WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Monkeypox Genome Analysis and Phylogenetic study using Biopython

***Pritam P. Sose*[1*]**

[1]Padmashree Institute of Management and Sciences, Bengaluru, Karnataka, India.
Email Id: pritamsose@gmail.com
Orcid Id: 0000-0002-1191-1200
Doi : https://doi.org/10.55248/gengpi.5.1024.2712

ABSTRACT:

The Human Monkeypox virus, an etiological agent of Monkeypox disease, manifests with clinical symptoms akin to smallpox but generally exhibits reduced severity. Since 2022, Monkeypox has emerged as a global health threat, raising significant concerns about its transmission. In this study, we conduct a comprehensive genomic analysis and phylogenetic examination of the Monkeypox virus, leveraging multiple sequence alignments of viral proteins, facilitated by the Biopython package. Key findings reveal that the GC content of the viral DNA stands at 33.03%, with a molecular weight of 60,910,683.70 Da. Notably, leucine is the most abundant amino acid in the viral proteome, followed by isoleucine. Additionally, this investigation elucidates the phylogenetic relationships of the Human Monkeypox virus, particularly the DNA ligase-2 protein, and its evolutionary kinship across various strains. A phylogram, based on protein accession AAU01367.1, was generated, incorporating amino acid sequence alignments for all Monkeypox strains available in NCBI. Furthermore, BLAST analysis of the UKV32022.1 strain revealed a 99.80% sequence similarity to AAU01367.1. The highest phylogenetic branch length was observed in the protein with accession WZB41963.1, showing a branch length of 0.022 and comprising 35 taxa. The AAU01367.1 protein features the conserved domain PHA02782 Superfamily (accession no. cl31504) with an e-value of 0 across its 505 amino acids, which was pivotal in constructing the phylogenetic tree. As the three-dimensional structure of this protein remains unavailable, an attempt was made to predict and visualize the 3D structure of the AAU01367.1 protein using Biopython.

KEYWORDS: Mpox; Human Monkeypox disease; Monkeypox virus; Biopython, Genome analysis; Phylogenetic analysis; DNA Ligase-2 protein [AAU01367.1]; Protein 3-D structure of Mpox.

INTRODUCTION:

In the wake of the SARS-CoV-2 pandemic, the emergence of the monkeypox virus (MPXV) has raised alarming global health concerns, marking yet another significant zoonotic threat to humanity. As a member of the Orthopoxvirus genus, MPXV is a double-stranded DNA virus with clinical manifestations akin to smallpox, though generally less severe. First documented in 1970 in a nine-year-old child from the Democratic Republic of Congo, the virus has since remained endemic in certain regions of Africa. However, in recent years, it has spread far beyond these borders, prompting a re-evaluation of global public health strategies [1-3].

Beginning in May 2022, an unprecedented outbreak of monkeypox has seen more than 3,000 confirmed infections across 50 countries within just a few months. As of October 15, 2022, over 73,426 cases have been reported across 110 nations, illustrating the virus's alarming capacity for widespread transmission. Echoing the dynamics observed during the COVID-19 pandemic, MPXV's transmission routes—particularly through human-to-human contact—pose a clear threat to global health stability, risking the onset of another public health crisis [4-6].

Despite being identified more than six decades ago, MPXV has recently resurged, largely due to waning immunity in populations no longer vaccinated against smallpox. The cessation of smallpox vaccination, following the disease's eradication, has been identified as a critical factor in the vulnerability of modern populations to monkeypox. Studies indicate that smallpox vaccination is approximately 85% effective in preventing monkeypox infection, underscoring the importance of immunological preparedness. In addition to decreased herd immunity, other contributing factors to the spread of MPXV include high population densities, increased global mobility, and heightened exposure to animal reservoirs of the virus [7-8].

From a virological standpoint, MPXV is morphologically distinct, presenting as a rectangular or ovoid-shaped virion, approximately 200 x 250 nm in size. Its structure and genome, detailed in this study, provide key insights into the mechanisms underlying its pathogenicity and transmission. With these growing concerns, comprehensive genomic analysis and phylogenetic studies are essential to understanding MPXV's evolution and guiding public health interventions. This paper aims to contribute to that endeavor by leveraging advanced computational tools such as Biopython to facilitate genome analysis and phylogenetic studies of the monkeypox virus [9-10].

