00	1.00	1.00	0.34	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		0.59	-1.02
Ursus arctos **(Brown bear)**	0.01	0.20	0.26	0.19	0.02	0.06	0.09	0.08	0.13	0.09	0.14	0.20	0.06	0.08	0.17	0.28		0.35
Equus caballus **(Horse)**	0.10	1.00	1.00	1.00	0.46	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.37	

Fig. 22: Positive selection of model of nucleotide sequences of *Myf6* gene using MEGA 11

	Zebrafish	Chicken	Dog	Cattle	Tropical clawed frog	Chimpanzee	Rhesus monkey	Cat	Blue whale	Gray seal	Sheep	Red deer	Humans	House mouse	Goat	Pig	Brown bear	Horse
Danio rerio **(Zebrafish)**		1.49	0.15	-0.66	0.46	-0.74	-0.44	-0.81	-0.12	-0.47	-0.59	-1.18	-0.88	-0.20	-0.45	-0.17	-2.23	-1.31
Gallus gallus **(Red junglefowl)**	0.07		-0.43	0.15	0.10	0.03	-0.58	-0.26	0.30	0.03	0.18	0.52	-0.18	0.71	-0.07	0.57	-0.85	0.10
Canis lupus familiaris **(Dog)**	0.44	1.00		0.08	0.09	-0.80	-0.53	-0.21	1.08	-0.26	0.51	0.14	-0.09	1.34	0.26	1.37	-0.64	1.77
Bos Taurus **(Cattle)**	1.00	0.44	0.47		-1.48	-0.72	0.05	-0.17	-1.03	0.16	-2.26	-1.30	-0.95	-0.09	-3.31	0.58	-0.89	0.05
Xenopus tropicalis **(Tropical clawed frog)**	0.32	0.46	0.47	1.00		-1.15	-1.22	-0.39	-0.56	-1.04	-1.14	-0.77	-1.05	-0.40	-1.03	-0.42	-2.02	-0.10
Pan_troglodytes **(Chimpanzee)**	1.00	0.49	1.00	1.00	1.00		2.00	-0.53	-0.79	-0.65	0.51	-1.04	0.33	-0.07	0.30	0.58	-1.58	1.51
Macaca mulatta **(Rhesus monkey)**	1.00	1.00	1.00	0.48	1.00	0.02		-0.14	-0.54	0.12	0.71	-0.30	1.29	0.66	0.46	0.80	-1.35	2.00
Felis catus **(Domestic cat)**	1.00	1.00	1.00	1.00	1.00	1.00	1.00		0.31	1.62	0.54	-0.48	-0.53	0.62	0.66	0.64	-1.40	1.98
Balaenopteramusculus **(Blue_whale)**	1.00	0.38	0.14	1.00	1.00	1.00	1.00	0.38		0.56	-0.12	-0.59	-0.97	0.03	-0.23	0.77	-1.15	1.14
Halichoerus_grypus **(Gray_seal)**	1.00	0.49	1.00	0.44	1.00	1.00	0.45	0.05	0.29		0.44	-0.01	-0.06	0.22	0.23	0.30	-1.34	1.87
Ovis aries **(Sheep)**	1.00	0.43	0.30	1.00	1.00	0.31	0.24	0.29	1.00	0.33		0.15	0.23	0.53	-1.11	0.81	-1.07	0.80
Cervus elaphus **(Red_deer)**	1.00	0.30	0.44	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.44		-1.02	0.10	-0.48	0.02	-0.84	0.53
Homo sapiens **(Humans)**	1.00	1.00	1.00	1.00	1.00	0.37	0.10	1.00	1.00	1.00	0.41	1.00		-0.17	0.12	0.34	-1.61	1.39
Mus musculus **(House mouse)**	1.00	0.24	0.09	1.00	1.00	1.00	0.25	0.27	0.49	0.41	0.30	0.46	1.00		0.10	0.32	-1.43	0.74
Capra hircus **(Goat)**	1.00	1.00	0.40	1.00	1.00	0.38	0.32	0.26	1.00	0.41	1.00	1.00	0.45	0.46		0.39	-0.95	0.38
Sus scrofa **(pig)**	1.00	0.29	0.09	0.28	1.00	0.28	0.21	0.26	0.22	0.38	0.21	0.49	0.37	0.38	0.35		-0.59	1.02
Ursus arctos **(Brown bear)**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		-0.35
Equus caballus **(Horse)**	1.00	0.46	0.04	0.48	1.00	0.07	0.02	0.03	0.13	0.03	0.21	0.30	0.08	0.23	0.35	0.15	1.00	

