Research Article

Identification of Genomic Signatures Underlying Human-Specific Skeletal Morphology During Primates Evolution

Syrut Nasir[1]*, Freeha Shahbaz[2], Zuha Ehsan[1], Tahira Shaheen Butt[1], Uswa Meenal[2] and Isra Arshad[2]

[1]Department of Biotechnology, University of Sialkot, Sialkot, Pakistan; [2]Department of Biochemistry and Molecular biology, University of Gujrat, Gujrat, Pakistan.

Abstract | Human evolution is driven by mutation and natural selection, yet the genomic basis of human-specific skeletal traits remains obscure. The primary objective was to investigate the genetic variations in the Duchenne Muscular Dystrophy (*DMD)* gene and their impact on protein function. The *DMD* gene, located on the X chromosome, is responsible for producing dystrophin, a key protein that stabilizes muscles, especially in the skeletal muscles. Its role in skeletal morphology is significant as it helps maintain muscle strength, preventing muscle degeneration and atrophy, which is crucial for movement and posture. By supporting the muscles needed for walking and upright posture, the *DMD* gene plays an essential role in human bipedalism, enabling efficient and stable locomotion. A combination of genomic sequencing and bioinformatics analysis were employed that includes genetic sequence comparisons, phylogenetic analyses, domain annotation, and protein structural evolution analysis to identify regions of accelerated evolution related to human skeletal morphology. The study conducted a comparative analysis of the *DMD* gene in humans and non-human primates, revealing human-specific amino acid replacements. The Ka/Ks ratio of human *DMD* is 1.5, indicating positive selection, suggesting adaptive evolution. Regions of the human and chimpanzee *DMD* gene involved in positive selection were identified using SWAKK software. Selective pressures were further supported by Tajima's D and Fu and Li's tests, which showed significant deviations from neutral evolution. Tajima's D test yielded a negative value of -1.71219 with a p-value less than 0.05, while Fu and Li's test produced a negative value of -1.95457, reinforcing the presence of selective pressures on *DMD*. Domain characterization across vertebrates highlighted key functional domains of *DMD*, providing insights into their evolutionary relationships. Structural comparisons using Chimera yielded an RMSD value of 2.916 angstroms and an SDM value of 169.437, indicating significant structural differences between the human and ancestor *DMD* proteins. These differences suggest divergent evolutionary paths and potential functional modifications in these proteins. The accelerated evolution of the *DMD* gene in humans may have led to alterations in the functional capabilities of the *DMD* protein.

Received | May 11, 2026; **Accepted** | June 16, 2026; **Published** | June 26, 2026
***Correspondence** | Syrut Nasir, Department of Biotechnology, University of Sialkot, Sialkot, Pakistan; **Email:** syrut.nasir@uskt.edu.pk
Citation | Nasir, S., F. Shahbaz, Z. Ehsan, T.S. Butt, U. Meenal and I. Arshad. 2026. Identification of genomic signatures underlying human-specific skeletal morphology during primates evolution. *Biologia (Lahore),* 72(1): 14-28.
DOI | https://dx.doi.org/10.17582/journal.biologia/2026/72.1.14.28
Keywords | Evolution, Mutation, Selection, DMD gene, Skeletal, Morphology, Genomics, Phylogenetic

Introduction

The evolution of human skeleton over millions of years has provided useful information about skeletal diversity, its origin, and the changes brought about by environment and other factors of evolution (Eckhardt, 1989; Kralick and Zemel, 2020; Kun *et al*., 2023). The research conducted dives into the evolution and genetic architecture of human skeleton, accentuating the intricate play of different evolutionary forces responsible for shaping the traits of the human skeleton as well as diversifying it (Kun *et al*., 2023). Additionally, the evidence from fossils indicates vital morphological modifications in the size of the torso, limbs, and body width, denoting that these transitions happened gradually over time, with gradual evolution over the span of 6 to 7 million years. Humans are the sole bipedal great ape's characteristic of their distinct skeletal structure, which holds an upright posture. Bipedalism is facilitated by distinct anatomy of skeletal system, like short arms as compared to legs, a narrow trunk and pelvic region, and orientation nature of the vertebral column (Kun *et al*., 2023).

All these modifications to structural proportions started to emerge around the separation of the human and chimpanzee heritage and could possibly have promoted the usage of tools as well as altered the higher brain functions and cognitive evolution (Kun *et al*., 2023). Modern-day human skeleton is gracile described in relation to previously found hominin fossils and great ape remains. Evolutionary biologists classify these influences into neutral and non-neutral forces. Neutral processes implicate mutations that evoke unique diversity without straightforward benefits or impediments to individuals, adding up to random shifts within populations. On the other hand, non-neutral processes happen when mutations impact the individual's fitness, affecting their capacity to adapt to environmental factors (Fournier-Level *et al*., 2011).

Genomic studies on human-specific skeletal traits

The genomic study of human-specific skeletal traits has made substantial advancements in recent years. A thorough scientific analysis uncovered that distinctive genetic loci concerned with human skeletal proportions were focused within areas of genome that have gone through evolutionary changes in humans, as well as in regulatory components of genes that display differential patterns between humans and our closest evolutionary lineages, the great apes (Kun *et al*., 2023). After studying the genes concerned with evolution of human skeletal morphology, a candidate gene has been specified. The *DMD* gene, seems to have exhibited some point mutations for positive selection. Duchenne muscular dystrophy (*DMD*) is a hereditary disease distinguished by progressive muscle deficiency due to mutations in the *DMD* gene, which encodes the protein dystrophin (Farrokhi, 2021). The *DMD* gene displays an elevated rate of unique mutations, contributing to the preponderance of the disorder across all populations (Fischer et al., 2002). This fascinating discovery indicates that hereditary modifications that happened in these pivotal genomic regions may have played a climactic function in navigating the development of the distinct skeletal morphology found in human lineage (Buckler *et al*., 2009).

Materials and Methods

Sequence data collection across primates

Based on the comprehensive literature review, the genes *DMD*, *ZMIZ1*, and *FN1* have been identified to human skeletal development. *ZMIZ1* (Zinc Finger MIZ-Type Containing 1) is a crucial transcriptional co-activator that plays a pivotal role in regulating gene expression. Fibronectin, encoded by the *FN1* gene, is a crucial glycoprotein component of the extracellular matrix (ECM) that plays a vital role in various cellular processes essential for bone development and repair. After investigating these genes for human-specific mutations, it was discovered that *ZMIZ1* did not exhibit any human-specific non-synonymous mutations. Consequently, *DMD* gene has been determined as a candidate for analyzing positive selection in human skeletal morphology. The *GTEx* portal (https://gtexportal.org/home/) authorizes the exploration of gene expression patterns, study of genetic variants, knowledge of tissue-specificity, analysis of diseases, collection of data from numerous references, and visualization of the results (Stanfill and Cao, 2021).

Selection of orthologs sequences

To further scrutinize the evolutionary relationships of the *DMD* gene, the orthologs protein sequences of human *DMD* and its paralogs (*UTRN* and *SPTB*) were retrieved from *Ensembl* utilizing the *BLASTP* (https://blast.ncbi.nlm.nih.gov/Blast.cgi) sequence alignment tool (Pervaiz and Abbasi, 2016).

Species that were employed in the study are *Homo sapiens* (humans), *Macaca mulatta* (macaque), *Pongo abelii* (orangutan), *Pan troglodytes* (chimpanzee), *Pan paniscus* (bonobo), and *Saimiri boliviensis* (squirrel monkey), *Rattus norvegicus* (rat), *Mus musculus* (mouse) Gorilla gorilla (gorilla), *Equus caballus* (horse), *Ornithorhynchus anatinus* (platypus), *Gallus gallus* (chiken), *Loxdonta africana* (elephant), *Dasypus novemcinctus* (armadillo), *Sarcophilus hairrisii* (tasmanian devil), *Felis Catus* (cat), *Xenopustropicalis* (Tropical clawed frog), *Gasterosteus aculeatus* (stickleback), and *Denio rerio* (zebra fish).