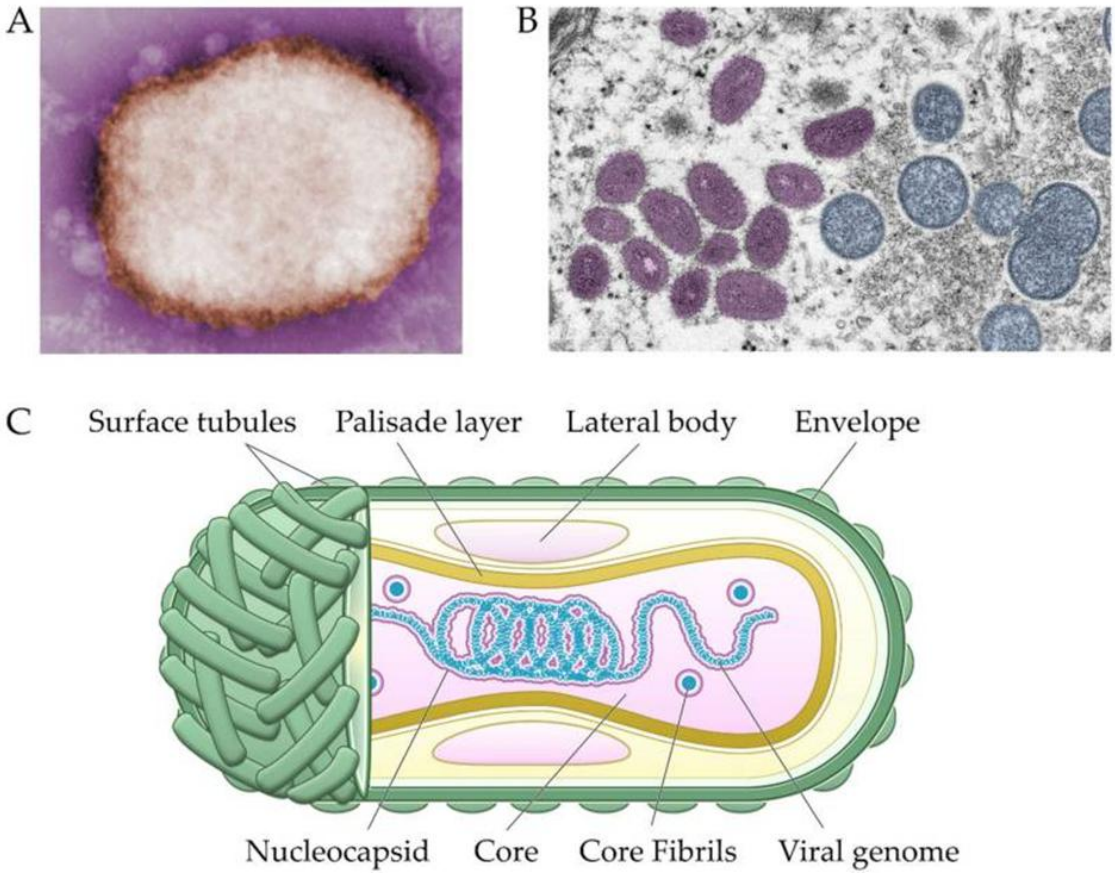


Figure 1. (A)Negative-stained transmission electron micrograph of M-type Monkeypox viral particle. (B) A thin section of the sample contains an ovoid virus and below is its structure. (C) Micrographs are courtesy of the Centres for Disease Control and Prevention (CDC) Public Health Image Library (PHIL). Schematic representation of Mpox virus structure [11].

PRESENT FINDINGS:

```
!pip install Biopython
from Bio import Entrez, SeqIO
Entrez.email = ""
handle = Entrez.efetch(db="nucleotide", id="NC_063383", rettype="gb", retmode="text")
recs = list(SeqIO.parse(handle, 'gb'))
handle.close()
```

```
Mpox_dna = recs[0].seq
print(f'The genome of Mpox consists of {len(Mpox_dna)} nucleotides.')
```

```
# molecular weight
from Bio.SeqUtils import molecular_weight
molecular_weight(Mpox_dna)
```

```
# GC content - higher GC content implies more stable molecule due to G and C forming triple hydrogen bonds
from Bio.SeqUtils import gc_fraction
print(f'The GC content of Mpox genome is {gc_fraction}%')
a = gc_fraction(Mpox_dna)
GC = a*100
GC
```

```
The GC content of Mpox genome is <function gc_fraction at 0x7e1468339f30>%
33.029932710981754
```

```
count_nucleotides = {
    'A': Mpox_dna.count("A"),
    'T': Mpox_dna.count("T"),
    'C': Mpox_dna.count("C"),
    'G': Mpox_dna.count("G")
}

print(count_nucleotides)
```

```
{'A': 66193, 'T': 65878, 'C': 32602, 'G': 32536}
```

Figure 2: Biopython program to get sequence analysis

The comprehensive genome analysis of the Monkeypox virus (MPXV), particularly strain NC_063383, reveals that its genome comprises 197,209 nucleotides. The calculated molecular weight of the viral genome is approximately 60,910,683.70 Daltons. Notably, the GC (guanine-cytosine) content of the MPXV genome stands at 33.03%, a figure that suggests a higher mutation rate compared to SARS-CoV-2, which exhibits an approximate GC content of 37%. This lower GC percentage is a crucial indicator of the genetic flexibility and mutability of MPXV, potentially facilitating its adaptation and transmission within varied host environments.

The nucleotide composition of MPXV presents a total of 66,193 adenosine (A) bases, 65,878 thymidine (T) bases, 32,602 cytidine (C) bases, and 32,536 guanosine (G) bases. The relative abundance of adenine and thymine over cytosine and guanine further corroborates the low GC content, suggesting a genome structure more prone to mutations, which could influence viral evolution and pathogenicity. The detailed nucleotide frequency of each base is provided in the subsequent table, accompanied by a corresponding nucleotide frequency plot.

This intricate genomic profiling of MPXV not only enhances our understanding of its molecular architecture but also serves as a foundational basis for studying its phylogenetic relationships. Such insights are pivotal for predicting viral behavior, monitoring potential genetic shifts, and devising effective control strategies, particularly as MPXV continues to pose a substantial threat to global public health.

Table 1: Occurrence and content of A, T, C, G base pairs

Serial No.	**Nitrogen base**	**Representation**	**Complementary base in DNA**	**Chemical formula**	**Pairing where : is One Hydrogen Bond**	**Occurrence in Genome in number**
1	Adenine	A	Thymine	C 5H5N5	A::T	66193
2	Guanine	G	Cytosine	C5H5N5O	G:::C	65878
3	Thymine	T	Adenine	C5H6N2O2	T::A	32602
4	Cytosine	C	Guanine	C4H4N3O	C:::T	32536
						Total =197209