Fig. 23: Purifying selection of model of nucleotide sequences of Myf6 gene using MEGA 11

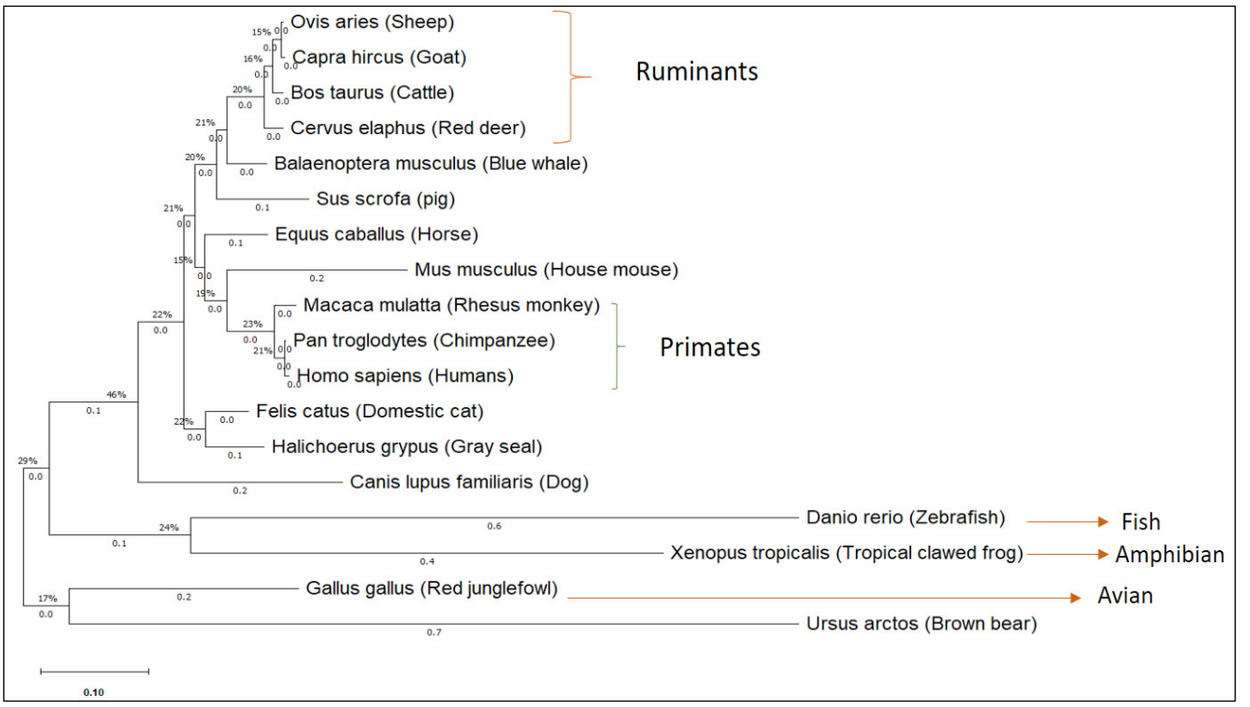


Fig. 24: Construction of Phylogenetic tree of divergent *Myf6* gene sequence using MEGA 11