Node sequences

After creating the phylogenetic tree, which illustrates the branching pattern and transformation of species over time, downloading the hereditary sequences from the internal nodes of the tree is the next step. This is accomplished by employing the ExtAncSeqMEGA file, which collects hereditary information for these ancestral origins (https://sourceforge.net/projects/extancseqmega/). The commands used in research are as follows Figure 1:

```
C:\Users\ALAZIZ\Desktop\New folder>ExtAncSeqMEGA.exe dmdancestors.txt
Extracting node Node 8
Extracting node Node 9
Extracting node Node 10
Extracting node Node 11
Extracting node Node 12

C:\Users\ALAZIZ\Desktop\New folder>
```

Figure 1: *Commands used in command prompt for extracting nodes.*

The last step in the procedure was to align the sequences of the modern human and non-human primate species with the downloaded node sequences. The human Duchenne muscular dystrophy (*DMD*) gene has two closely affiliated putative paralogs, as discerned by the paralog prediction tool in the *Ensembl* database (http://www.ensembl. Org). These are the *SPTB* (spectrin beta chain) and *UTRN* (utrophin) genes (Hubbard *et al*., 2002).

Comparative genomic analysis across primates

A comparative analysis was executed for the human *DMD* gene across different primate species, like chimpanzee, gorilla, macaque, marmoset, orangutan, and bonobo. The *MEGA* software (https://www.megasoftware.net/) was used to execute this comparative analysis with the prior goal of determining any human-specific mutations in this gene and then computing the Ka/Ks ratio, which is a criterion used to evaluate whether the gene has gone through positive selection during the process of evolution over time. By comparing the sequence of this gene across multiple primate species, the analysis aimed to reveal the molecular mechanisms and evolutionary influences that have led to the hereditary multiplicity and transformation of these organisms. In Figure 2 and 3, the interface of MEGA 11 software represented human-specific mutations in coding and protein sequences.

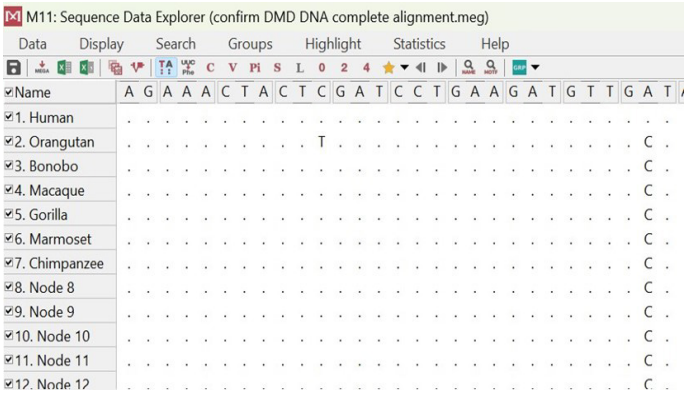

Figure 2: *Represented the human-specific mutations in coding sequences.*

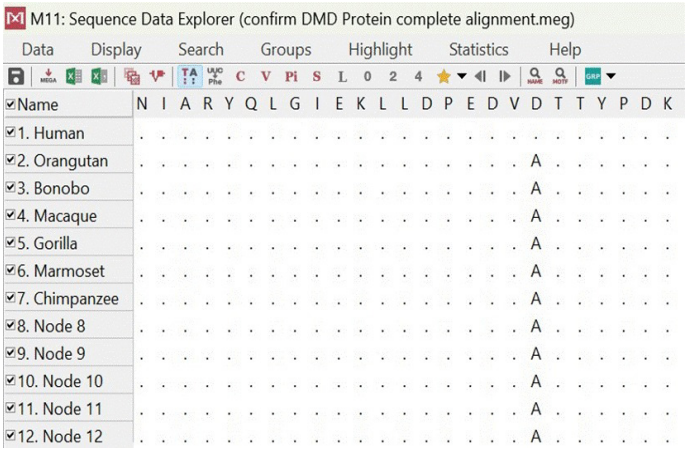

Figure 3: *Represented the human-specific mutations in protein sequences.*

Calculation of Ka/Ks ratio

The synonymous (Ks) and non-synonymous (Ka) mutations were calculated to discern the Ka/Ks ratio, which is a valuable measurement for assessing the evolutionary stresses acting on a gene. The Ka/Ks ratio, where Ka is the rate of non-synonymous mutations (those that alter the amino acid sequence) and Ks is the rate of synonymous mutations (those that do not influence the amino acid sequence), assists in determining whether the gene has undergone positive selection.

Synonymous and non-synonymous mutation identification

*MEGA*11 was used to align the coding sequences of *DMD* gene using the *CLUSTALW* algorithm. This process allowed the sequences to be compared and their similarities and disparities to be specified and identify the synonymous and non-synonymous mutations.

Sliding window analysis of Ka/Ks ratio

Sliding window analysis of the Ka/Ks ratio (*SWAKK*) was employed to determine regions under positive selection in the coding sequences of the gene commonly found in humans and chimpanzees. In *SWAKK* server (https://ibl.mdanderson.Org/swakk/1dkaks.html), the coding sequences of gene of interest into overlapping windows of 10 codons (30 nucleotides), with each window shifted by 10 codons were inserted. To achieve a balance between localized evolutionary changes and statistical reliability in Ka/Ks estimation a window size of 10 codons was selected on SWAKK server. It helps to reduce the excessive noise and identify region-specific selection signals precisely.

The estimation of the Ka (non-synonymous substitution rate) and Ks (synonymous substitution rate) values for each window across a gene define the deviations in the selective stress influencing the gene. A Ka/Ks ratio greater than 1 indicates the presence of positive selection, which demonstrates that the gene has experienced adaptive transformations. On the contrary, a Ka/Ks ratio less than or equal to 1 suggests either neutral or purifying selection.

Plot data preparation

In order to yield the desired plot, the analysis used the Ka and Ks values that were previously acquired through the *SWAKK* (Sliding Window Analysis of Ka/Ks) method. This ensured that the data was structured in a symmetric way that could be easily plotted and depicted visually, facilitating additional analysis and interpretation of the Ka/Ks proportions across the various window positions.

Visualizing data with GNUPLOT

The structured and categorized data were then fed into the GNUPLOT software, which is a strong open-source data visualization tool (https://gnuplot.en.lo4d.com/windows). The precise commands and code that were used to create and depict this graphical illustration are as follows in Figure 4:

```
# Set the style of the plot
set style data lines

# Set axis labels
set xlabel "Codon Number"
set ylabel "Ka-Ks"

# Set plot title
set title "Ka-Ks Values with Sliding Window (30 nucleotides)"

# Set the sliding window size
window_size = 10

# Load data from file and apply sliding window
plot 'rttn.txt' using 1:2 smooth csplines every window size with lines title 'Ka-Ks'
```

Figure 4: *Representing commands used to create GNUPLOT.*

Selection scans

Statistical trials: Tajima's D, Fu and Li's Test

The neutrality tests, such as Tajima's D and Fu and Li's D, were used to evaluate the genetic variability patterns within coding sequences of gene across human populations and their primate counterparts. These analyses were conducted using the DnaSP6 software (http://www.ub.edu/dnasp/).

Determination of statistical significance

Statistical importance was demonstrated through the calculation of p-values. Conventionally, a p-value lower than 0.05 is considered statistically substantial, symbolizing significant variations from the neutral model's expectations. Such divergences indicate the impact of factors beyond genetic drift, such as natural selection or demographic shifts, in shaping observed genetic deviation across human populations and primates (Wang *et al.*, 2023). This understanding carries substantial significance for explaining genetic variation, susceptibility to diseases, and the evolutionary dynamics describing the genes.