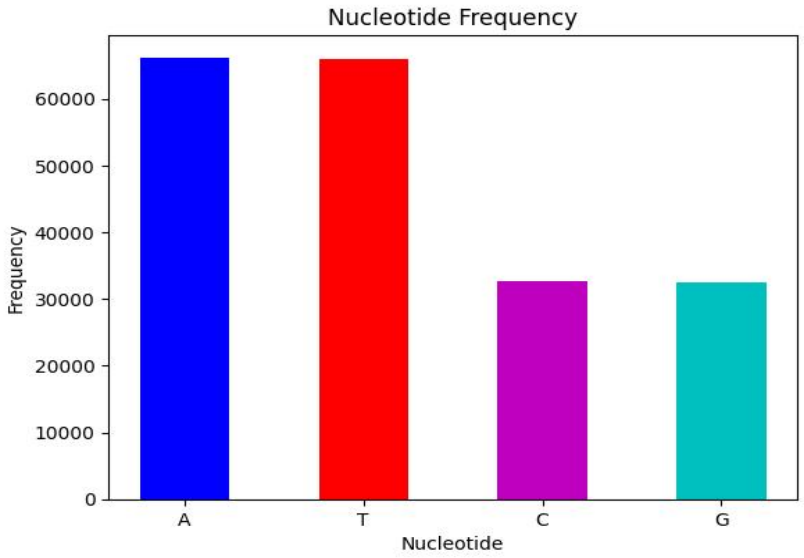


Figure 3: Bar plot indicating Nucleotide Frequency in viral Genome

The frequency of most common amino acid is:

'Leucine(L)', 6920)>('I(Isoleucine)',6772)>('S(Serine)', 5568)>('STOP CODONS', 4471)>('Y(Tyrosine)', 3834)>('R(Arginine)', 3793)>('F(Phenylalanine)', 3676)>('T'(Threonine), 3578)>('V(Valine)', 3551)>('N(Asparagine', 3437)]

which can be visualized graphically as

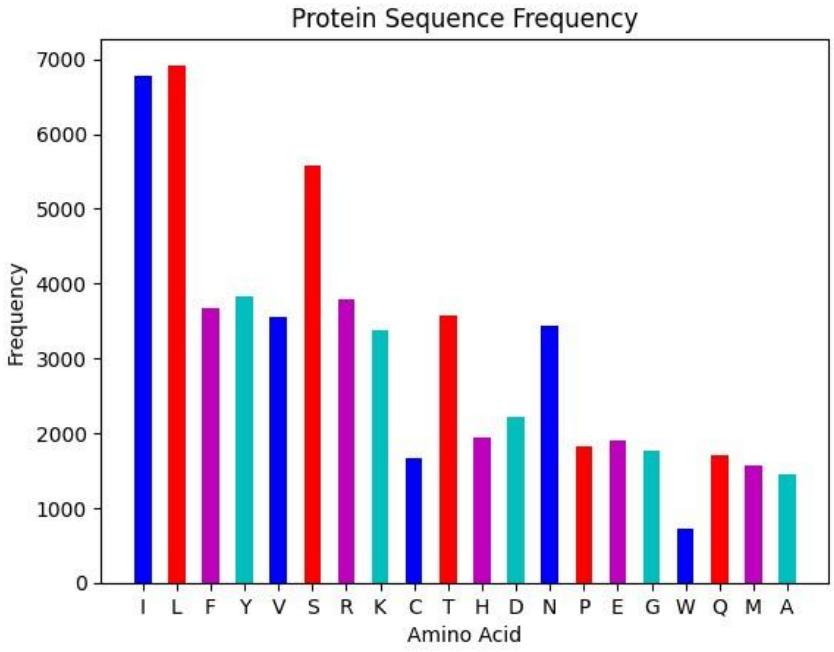


Figure 4: Bar plot of amino acid frequency in proteins

Table 2: AMINO ACIDS with their content in number.

Sr No.	Amino Acids	Three Letter Code	Single Letter Code	Content in number
1	Leucine	LEU	L	6920
2	Isoleucine	ILE	I	6772
3	Serine	SER	S	5568

4	Stop codon	*	*	4471
5	Tyrosine	TYR	Y	3834
6	Arginine	ARG	R	3793
7	Phenylalanine	PHE	F	3676
8	Threonine	THR	T	3578
9	Valine	VAL	V	3551
10	Asparagine	ASN	N	3437

The monkeypox virus (MPXV) genome encodes a total of 61,265 amino acids, with leucine emerging as the most predominant amino acid, comprising 6,120 residues. Following closely is isoleucine, with a total count of 6,772, reflecting a similar pattern of dominance across the entire proteome. This amino acid prevalence is mirrored in MPXV's largest protein, further affirming the structural and functional importance of these hydrophobic amino acids within the viral architecture.

The total protein repertoire of MPXV consists of approximately 4,472 individual proteins, among which 2,473 proteins exceed 20 amino acids in length and are classified as functional proteins due to their structural complexity and biological significance. Notably, the largest protein encoded by the MPXV genome is 785 amino acids long, and, consistent with the overall proteome, leucine emerges as the most frequent amino acid within this large protein as well. This leucine dominance is indicative of its role in maintaining protein stability, facilitating hydrophobic interactions, and potentially influencing protein folding and function in the virus's replication cycle.

A detailed analysis of the amino acid distribution, as represented graphically, reveals leucine's significant presence across MPXV proteins, reinforcing its critical contribution to the virus's protein architecture. This extensive presence of leucine within MPXV's largest and most functional proteins may offer insights into the virus's pathogenic mechanisms, warranting further exploration into how amino acid composition influences the viral lifecycle and host interactions.