Implications for the Evolutionary Dynamics of the *Myf6* Gene

The phylogenetic analysis of the *Myf6* gene sequences across diverse species provide insights in understanding of the gene's evolutionary history, functional conservation, and adaptive changes. The genetic distances inferred from the branch lengths offer a window into the intricate interplay of genetic conservation, adaptation, and selective pressures revealed by the branch lengths in the phylogenetic tree of the *Myf6* gene sequences provides a multifaceted perspective on the gene's evolution. While this analysis offers a compelling starting point, a comprehensive understanding would necessitate integrating additional data, such as functional studies, genomic context, and other molecular analyses. Researchers should consider these implications in a broader evolutionary and functional framework to unravel the nuanced complexities of the *Myf6* gene's evolutionary journey across species.

1. Functional Conservation and Divergence

The observation of short branch lengths between closely related species, such as *Pan troglodytes* (Chimpanzee) and *Homo sapiens* (Humans), underscores the notion of functional conservation of the *Myf6* gene. The small genetic divergence implies that the essential functional roles of *Myf6*, likely associated with muscle development and regulation, have been largely preserved in these lineages over evolutionary time. This conservation could suggest that alterations to *Myf6* function might carry significant fitness costs, prompting its retention.

2. Adaptive Evolution and Functional Innovation

Conversely, the longer branch lengths between more distantly related species, such as *Gallus gallus* (Red junglefowl) and *Ursus arctos* (Brown bear), suggest the potential for adaptive changes and functional innovations in the *Myf6* gene. The genetic divergence over extended evolutionary periods might reflect species-specific adaptations in muscle development, metabolism, or other related physiological processes. These adaptations could have arisen due to differences in ecological niches, lifestyles, or dietary preferences among these species.

3. Selective Pressures and Molecular Evolution

The variations in branch lengths could also be indicative of the diverse selective pressures acting on the *Myf6* gene across species. Species experiencing similar environmental challenges or ecological roles might exhibit comparable selective pressures, leading to similar rates of genetic divergence. Conversely, unique selective pressures could drive distinct evolutionary trajectories, resulting in differences in *Myf6* sequence conservation or divergence.

4. Insights into Human Evolution

The proximity of branch lengths between *Macaca mulatta* (Rhesus monkey) *and Canis lupus familiaris* (Dog) provides intriguing insights into the genetic relationships between primates and domesticated mammals. Understanding the *Myf6* gene's evolutionary dynamics in these species can shed light on the molecular basis of muscle development and related functions in both primates and other mammals, including humans.

5. Evolutionary Genomics and Comparative Studies

The *Myf6* gene serves as a valuable model for investigating broader questions in evolutionary genomics and comparative biology. By comparing gene sequences across diverse species, researchers can uncover conserved genetic elements, functional domains, and potential regulatory regions within the *Myf6* gene. Such insights could extend our understanding of the genetic basis of muscle development, which has implications not only for basic research but also for applied fields such as biomedicine and agriculture.

CONCLUSION

The investigation unveiled transversion mutations and variations within the Non-descript pig sequence, highlighting the dynamic genetic diversity within the population. These mutations could potentially impact phenotypic traits, contributing to the breed's genetic variability and adaptability.

The discovery of new single nucleotide polymorphisms (SNPs) within the Myf6 gene, in contrast to the reference sequence NC_010447.5, holds significant significance due to the gene's pivotal role in muscle cell differentiation and proliferation. These identified SNPs could potentially influence muscle development, thereby affecting pork quality and production. As such these findings are relevant

not only for understanding indigenous pig genetics but also for improving meat traits through selective breeding programs.

Variations in the *Myf6* gene across species emphasize its evolutionary importance. While the protein coding sequence remains mostly conserved, these variations suggest adaptations to environmental or physiological demands, underscoring the gene's essential role in muscle development across taxa. Genetic mutations, including SNPs and indels, might impact regulatory elements or gene expression, influencing muscle development pathways. In-depth investigations, including gene expression analyses and in vitro experiments, are needed to elucidate how these genetic variations precisely affect muscle function and development. Furthermore, exploring the functional implications of positive selection signatures could shed light on the adaptive significance of these genetic changes across species.