Phylogenetic analysis

The evolutionary relationship among the target gene, as well as their putative paralogs and orthologs, were specified using phylogenetic tree that employed the Maximum Likelihood (ML) method (Yang and Evolution, 2007). ML tree was yielded based on the Whelan and Goldman (WAG) model, which is a well-established statistical prototype for emulating amino acid substitutions throughout evolutionary cycles.

Comparative protein domain annotation

A thorough study of relevant academic literature was

done to determine distinct domains present within the gene of interest. Subsequent to this extensive literature review, the SMART (Simple Modular Architecture Research Tool) bioinformatics resource (http://smart.embl-heidelberg.de/) was employed to execute a more thorough analysis and depiction of these specified domains (Schultz *et al.*, 1998).

Protein structural evolution analysis

Structure prediction using PHYRE2

A computational tool *PHYRE2* was utilized to predict the structure of target gene shared among humans, chimpanzees, and their common ancestor. *PHYRE2* employs a technique called homology modeling, which involves predicting the 3D structure of a *DMD* domain protein. The C-terminal domain of *DMD* gene was selected due to presence of mutation. The sequences of selected domains were then uploaded in PHYRE2 server. *PHYRE2* provide a more comprehensive understanding of the structural characteristics of the target gene shared among the studied species.

Visualization and analysis with UCSF chimera

The predicted structures were then analyzed and compared by utilizing Chimera, a technological software program developed for visualizing and investigating complicated biomolecular structures (https://www.cgl.ucsf.edu/chimera/)(Sgro, 2017). The comparative study was done using chimera's matchmaker tool.

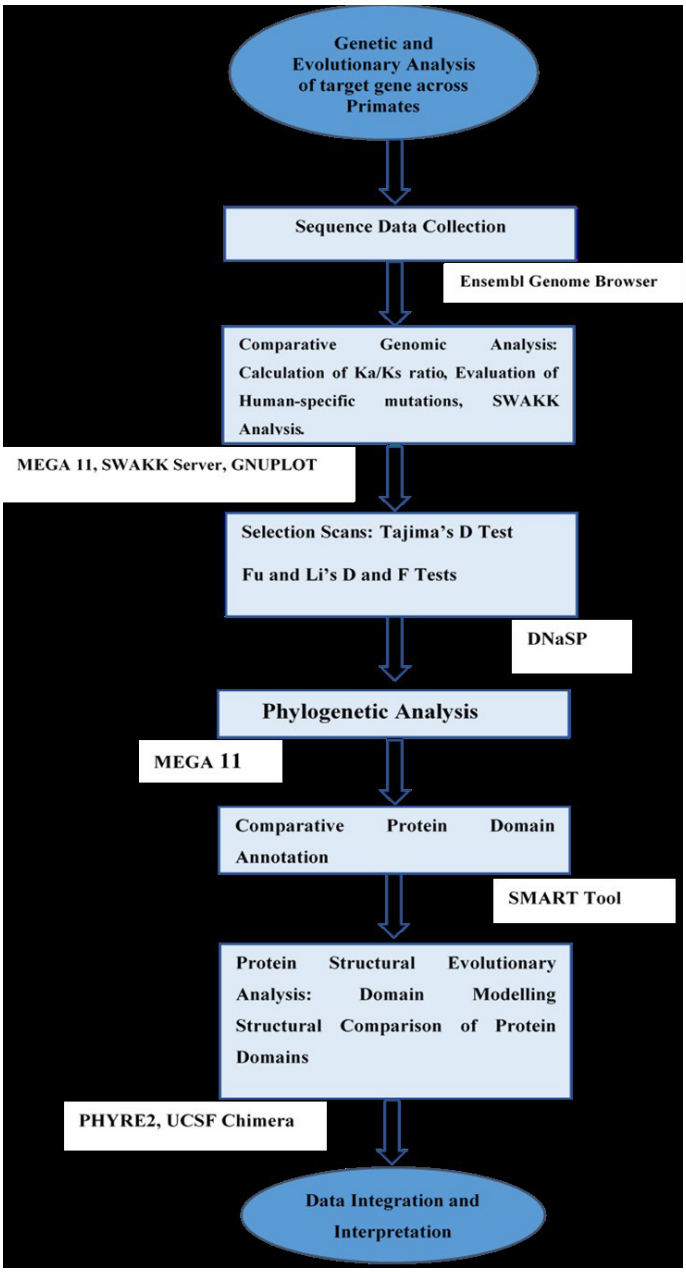

Figure 5: *Flowchart providing an overview of the methodology.*

Table 1: *Utilization of online resources for current study.*

Resources	URLs	Last accessed
Ensembl	https://www.ensembl.org/index.html	25-05-2024
Uniprot	https://www.uniprot.org/	02-06-2024
GTEx	https://gtexportal.org/home/	10-05-2024
CLUSTALW or MSA	https://megasoftware.net/dload_win_beta	15-05-2024
SWAKK	https://ibl.mdanderson.org/swakk/1dkaks.html	20-05-2024
GNUPlot	http://www.gnuplot.info/	18-04-2024
DnaSP	http://www.ub.edu/dnasp/	28-05-2024
SMART	http://smart.embl.de	04-06-2024
I-TASSER	https://zhanggroup.org/I-TASSER/	22-05-2024
PHYRE2	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	14-06-2024
UCSF Chimera	https://www.cgl.ucsf.edu/chimera/download.html	16-06-2024

Results and Discussion

Identification and analysis of species unique mutations and positive selection across primates

The study of Ka/Ks ratios in primates provides insights into how gene has evolved under the forces of natural selection. The marmoset, macaque, orangutan, gorilla, and bonobo heritage, as well as their common ancestral nodes, show lower Ka/Ks ratios as compared to humans. This suggests that these primate species have experienced fewer adaptive modifications at the protein level comparative to humans. The distinctive Ka/Ks ratios for the human and non-human primates, as well as their common ancestral nodes, are provided in Figure 6. Furthermore, Table 2 illustrates the places within the genome where these non-synonymous mutations have occurred, proposing additional insights into the evolutionary trajectories of these species.

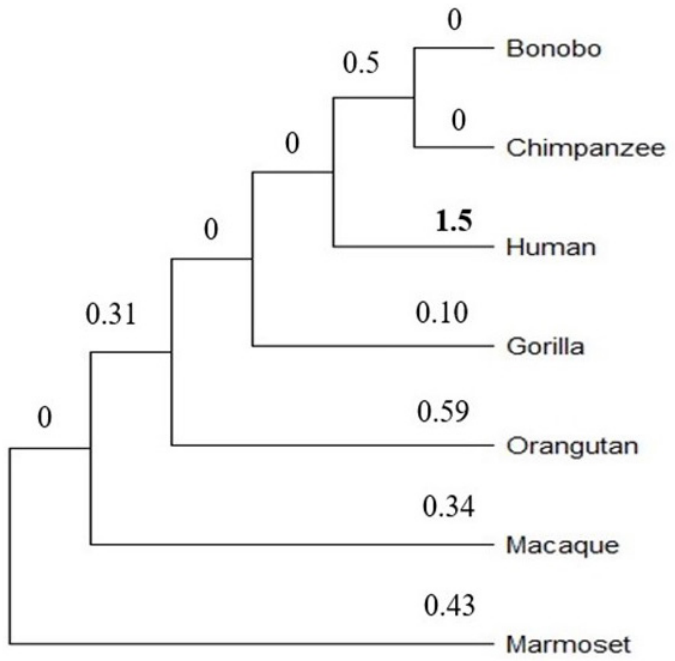

Figure 6: *Molecular evolution of DMD gene across primates.*

SWAKK assessment of human and chimpanzee genome

This relative analysis of the coding sequences between humans and chimpanzees, enabled by the *SWAKK* server, bared patterns of conservation and deviation, eventually shedding light on the complex evolutionary affinity between these two primates.