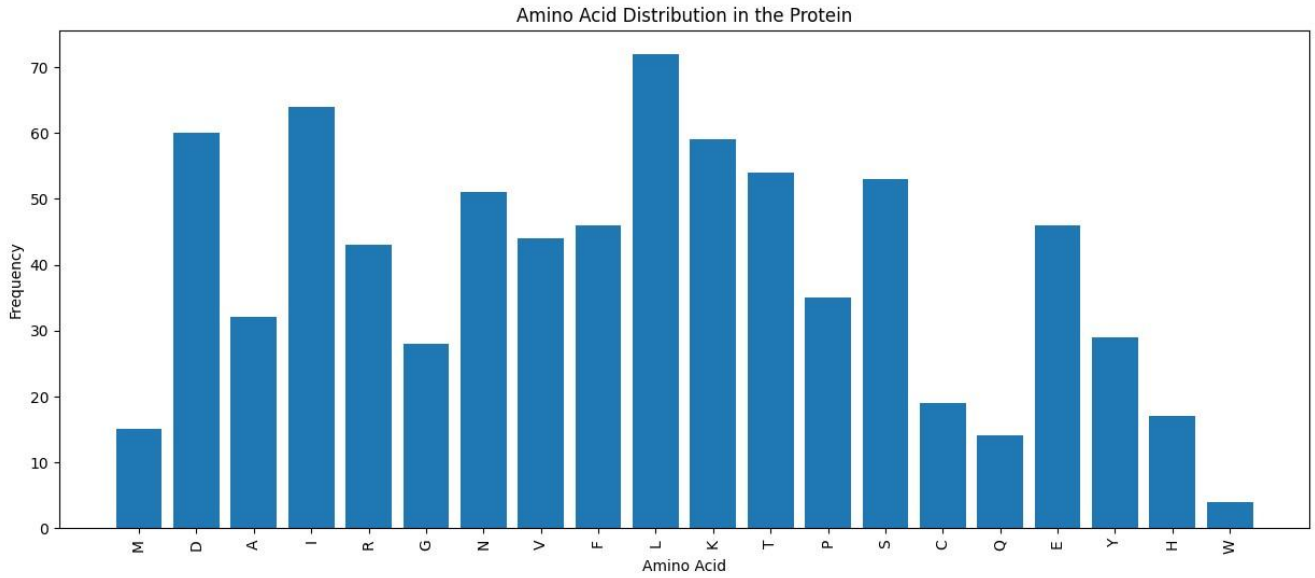


Figure 5: Amino acid distribution in Largest protein

And its 3D structure is as follows.

3-D STRUCTURE OF MPXV-WRAIR157 [Monkeypox virus] DNA LIGASE-2:-

ACCESSION NO :- AAU01367.1

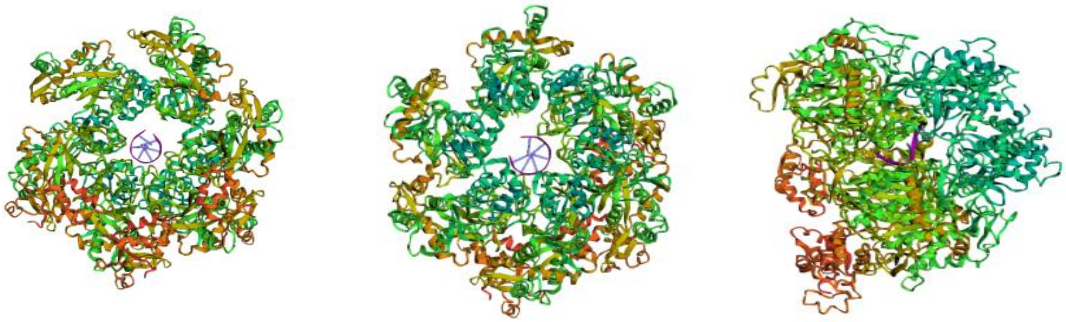


Figure 6: The 3-D structures of the Protein.

The phylogenetic analysis reveals that the sequences UKV32022.1 and WHP53763.1, positioned as sister nodes, exhibit the closest evolutionary proximity to AAU01367.1, with taxa and branch lengths measured at approximately 1.78/0.0020 and 1.95/0.00198, respectively. This close genetic relationship suggests that these strains share a recent common ancestor, reinforcing their classification within the same clade.

Among the analyzed taxa, taxa 38 displays the greatest genetic diversity, as evidenced by its extensive branching within the phylogram. This heightened diversity may indicate a wider range of mutations or adaptations within this clade. Furthermore, the strain WZB41963.1 stands out with the highest branch length, measured at approximately 0.022, and is associated with taxa 35. The extended branch length here suggests a significant evolutionary divergence, possibly indicative of unique mutations or evolutionary pressures acting on this strain.

The second-longest branch length was observed in protein accession UXL69418.1, with a taxa count of 52 and a branch length of 0.018. Notably, UXL69418.1 and UW044944.1 appear to be closely related, with UW044944.1 possessing a taxa count of 54 and a branch length of 0.016. These findings highlight the close evolutionary relationship between these two sequences, which may have diverged from a common ancestral strain in recent evolutionary history.

In summary, the phylogenetic analysis conducted in this study underscores the evolutionary diversity within the monkeypox virus genome, with specific clades showing varying degrees of divergence. The use of Biopython has enabled a precise and comprehensive assessment of genetic relationships, providing critical insights into the evolutionary trajectories of the MPXV strains analyzed.

select all 100 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
MPXV-WRAIR157 [Monkeypox virus]	Monkeypox virus	1048	1048	100%	0.0	99.80%	505	AAU01367.1
MPXVgp165 [Monkeypox virus]	Monkeypox virus	1048	1048	100%	0.0	100.00%	503	UVK32022.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1047	1047	100%	0.0	99.80%	503	YP_010377160.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	XBJ99146.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	UXP66363.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	UYD51645.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	WOW77273.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	UXX14004.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	UZZ60652.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	UTS55304.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	XDT98487.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	UWK03881.1
MPXVgp165 [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	WNN26079.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	WCS72415.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	WHP53612.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	WHP57320.1
Schlafen (1) [Monkeypox virus]	Monkeypox virus	1046	1046	100%	0.0	99.60%	503	WCB61882.1

Figure 7: Protein sequence AAU01367.1 BLAST result

BLAST i.e. Basic local alignment searching tool also showed that the similarity between UKV32022.1 with AAU01367.1 is 99.80% in percentage. This result indicates that phylogenetic analysis is successfully carried out with more correctness.