REFERENCES

Benton, M.J., Donoghue, P.C., Asher, R.J., Friedman, M., Near, T.J. and Vinther, J. 2015. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron.,* **18**(1): 1-106.

Bhadauria, P., Sharma, A., Verma, H.K., Singh, I. and Singh, R. 2019. Pig farming: Promising Agri-Business in Punjab. ICAR-ATARI, Ludhiana, Punjab, India, pp. 1-80.

Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Todd Hubisz, M., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., Civello, D., Adams, M.D., Cargill, M. and Clark, A.G. 2005. Natural selection on protein-coding genes in the human genome. *Nature*, **437**(7062): 1153-1157.

Carroll, S.B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, **134**(1): 25-36.

Cvijović, I., Good, B.H. and Desai, M.M. 2018. The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics*, **209**(4): 1235-1278.

Felsenstein, J. 1985. Confidence Limits On Phylogenies: AN approach using the bootstrap. *Evolution,* **39**(4): 783-791.

Kadirvel, G., Devi, Y.S., Naskar, S., Bujarbaruah, K. M., Khargariah, G., Banik, S. and Gonmei, C. 2023. Performance of crossbred pigs with indigenous and Hampshire inheritance under a smallholder production system in the Eastern Himalayan hill region. *Front. Genet.*, **14**: 1042554.

King, M.C. and Wilson, A.C. 1975. Evolution at two levels in humans and chimpanzees. *Science,* **188**(4184): 107-116.

Nielsen, R. and Yang, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**(3): 929-936.

Hughes, A.R., Inouye, B.D., Johnson, M.T., Underwood, N. and Vellend, M. 2008. Ecological consequences of genetic diversity. *Ecol. Lett.*, **11**(6): 609-623.

Kosiol, C., Vinar, T., da Fonseca, R.R., Hubisz, M.J., Bustamante, C.D., Nielsen, R. and Siepel, A. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet.*, **4**(8): e1000144.

Mayrose, I., Stern, A., Burdelova, E.O., Sabo, Y., Laham-Karam, N., Zamostiano, R., Bacharach, E. and Pupko, T. 2013. Synonymous site conservation in the HIV-1 genome. *BMC Evol Biol.,* **13**: 164.

Nei, M. and Kumar, S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York.

Riedel, S., Schiborra, A., Huelsebusch, C., Huanming, M. and Schlecht, E. 2012. Opportunities and challenges for smallholder pig production systems in a mountainous region of Xishuangbanna, Yunnan Province, China. *Trop. Anim. Health Prod.,* **44**(8): 1971-1980.

Pearson, H. and Stirling, D. 2003. DNA Extraction from Tissue. In: PCR Protocols. Methods in Molecular Biology™, Bartlett, J.M.S., Stirling, D., 2nd Ed., Humana Press, Totowa, NJ, pp. 33-34.

Sodhi, S.S., Sharma, N., Ghosh, M., Sethi, R.S., Jeong, D. K. and Lee, S.J. 2021. Differential expression patterns of myogenic regulatory factors in the postnatal longissimus dorsi muscle of Jeju Native Pig and Berkshire breeds along with their co-expression with Pax7. *Electron J. Biotechnol.*, **51**: 8-16.

Tamura, K., Stecher, G. and Kumar, S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.*, **38**(7): 3022-3027.

Te Pas, M.F., Hulsegge, I., Coster, A., Pool, M.H., Heuven, H.H. and Janss, L.L. 2007. Biochemical pathways analysis of microarray results: regulation of myogenesis in pigs. *BMC Dev. Biol.*, **7**: 66.

Wyszyńska-Koko, J. and Kurył, J. 2004. Porcine MYF6 gene: sequence, homology analysis, and variation in the promoter region. *Anim. Biotechnol.*, **15**(2): 159-173.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.,* **24**(8): 1586-1591.