Regions exhibiting positive selection

After executing a thorough analysis of the hereditary sequences, many distinctive regions were specified that

Table 2: *Depicts non-synonymous substitutions and ka/ks ratios for human and non-human primates.*

Species	Amino acid replacement	Positions	TM	No of Ka	No of Ks	Ka/ks
Human	A → D	219	20	12	8	1.5
	E → D	378				
	N → Y	998				
	N → S	1097				
	S → L	2013				
	N → K	2243				
	I → V	2366				
	R → Q	2492				
	I → V	2614				
	I → T	2636				
	G → R	2672				
	Q → R	2938				
Bonobo	None	None	7	0	7	0
Marmoset	V → I	25	190	58	132	0.43
	C → S	188				
	F → Y	266				
	Y → F	302				
	T → A	309				
	S → A	320				
	Q → R	402				
	Q → K	639				
	T → A	673				
	T → A	865				
	R → K	929				
	M → T	943				
	A → T	951				
	T → A	961				
	D → E	972				
	I → V	1015				
	N → S	1049				
	L → I	1079				
	Q → K	1114				
	K → T	1196				
	V → A	1232				
	V → I	1360				
	K → R	1429				
	M → T	1484				
	M → T	1593				
	D → N	1609				
	I → V	1913				
	T → M	1982				
	M → T	1984				
	P → S	2022				

Table continues on next page.....

Species	Amino acid replacement	Positions	TM	No of Ka	No of Ks	Ka/ks
	D → N	2031				
	S → I	2049				
	I → V	2055				
	T → K	2068				
	R → H	2109				
	T → I	2177				
	L → P	2239				
	K → R	2322				
	Q → L	2326				
	D → A	2333				
	L → F	2334				
	E → D	2336				
	Q → H	2398				
	R → H	2632				
	H → Q	2675				
	R → H	2721				
	V → I	2752				
	K → R	2762				
	Q → R	2807				
	T → A	2933				
	I → V	2986				
	K → R	2990				
	E → D	2991				
	N → S	2992				
	S → G	2994				
	A → S	3049				
Chimpanzee	None	None	7	0	7	0
Macaque	E → K	477	90	23	67	0.34
	A → V	582				
	K →R	585				
	T → M	641				
	A → V	787				
	K → R	803				
	Q → E	885				
	A → T	897				
	V → M	926				
	A → V	985				
	G → T	1028				
	V → A	1038				
	A → V	1148				
	A → T	1395				
	P → S	1711				
	L → I	2082				
	Q → R	2298				
	L → F	2390				

Table continues on next column.....

Species	Amino acid replacement	Positions	TM	No of Ka	No of Ks	Ka/ks
	A → T	2413				
	D → G	2491				
	R → Q	2681				
	I → V	2763				
	A → S	3224				
Gorilla	R → C	489	21	2	19	0.10
	Q → R	594				
Orangutan	I → V	228	43	16	27	0.59
	S → T	330				
	R → Q	396				
	N → T	426				
	Q → E	503				
	V → L	516				
	K → Q	535				
	T → K	603				
	T → A	631				
	I → M	1116				
	S → L	1706				
	S → G	2180				
	N → H	2190				
	N → T	2300				
	D → E	2323				
	H → R	2389				

exhibited a Ka/Ks ratio greater than 1. The marmoset, macaque, orangutan, gorilla, and bonobo heritage, as well as their common ancestral nodes, show lower Ka/Ks ratios as compared to humans. This suggests that these primate species have experienced fewer adaptive modifications at the protein level comparative to humans. Studies like those on HR, HAR1, and RUNX2, among others, highlight how specific genes underwent positive selection to promote human-specific traits, offering deeper understanding into the molecular mechanisms underlying human evolution. These pivotal genes have been determined as playing significant roles in the transition of the human skeletal structure, especially the evolution of bipedalism and the S-shaped backbone including Runt genes (*Runx1, Runx2, Runx3*), these genes express during cartilage construction and are critical for skeleton-genesis in vertebrates. Runt Genes, specifically *Runx2* is involved in BMP signaling pathway. BMPs bind to receptors, phosphorylating SMAD1/5/8. The SMAD complex activates *RUNX2*, promoting osteoblast differentiation and bone formation. Genes in human accelerated regions

(*HARs*), and loci of skeletal ratios were enormously enhanced in *HARs*, which are components of the genome that have developed rapidly in humans in relation to other species. Similar to these genes, the regions exhibiting positive selection depicted the possibility of changes in skeletal development that permit bipedalism.

Regions exhibiting purifying selection

In contrast, other genomic domains were also determined that displayed a Ka/Ks ratio less than 1. This pattern is characteristic of purifying selection, a procedure in which deleterious mutations are efficiently extracted from the population, conserving the integrity and functionality of vital genetic components.

GNUPLOT visualization of genomic evolutionary dynamics

The study utilized the GNUPLOT software, which is integrated into the *SWAKK* server, to visually depict the selective pressures exerted across various genomic regions. The resulting graphical depiction, as shown in Figure 7, delivers an explicit and thorough knowledge of the selective forces at play.

The distinct substitutions, where the encoded amino acid was transformed, are complicated and illustrated in Table 3.

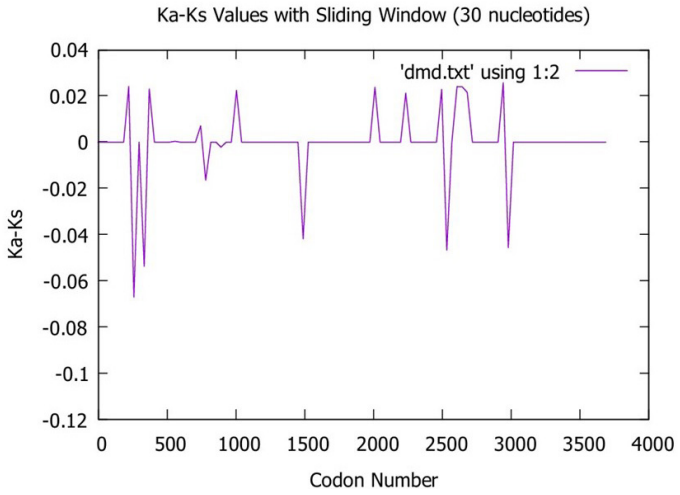

Figure 7: *Illustrates the specific regions of the DMD gene that exhibit evidence of undergoing positive selection.*

The Table 3 depicts that since the variation of the human origin from chimpanzees, there have been 13 specific amino acid substitutes in proteins that are distinctive to humans. These amino acid alternates signify modifications that have arisen at the genetic level, which may have contributed to the evolution of specific functional or structural traits in humans compared to chimpanzees.

Table 3: *Represents ancestral amino acid residues and substitutions in human and chimpanzee linages.*

Regions	Position	Ancestral residue	Replacement in chimpanzee	Replacement in human
Region 1	219	A		D
Region 2	343	R	G	
	373	E	D	
	377	E		D
Region 3	882	G		D
Region 4	997	N		Y
	1034	S	C	
	1096	N		S
	1129	T	K	
Region 5	2012	S		L
Region 6	2161	T	A	
	2242	N		K
Region 7	2365	I		V
Region 8	2491	R		Q
	2613	I		V
	2635	I		T
	2671	G		R
Region 9	2937	Q		R

It suggests that there have been five amino acid alterations that have ensued particularly within the chimpanzee lineage. This implies that both human and chimpanzee genomes have gone through distinct evolutionary modifications since their split, resulting in the observed disparities between the two species.

Evaluating genetic diversity and neutrality utilizing Tajima's D and Fu and Li's D and F tests

Analysis of Tajima's Test for the DMD Gene

The Tajima's Test analysis operated on the *DMD* gene provides insights into the selective pressures affecting this genetic locus. The outcomes suggest that the *DMD* gene is not going through neutral selection, implying that the observed genetic mutations are probably impacted on by other evolutionary means. The analysis was conducted on an alignment file comprising coding sequences, which covered a region of 14,014 areas.