SUPPLEMENTARY MATERIALS: -

1) Aligned_sequences (4).fasta contains alignment format of protein file.

2) Mpox_protein.xlsx contains Protein sequence and its length in amino acids.

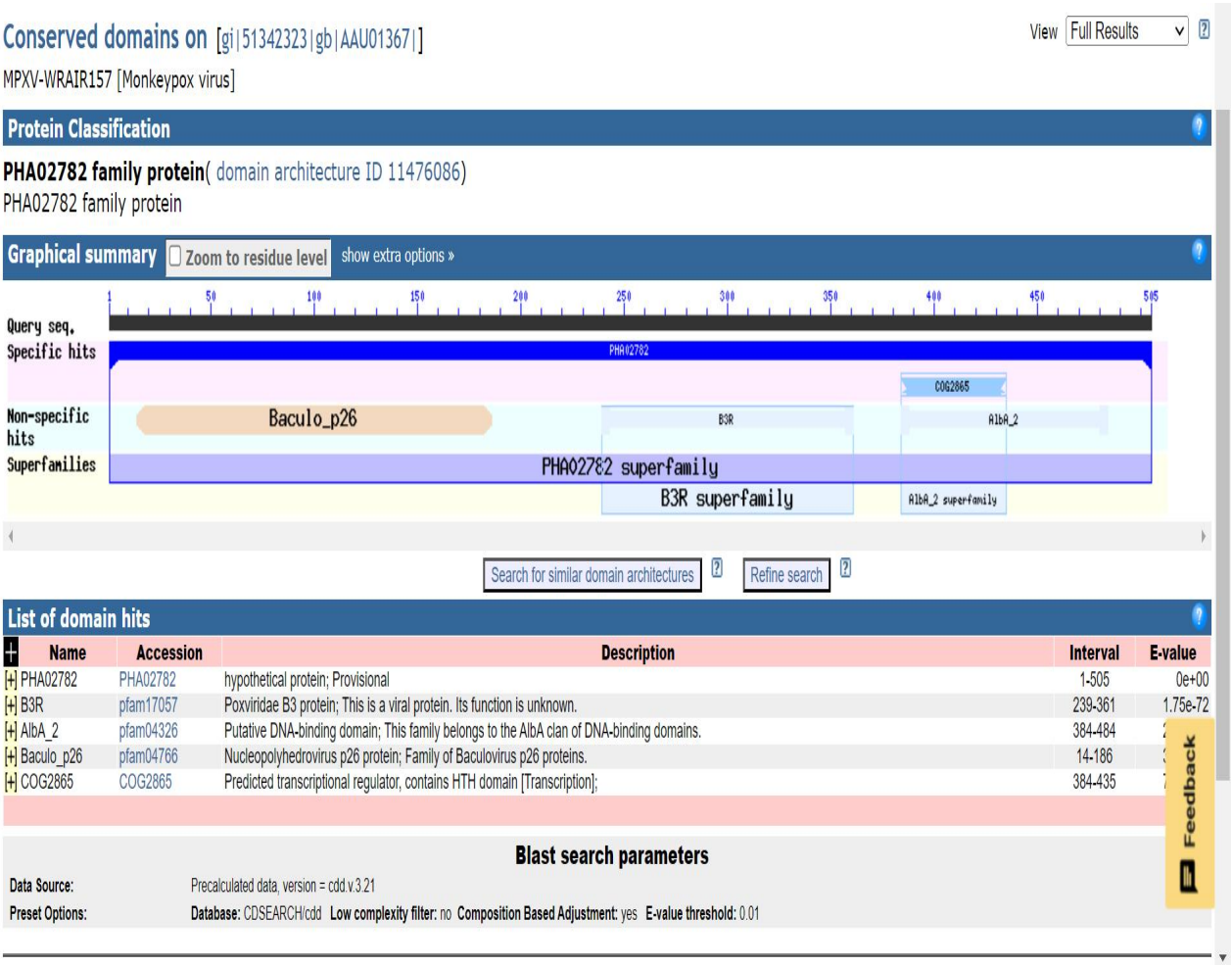


Figure 8: Conserved Domain analysis of Protein AAU013.67.1

CONCLUSION AND FUTURE RESEARCH SCOPES

In conclusion, this comprehensive genomic and phylogenetic analysis of the Human Monkeypox virus (MPXV) using the Biopython package offers critical insights into the virus's genetic composition and evolutionary dynamics. The study underscores the utility of advanced bioinformatics tools, such as multiple sequence alignment and phylogenetic tree construction, in unraveling the complexities of viral evolution. Through meticulous analysis, the study highlights several salient findings, including the notable GC content of 33.03% in the viral genome and the prevalence of leucine in the viral proteome. Furthermore, the phylogenetic investigation identifies key strains closely related to the reference sequence AAU01367.1, alongside significant evolutionary divergences, such as those observed in WZB41963.1 with the longest branch length, suggesting potential mutational hotspots or unique evolutionary pressures.

Moreover, the identification of the conserved domain PHA02782 in the AAU01367.1 protein, along with its 99.80% sequence similarity to UKV32022.1, offers valuable insight into the structural and functional conservation across MPXV strains. However, the absence of a three-dimensional protein structure for the AAU01367.1 protein emphasizes a critical gap in the current understanding of the virus's molecular architecture, which could further elucidate its mechanisms of pathogenicity and immune evasion.

Scopes for Future Research

The findings from this study lay a robust foundation for future research endeavors in the field of Monkeypox virology. Several key areas of investigation warrant immediate attention:

1. **Three-Dimensional Structural Elucidation**: Given the absence of a resolved 3D structure for key viral proteins, particularly DNA ligase-2, future studies should prioritize experimental and computational methods, such as X-ray crystallography or cryo-electron microscopy, in

conjunction with advanced molecular modeling techniques. A precise 3D structure will provide deeper insights into the functional domains of the virus, enhancing drug target identification and vaccine design.