This conclusion is illustrated in the study's Figure 8, providing a pictorial manifestation of Tajima's D result and its importance for the evolutionary dynamics of the *DMD* gene.

Tajima's Test

Input Data File: C:\...\dmdalignment.meg
Number of sequences: 7 Number of sequences used: 7
Selected region: 1-14014 Number of sites: 14014
Total number of sites (excluding sites with gaps / missing data): 1594

Number of polymorphic (segregating) sites, S: 816
Total number of mutations, Eta: 847

Average number of nucleotide differences, k: 245.28571
Nucleotide diversity, Pi: 0.15388
Theta (per sequence) from Eta: 345.71429
Theta (per site) from Eta: 0.21688

Tajima's D: -1.71219 Statistical significance: *, P < 0.05

========== Synonymous and NonSynonymous Substitutions ==========
No coding region defined
Calculated using the total number of mutations

Figure 8: *Evaluation of genetic variation and deviation from neutrality using Tajima's D.*

Analysis of Fu and Li's test for the DMD gene

The outcomes of the Fu and Li's tests for the *DMD* gene are illustrated in Figure 9 These statistical tests provide helpful insights into the evolutionary patterns of the *DMD* gene, demonstrating explicit evidence that it has not evolved under neutral selection. Several key findings support this conclusion:

Fu and Li's Tests

Input Data File: C:\...\dmdalignment.meg
Number of sequences: 7 Number of sequences used: 7
Selected region: 1-14014 Number of sites: 14014
Total number of sites (excluding sites with gaps / missing data): 1594

==== Analysis using the total number of mutations ====
Number of polymorphic (segregating) sites, S: 816
Total number of mutations, Eta: 847
Total number of singleton mutations, Eta(s): 822

Average number of pairwise nucleotide differences, k: 245.286
Nucleotide diversity, Pi: 0.15388

Theta estimated from k: 245.286
Theta estimated from Eta(s): 704.571
Theta estimated from Eta: 345.714
Number of Haplotypes, h: 7
Haplotype (gene) diversity, Hd: 1.000
Variance of Haplotype diversity: 0.00583

Fu's Fs statistic: 2.424
Strobeck's S statistic: 1.000
(Probability that NHap <= 7)
Probability that [NHap = 7]: 0.919

==== Analysis using only biallelic positions ====
Number of segregating sites analyzed, S: 785
Fu and Li's D* test statistic, FLD*: -1.78499
Fu and Li's F* test statistic, FLF*: -1.77998
Achaz Y* test statistic, AY*: -0.93165

=== Results from DnaSP v5 ===
Fu and Li's D* test statistic: -1.77525
Statistical significance: **, P < 0.02
Fu and Li's F* test statistic: -1.95457
Statistical significance: **, P < 0.02

Figure 9: *Deviation from neutrality using Fu and Li's.*

Significant negative Fu and Li's D and F values

The Fu and Li's D* and F* statistics, which are employed to witness divergences from neutral evolution, exhibit highly substantial negative values of -1.77525 and -1.95457, respectively, with p-values smaller than 0.02. These negative values suggest an excess of rare genetic variants in the *DMD* gene sequence described in relation to what would be anticipated under a neutral evolutionary model.

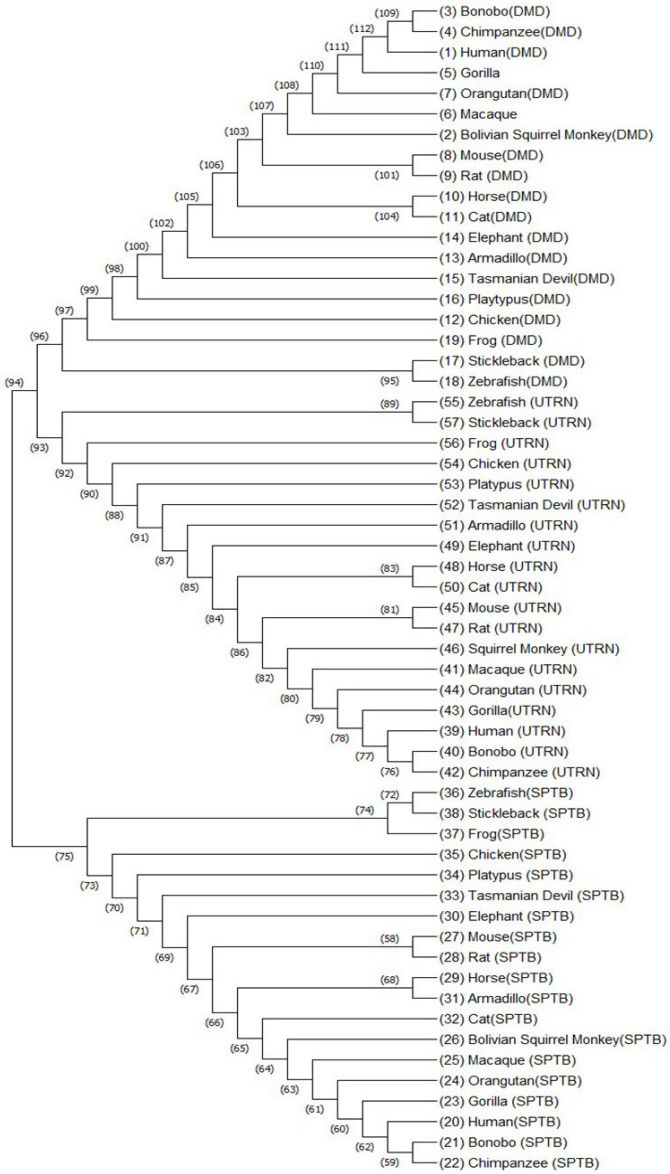

Figure 10: *Evaluation of genetic variation and deviation from neutrality using Fu and Li's test.*

High level of haplotype diversity

The population under investigation displays an incredibly high level of haplotype multiplicity, as evidenced by the haplotype diversity index (Hd) being analogous to 1.000. This conclusion indicates the existence of a significant number of unique haplotypes within the studied population. Such an

increased extent of genetic diversity implies the gene being subjected to selective forces, rather than just going through random genetic drift.

Excessive number of singleton mutations

The analysis demonstrates a considerable excess of singleton mutations, with 822 out of the total 847 mutations being distinctive to single individuals. The existence of a substantial number of singleton mutations implies the population has encountered somewhat recent genetic changes, potentially steered by adaptive procedures.

Nucleotide diversity (Pi)

The high haplotype diversity, overabundance of singleton mutations, and the nucleotide variation observed in the data collectively indicate that the gene under investigation is subject to selective forces, potentially driven by recent adaptive circumstances, rather than being impacted exclusively by random genetic drift. These conclusions provide valuable understanding into the evolutionary dynamics and genetic arrangement of the studied population.

Analysis of phylogenetic tree

The *UTRN* (Utrophin) gene comprises a distinct clade below the *DMD* clade with multiple species. *SPTB* (Spectrin Beta) gene constructs another singular clade further below the *UTRN* clade. Since the *UTRN* clade is placed between the *DMD* clade and the SPTB clade, this implies that *UTRN* is more closely associated to *DMD* than *SPTB*. These results are visually portrayed in the phylogenetic tree displayed in Figure 10.

Functional annotation analysis of candidate gene

The Duchenne muscular dystrophy (*DMD*) protein is a complicated and multifunctional molecule that plays a critical role in different cellular functions. Using the *SMART* (Simple Modular Architecture Research Tool) bioinformatics mechanism, several different functional domains and areas are determined within the *DMD* protein. These determined domains are thought to be accountable for the protein's diverse functional capabilities, as they likely contribute to its structural integrity, interactions with other molecules, and overall biological actions.