2. **Viral Evolution and Mutational Analysis**: With the identification of strains exhibiting significant branch lengths, such as WZB41963.1, future genomic studies should focus on mutational analyses to discern whether these evolutionary divergences correlate with changes in virulence, transmission efficiency, or immune escape. Whole-genome sequencing of emerging strains could further reveal the ongoing evolutionary pressures acting on the MPXV genome.
3. **Vaccine Efficacy and Immunogenicity**: Given the demonstrated effectiveness of smallpox vaccination in conferring protection against Monkeypox, future research should explore the long-term immunogenicity of existing orthopoxvirus vaccines, especially in the face of waning immunity in global populations. Additionally, the development of next-generation vaccines specifically targeting conserved domains within the MPXV proteome could be a promising avenue for preemptive pandemic preparedness.
4. **Global Surveillance and Zoonotic Reservoirs**: As the virus continues to spread, a comprehensive global surveillance network should be established to monitor viral transmission patterns and potential reservoirs of zoonotic origin. This will aid in identifying emerging strains that may present heightened risks to public health and facilitate the development of region-specific containment strategies.
5. **Therapeutic Development**: Finally, given the urgent need for antiviral treatments, further research should focus on high-throughput screening of small molecules and biologics that target essential viral proteins, such as DNA ligase-2. In silico drug discovery pipelines, integrated with Biopython, can accelerate the identification of lead compounds with the potential to inhibit viral replication.

REFERENCES:

1. Giovanni Rezza, 2019 Aug;19(8):797-799. Emergence of human monkeypox in west Africa [doi: 10.1016/S1473-3099(19)30281-6].
2. PLoS Negl Trop Dis. 2022 Feb; 16(2): e0010141.Published online 2022 Feb 11. The changing epidemiology of human monkeypox—A potential threat? A systematic review. [doi: 10.1371/journal.pntd.0010141].
3. Human Monkeypox Reserch by MDPI.
4. .E. Mathieu, F. Spooner, S. Dattani, H. Ritchie, M. Roser. Mpox (monkeypox) (2022) Published online at OurWorldInData.org. Retrieved from:[https://ourworldindata.org/monkeypox].
5. Nikola Sklenovská 1, Marc Van Ranst 1 [1], Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans. [doi: 10.3389/fpubh.2018.00241.
6. Anne W. Rimoin arimoin@ucla.edu, Prime M. Mulembakani, Sara C. JohnstonMajor increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo. [doi: 10.1073/pnas.1005769107].
7. Brett W Petersen 1, Joelle Kabamba 2, Andrea M McCollum 3, Robert Shongo Lushima 4, Emile Okitolonda Wemakoy 5, Jean-Jacques Muyembe Tamfum 6, Beatrice Nguete 5, Christine M Hughes 3, Benjamin P Monroe 3, Mary G Reynolds 3Vaccinating against monkeypox in the Democratic Republic of the Congo [doi: 10.1016/j.antiviral.2018.11.004].
8. Anne W. Rimoin arimoin@ucla.edu, Prime M. Mulembakani, Sara C. Johnston, +15, and Jean-Jacques MuyembeMajor increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo. [doi: 10.1073/pnas.1005769107]
9. [doi: 10.1016/j.ijid.2022.05.058]
10. Sklenovská N. Monkeypox Virus. In: Malik Y.S., Singh R.K., Dhama K., editors. Animal-Origin Viral Zoonoses. Springer; Singapore: 2020. pp. 39–68.
11. [https://phil.cdc.gov]

Largest Protein is: -

>MDAAIRGNDVIFVLKTIGVPSACRQNEDPRFVEAFKCDELERYIDNNPECTLFESLRDEEAYSIVRIFMDVDLDACLDEIDYLTAIQDFIIEVSN
CVARFAFTECGAIHENVIKSMRSNFSLTKSTNRDKTSFHIIFLDTYTTMDTLIAMKRTLLELSRSSENPLTRSIDTAVYRRKTTLRVVGTRKNPN
CDTIHVMQPPHDNIEDYLFTYVDMNNNSYYFSLQRRLEDLVPDKLWEPGFISFEDAIKRVSKIFINSIINFNDLDENNFTTVPLVIDYVTPCALC
KKRSHKHPHQLSLENGAIRIYKTGNPHSCKVKIVPLDGNKLFNIAQRILDTNSVLLTERGDHIVWINNSWKFNSEEPLITKLILSIRHQLPKEYSS
ELLCPRKRKTVEANIRDMLVDSVETDTYPDKLPFKNGVLDLVDGMFYSGDDAKKYTCTVSTGFKFDDTKFVEDSPEMEELMNIINDIQPLTD
ENKKNRELYEKTLSSCLCGATKGCLTFFFGETATGKSTTKRLLKSAIGDLFVETGQTILTDVLDKGPNPFIANMHLKRSVFCSELPDFACSGSK
KIRSDNIKKLTEPCVIGRPCFSNKINNRNHATIIIDTNYKPVFDRIDNALMRRIAVVRFRTHFSQPSGREAAENNDAYDKVKLLDEGLDGKIQNN
RYRFAFLYLLVKWYKKYHIPIMKLYPTPEEIPDFAFYLKIGTLLVSSSVKHIPLMTDLSKKGYILYDNVVTLPLTTFQQKISKYFNSRLFGHDIES
FINRHKKFANVSDEYLQYIFIEDISSP