Identified functional regions and domains of DMD gene

Dystrophin (*DMD*) protein displays a diverse array of structural and functional regions that contribute to its vital biological functions. The specified domains and functional zones of the *DMD* protein include:

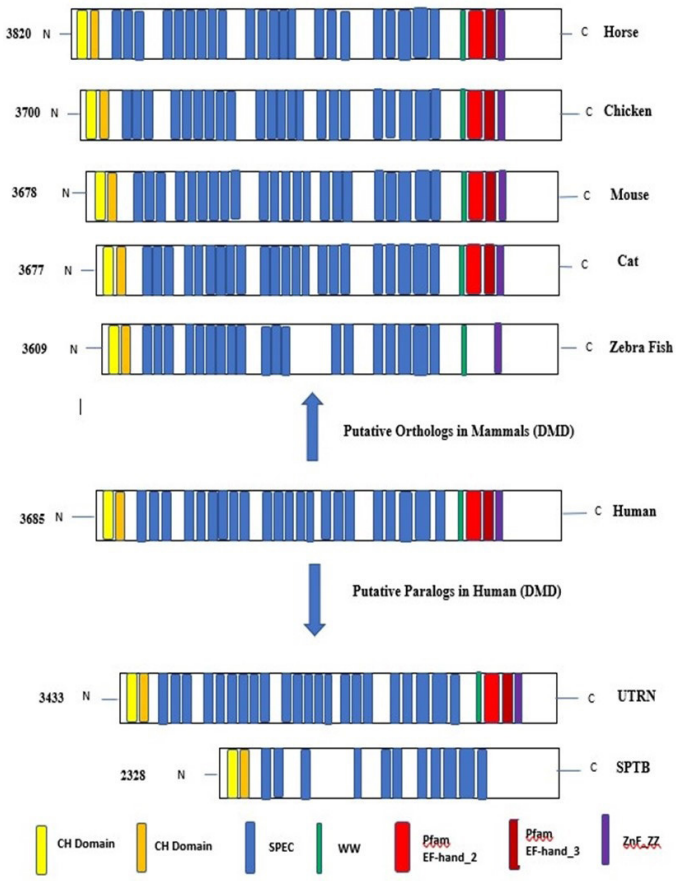

Figure 11: *Phylogenetic tree by maximum likelihood method.*

CH (calponin homology domain)

The calponin homology (*CH*) region is a structural motif found at the N-terminus of the *DMD* protein, mainly implicated in the organization and dynamics of the actin cytoskeleton. There are two *CH* domains in *DMD* protein. One is found at position (17-117) while the second *CH* domain is found at position (136-235). Depicted in yellow and orange color, these areas play a vital role in the binding and cross-linking of actin filaments, which is necessary for the construction of actin-based configurations, such as bundles and networks. The existence of a predicted *CH* domain has also been specified in the Cdc24 protein. Cdc24 is a pivotal regulator of cell division and polarization, indicating that the *CH* domain may be concerned with the spatial organization and dynamics of the actin cytoskeleton during these important cellular procedures.

SPEC (spectrin repeats)

These structural characteristics are important for the proper assembly and function of the cytoskeletal

network, which provides mechanical support and organization to the cell's interior. Moreover, twenty-two SPEC domains at the N-terminus of Human DMD were found between amino acid positions (1: 343-446, 2: 451-555, 3: 562-666, 4: 722-827, 5: 833-933, 6: 942-1044, 7: 1051-1153, 8: 1160-1262, 9: 1269-1366, 10: 1468-1567, 11: 1574-1675, 12: 1682-1779, 13: 1784-1873, 14: 1880-1970, 15: 1998-2100, 16: 2107-2207, 17: 2214-2317, 18: 2473-2576, 19: 2583- 2685, 20: 2692-2801, 21: 2808-2930, 22: 2973-3039).

WW (WWP or rsp5 domain)

Illustrated in green, the domain is distinguished by the existence of two positively conserved tryptophan (Trp or W) amino acid residues, and is thus often referred to as the *WWP* (Trp-Trp-Pro) or Rsp5 domain. The two tryptophan residues are thought to be vital for mediating these protein-protein interactions, accentuating the importance of the unique structural architecture of this domain. *WWP* domain is found at positions (3056-3088) portrayed in Figure 11.

EF-hand domains (Pfam Domain)

Depicted in Red and Maroon, the EF-hand domain adopts a distinctive helix-loop-helix tertiary configuration, identical to other EF-hand motifs found in different proteins. The preliminary role of this EF-hand-like domain seems to be the provision of particularity in the distinction and binding of beta-dystroglycan, a pivotal element of the dystrophin-glycoprotein complex. This interchange enables to promote the association between dystrophin and the dystroglycan complex.

ZnF (zinc finger domain)

Displayed in purple, the zinc-binding territory found in the dystrophin protein plays a critical part in its function. The distinct "ZZ" motif within the dystrophin configuration seems to have the ability to attach to the calcium-binding protein calmodulin. This ZZ-type zinc finger field is strategically placed between a WW domain (which is flanked by EF-hand) of the zinc-binding domain in dystrophin are vital for retaining the integrity of dystrophin between the intracellular and extracellular components of the muscle, and any disturbances in this domain or its interactions can have harsh outcomes for muscle function and fitness. The ZnF domain is found at positions of (3307-3352).

Functional characteristics of structural domains

N-terminal domain

The N-terminal domain of dystrophin is found from 1-240 amino acids. The N-terminal domain encodes the dystrophin protein, which binds to F-actin in the cytoskeleton of muscle cells.

C-terminal domain

The dystrophin protein is composed of 3685 amino acids. Particularly, the C-terminal domain spans roughly from 3240-3685 amino acids. The major role of C-terminal domain is to anchor dystrophin to the muscle cell membrane by attaching to β-dystroglycan, thereby retaining muscle cell integrity and strength during contraction and relaxation.

Comparative domain analysis across different species

Analyzing the domains across different species highlights considerable conservation with some deviation in species. The results are as follows:

Putative orthologs in human DMD

The proteins from different species, such as horse (3820 amino acids), chicken (3700 amino acids), mouse (3678 amino acids), and cat (3677 amino acids), display a favorably conserved domain configuration. These orthologous proteins, which have developed from a common ancestral gene, share an incredible resemblance in their structural organization, constituting key domains such as the CH (calponin homology), SPEC (spectrin repeat), WWP (WW domain-containing protein), EF-hand 2, EF-hand 3, and ZnF (zinc finger) domains. This implies that the arrangement and structural characteristics of zebra fish protein variants vary from other organisms, potentially indicating functional variation.

Figure 12: *Comparative domain analysis of DMD gene across orthologs and paralogs.*

Putative paralogs in human DMD

UTRN (3433 amino acids) indicates a high degree of resemblance to *DMD*, especially in the conservation of the CH domains and spectrin repeats, implying functional redundancy and compensation in the absence of dystrophin. *SPTB* (2328 amino acids) has a shorter sequence with lesser spectrin repeats and absence of WW, EF-hand 2, EF-hand 3 and ZnF domains. It reflected functional deviation where spectrin might play more restricted roles described in relation to dystrophin and utrophin.