While Second Largest: -

>MNTGIIDLFDNHVDSIPTILPHQLATLDYLVRTIIDENRSVLLFHIMGSGKTIIALLFALVASRFKKVYILVPNINILKIFNYNMGVAMNLFNDEF
IAENIFIHSTTSFYSLNYNDNVINYNGLSRYNNSIFIVDEAHNIFGNNTGELMTVIKNKNKIPFLLLSGSPITNTPNTLGHIIDLMSEETIDFGEIISR
GKKVIQTLLNERGVNVLKDLLKGRISYYEMPDKDLPTIRYHGRKFLDTRVVYCHMSKLQEKDYMITRRQLCYHEMFDKNMYNVSMAVLGQ
LNLMNNLDTLFQEQDKELYPNLKINNGVLYGEELVTLNISSKFKYFINRIQTLKGKHFIYFSNSTYGGLVIKYIMLSNGYSEYNGSQGTNPHMI
NGKPKTFAIVTSKMKSSLEDLLDVYNSPENDDGNQLMFLFSSNIMSESYTLKEVRHIWFMTIPDTFSQYNQILGRSIRKFSYVDISEPVNVYLLA
AVYSDFNDEVTSLNDYTQDELINVLPFDIKKLLYLKFKTKETNRIYSILQEMSETYSLPPHPSIVKVLLGELVRQFFYNNSRIKYNDSKLLKMVT
SVIKNKEDARNYIDDIVNGHFFVSNKVFDKSLLYKYKNDIITVPFRLSYEPFVWGVNFRKEYNVVSSP

PROTEIN USED FOR PHYLOGENETIC ANALYSIS :-

MPXV-WRAIR157 [Monkeypox virus]

GenBank: AAU01367.1

>AAU01367.1 MPXV-WRAIR157 [Monkeypox virus]