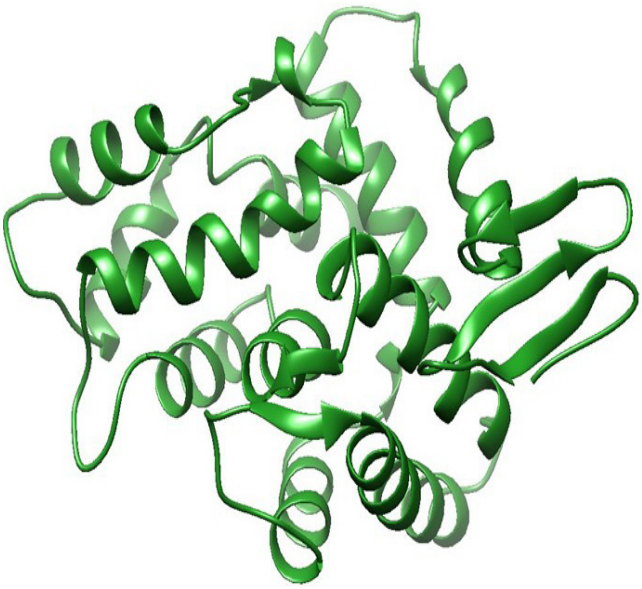

Figure 13: *Human DMD Domain protein structure.*

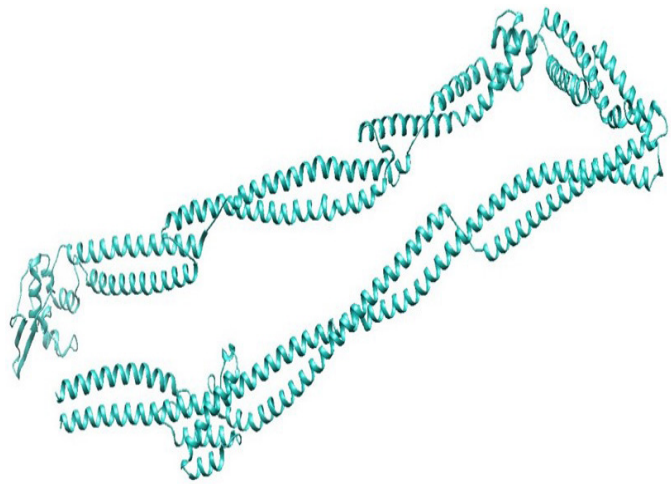

Figure 14: *Chimpanzee DMD protein structure.*

Assessing evolutionary conservation of protein structures

PHYRE2, a computational tool was utilized to specify the three-dimensional protein configurations of *DMD* protein in humans Figur 13, chimpanzees, and their common evolutionary predecessor displayed in Figure 12 and Figure 14. These protein configurations are visually demonstrated below.

Comparative analysis of human and ancestral protein structures

The investigation demonstrates that while the human *DMD* protein and its ancestral companion share well-aligned area, there are substantial divergences in other areas, as shown by the high overall *RMSD* and low Q-score in Figure 15. These findings suggest that although certain regions of the proteins are structurally similar, the overall structures diverge considerably due to differences in sequence length and domain composition Figure 16.This comprehensive analysis highlights the complexity of protein structure-function relationships, where sequence similarity does not always correlate with structural congruenceshown in Figure 17 and Figure 18.

Computing secondary structure assignments for model(s) #0, #1
using ksdssp (Kabsch and Sander Define Secondary Structure
of Proteins) with the parameters:
energy cutoff -0.5
minimum helix length 3
minimum strand length 3

Matchmaker Ancestor_SPEC_13-ZnF.pdb (#0) with Human_Spec_13-ZnF.pdb (#1),
sequence alignment score = 99.7
with these parameters:
chain pairing: bb
Needleman-Wunsch using BLOSUM-62
ss fraction: 0.3
gap open (HH/SS/other) 18/18/6, extend 1
ss matrix: (O, S): -6 (H, O): -6 (H, H): 6 (S, S): 6 (H, S): -9 (O, O): 4
iteration cutoff: 2
RMSD between 6 pruned atom pairs is 1.519 angstroms; (across all 240 pairs: 87.823)

Match→Align cutoff: 5.0, in column if within cutoff of: any
25 residue pairs aligned
25 fully populated columns

Evaluating superpositions across all 25 fully populated columns in the final alignment:
RMSD of Node_SPEC_13-ZnF.pdb with Human_Spec_13-ZnF.pdb: 2.916
Overall RMSD: 2.916
Sequence lengths: 1009 245
SDM (cutoff 5.0): 169.437
Q-score: 0.001

Figure 15: *Ancestral DMD domain protein structure.*

Figure 16: *Matchmaker tool results for structural comparison of human and ancestor.*

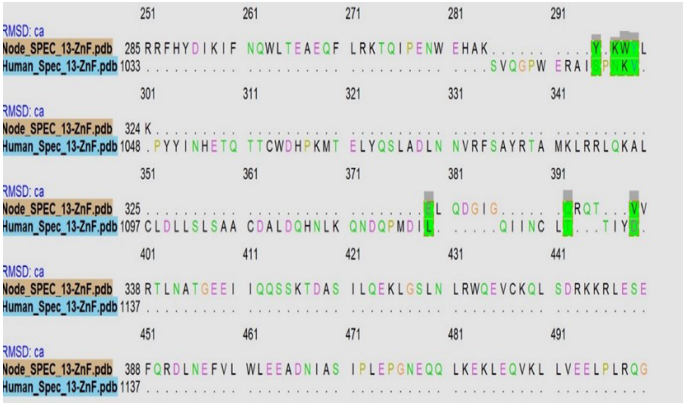

Figure 17: *Comparison of human and ancestor DMD protein structure.*

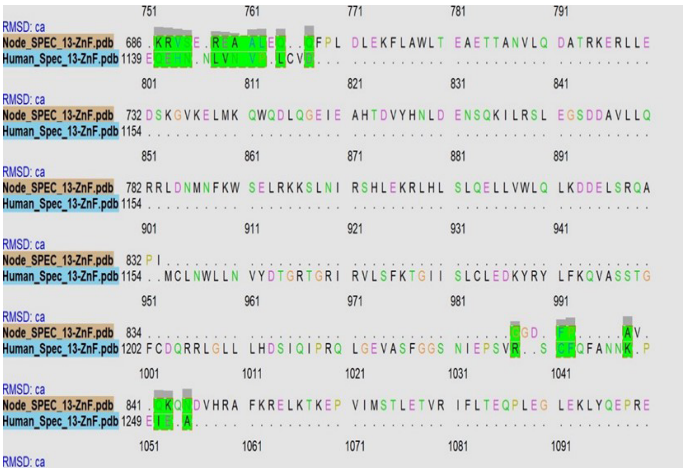

Figure 18: *Sequence alignment of human and ancestor DMD protein.*

```
Computing secondary structure assignments for model(s) #0, #1
using ksdssp (Kabsch and Sander Define Secondary Structure
of Proteins) with the parameters:
  energy cutoff -0.5
  minimum helix length 3
  minimum strand length 3

Matchmaker Ancestor_SPEC_13-ZnF.pdb (#0) with ChimP_Spec_13-ZnF.pdb (#1),
sequence alignment score = 3544
with these parameters:
        chain pairing: bb
        Needleman-Wunsch using BLOSUM-62
        ss fraction: 0.3
        gap open (HH/SS/other) 18/18/6, extend 1
        ss matrix: (O, S): -6 (H, O): -6 (H, H): 6 (S, S): 6 (H, S): -9 (O, O): 4
        iteration cutoff: 2
RMSD between 49 pruned atom pairs is 1.122 angstroms; (across all 808 pairs: 132.767)

Match→Align cutoff: 5.0, in column if within cutoff of: any
67 residue pairs aligned
67 fully populated columns

Evaluating superpositions across all 67 fully populated columns in the final alignment:
RMSD of Node_SPEC_13-ZnF.pdb with ChimP_Spec_13-ZnF.pdb: 1.872
Overall RMSD: 1.872
Sequence lengths: 1009 808
SDM (cutoff 5.0): 129.018
Q-score: 0.004
```

Figure 19: *Matchmaker tool results for structural comparison of ancestor and chimpanzee.*

Comparative analysis of ancestor and chimpanzee protein structures

Overall, the analysis reveals a high degree of sequence similarity between both proteins with significant structural differences in Figure 20. The conserved core regions exhibit strong structural similarity, evidenced by the low *RMSD* of 1.122 Å in pruned areas and 1.872 Å in fully aligned columns. However, the overall high *RMSD* of 132.767 Å for all pairs, the *SDM* value of 129.018, and the low Q-score indicated substantial structural variations beyond these well-aligned areas in Figure 19 The protein comparison suggests that, despite evolutionary conservation in key functional domains, there are species-specific structural adaptations in these proteins shown in Figure 21.

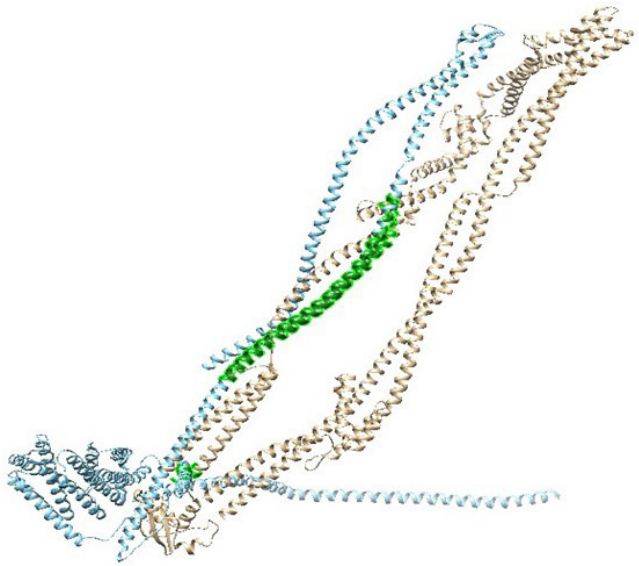

Figure 20: *Comparison of chimpanzee and ancestor DMD protein structure.*

Figure 21: *Sequence alignment of chimpanzee and ancestor DMD protein.*

Conclusion

The rapidly progressing field of genomics has

facilitated extensive bioinformatics analyses across various animal species, providing valuable insights into the evolutionary affinities of genes associated with distinct phenotypic characteristics and human skeletal diseases. A meticulous analysis of the *DMD* gene across primates has shed light on its evolutionary dynamics and functional significance. The presence of human-specific mutations and positive selection signals, as demonstrated by the Ka/Ks ratio, implies that this gene has experienced adaptive evolution and functional transformations within the human heritage. The calculated Ka/Ks value was found to be 1.5 in humans. Further research through SWAKK analysis, Tajima's D, and Fu and Li's tests has yielded a negative value of -1.71219 with a p-value less than 0.05, while Fu and Li's test produced a negative value of -1.95457, negative values for both tests indicated deviation from neutral evolution, all of which lead to an excess of rare alleles and genetic diversity at specific loci. Phylogenetic estimation of the *DMD* gene and its paralogs, *UTRN* and *SPTB*, has demonstrated powerful evolutionary affinities across vertebrates. The maximum likelihood method has confirmed the close evolutionary affinity between *DMD* and *UTRN*.

Recommendations

The study of the *DMD* gene, which is responsible for the growth and function of muscles, and its evolutionary dynamics supplies a valuable prospect to extend our understanding across various scientific domains. Investigating the distinctive mutations and adaptive evolution of the *DMD* gene in human sheds light on its key function in muscular disorders and skeletal development, potentially directing the development of more targeted and useful therapies. Moreover, advanced bioinformatics can enable the extension of these analyses to other genes associated with skeletal disorders and morphological characteristics across a wider spectrum of species, facilitating a more thorough awareness of the genetic signatures of these traits. Beyond the significance for human health and biology, the findings from the evolution of the *DMD* gene can also shape conservation methods for endangered species, supporting the maintenance of hereditary diversity and improving the strength of these species in the face of environmental challenges. For further experimental validation, a site directed mutagenesis in vitro protein expression maybe done in future to analyze structural and functional effects of *DMD* substitutions. Additionally, the biological relevance may also be confirmed by doing the *in-vivo* studies in engineered model organisms (like mouse) carrying the human-specific mutations.

Acknowledgement

I acknowledge faculty and HOD of Biotechnology department, University of Sialkot for insightful suggestion regarding the research.

Novelty Statement

The present study uniquely integrates positive selection signatures, comparative domain architecture, and structural divergence analyses to highlight candidate mutations that may have contributed to human-specific skeletal adaptations. These findings provide a basis for experimental validation, advanced structural studies and functional analysis of other novel skeletal genes.

Author's Contribution

SN designed the methodology and written the manuscript. FS collected the data and TS assisted in result interpretation.
ZI, UM and IA contributed to manuscript editing, validation of findings, and reference management.
All authors have read and approved the final version of the manuscript.

Generative AI and AI assisted technology statement
The authors declare that no generative AI and AI assisted technology was used in the creation of this manuscript.

Conflict of interest
The authors have declared no conflict of interest.

References

Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C. and Glaubitz, J.C.J.S., 2009. The genetic architecture of maize flowering time. Sciences, 325(5941): 714-718. https://doi.org/10.1126/science.1174276

Crompton, R.H., Sellers, W.I., Thorpe, S.K.J.P.T. o.t. R.S.B.B.S. (2010). Arboreality, terrestriality and bipedalism. Philosophical Transactions of the Royal Society B: Biological Sciences,

365(1556), 3301-3314. http://doi:10.1098/rstb.2010.0035

Eckhardt, R.B.J.A.A., 1989. Evolutionary morphology of human skeletal characteristics. pp. 193-228. https://doi.org/10.1127/anthranz/47/1989/193

Farrokhi, A., Guin., Naidu., (2021). Investigation on the impact of different vertical and horizonyal resolutions on the applicability of linear hybrid inverter.

Fischer, M.D., Gorospe, J.R., Felder, E., Bogdanovich, S., Pedrosa-Domellof, F., Ahima, R.S., Khurana, T.S.J.P.g. (2002). Expression profiling reveals metabolic and structural components of extraocular muscles. Physiological Genomics, 9(2), 71-84. https://doi.org/10.1152/physiolgenomics.00115.2001

Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J. and Wilczek, A.M.J.S., 2011. A map of local adaptation in Arabidopsis thaliana. Sciences. 334(6052): 86-89. https://doi.org/10.1126/science.1209271

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L. and Down, T.J.N.A.R., 2002. The Ensembl genome database project. Nucleic Acids Res., 30(1): 38-41. https://doi.org/10.1093/nar/30.1.38

Kralick, A.E. and Zemel, B.S.J.F.I.E., 2020. Evolutionary perspectives on the developing skeleton and implications for lifelong health. Front. Endocrinol.. 11: 513066. https://doi.org/10.3389/fendo.2020.00099

Kun, E., Javan, E.M., Smith, O., Gulamali, F., de la Fuente, J., Flynn, B.I. and Tucker-Drob, E.E.M.J.S., 2023. The genetic architecture and evolution of the human skeletal form. Sciences, 381(6655): eadf8009. https://doi.org/10.1126/science.adf8009

Pervaiz, N. and Abbasi, A.A.J.M.G., 2016. Molecular evolution of WDR62, a gene that regulates neocorticogenesis. Meta Gene, 9: 1-9. https://doi.org/10.1016/j.mgene.2016.02.005

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P.J.P.O.T.N.A.O.S., 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. Proceed. Nat. Acad. Sci., 95(11): 5857-5864. https://doi.org/10.1073/pnas.95.11.5857

Sgro, J.Y., 2017. UCSF Chimera-I-Introduction.

Stanfill, A.G. and Cao, X.J.B.R.F.N., 2021. Enhancing research through the use of the genotype-tissue expression (GTEx) database. Biol. Res. Nursing. 23(3): 533-540. https://doi.org/10.1177/1099800421994186

Yang, Z.J.M.B., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mole. Biol. Evolut., 24(8): 1586-1591. https://doi.org/10.1093/molbev/msm088