MAMFYAHAFGGYDENLHAFPGISSTVANDVRKYSVVSVYNKKYNIVKNKYMWCNSQVNKRYIGALLPMFE

CNEYLQIGDPIHDLEGNQISIVTYRHKNYYALSGIGYESLDLCLEGVGIHHHVLETGNAVYGKVQHEYST

IKEKAKEMNALKPGPIIDYHVWIGDCVCQVTTVDVHGKEIMRMRFKRGAVLPIPNLVKVKVGEENDTINL

STSISALLNSGGGTIEVTSKEERVDYVLMKRLESIHHLWSVVYDHLNVVNGEERCYVHMHSSHQSPMLST

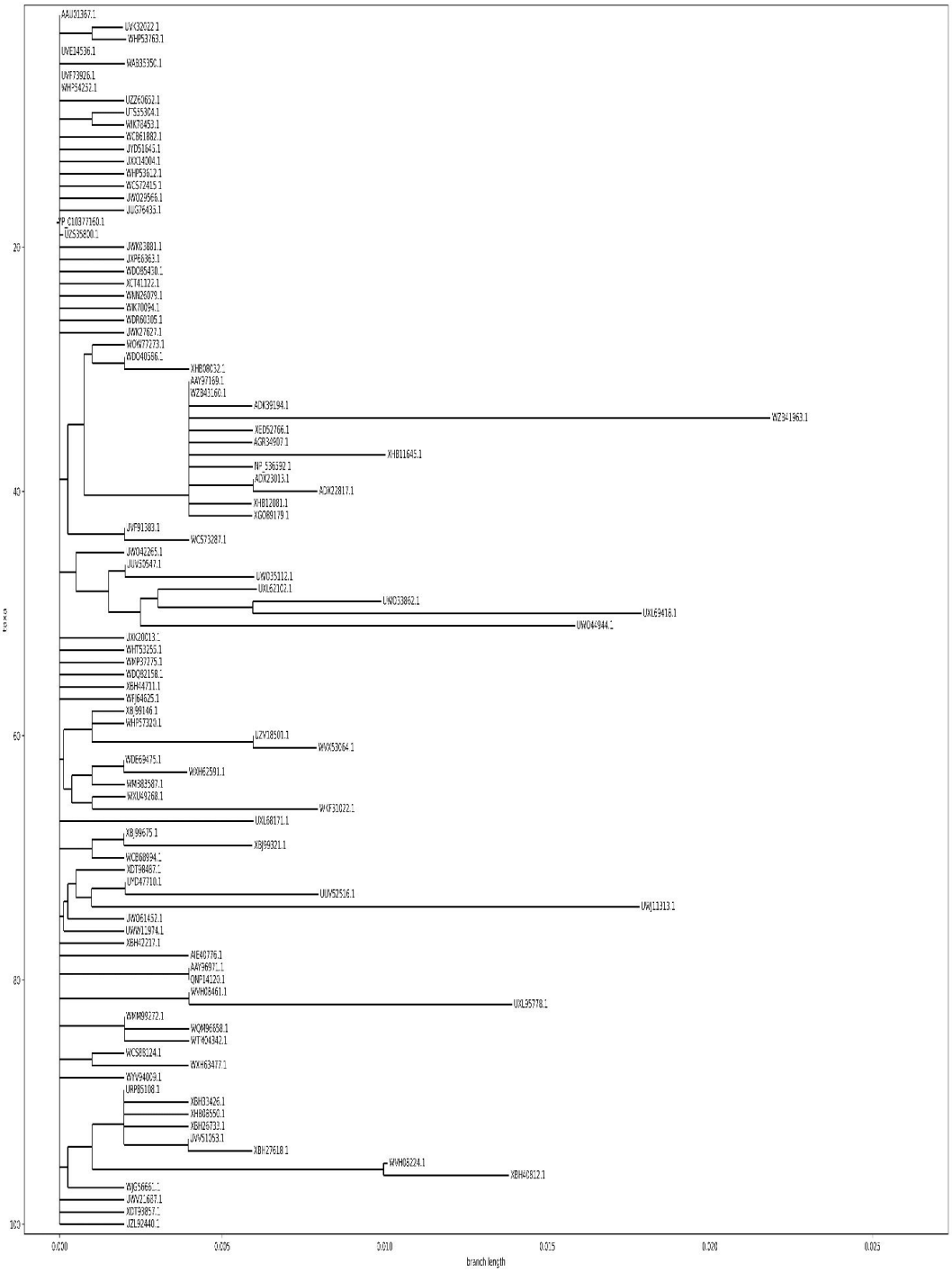
VKTNLYMKTMGACLQMDSMEALEYLSELKESGGRSPRPELQKFEYPDGVKDTESIERLAEEFFNRSELQA

GESVKFGNSINVKHTSVSAKQLRTRIRQQLPSILSSFANTKGGYLFIGVDNNTHKVIGFTVGHDYLKLVE

SDIEKYIQKLPVVHFCKKKEDIKYACRFIKVYKPGDETTSTYVCAIKVERCCCAVFADWPESWYMDTSGS

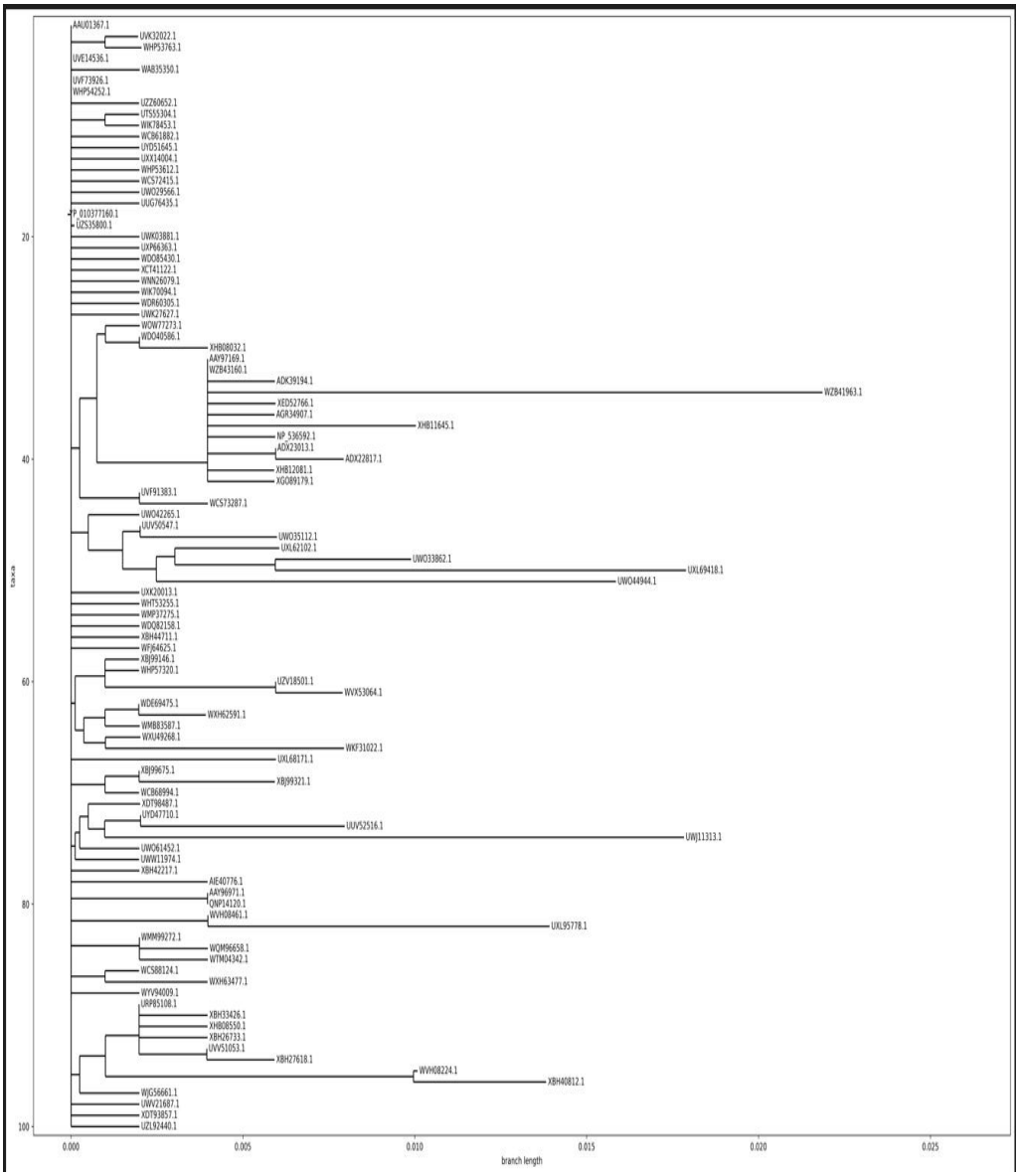
MKKYSPDEWVSHIKF

Figure 9: - Phylogram of Mpox Viral Protein



```
20178557     1 MAMFYAHAFGGYDENLHAFPGISSTVANDVRKYSVVSVYNNKYDIVKDKYMWCYSYVNKRYIGALLPMFECNEYLQIGDP 80
113195363    1 MAMFYAHALGGYDENLHAFPGISSTVANDVRKYSVVSVYNNKYDIVKDKYMWCYSQVNKRYIGALMPMFECNEYLQIGDP 80
18640413     1 MAMFYAHALGGYDENIHAFPGISSTVANDVRKYSVVSVYNNKYDIIKDKYMWCYSQVNKRYIGALLPMFECNEYLQIGDP 80
22164759     1 MAMFYTHAFGGYDENLHAFPGISSTVANDVRKYSVVSVYNKKYKIVKNKYMWCYSQVNKRYIGALLPMFECNEYLQIGDP 80
17975078     1 --MFYAHAFGGYDENLHAFPRISSTVANDVRKYSVVSVYNKKYNIVKNKYMWCNSQVNKRYIGALLPMFECNEYLQIGDP 78

20178557    81 IHDLEGNQISIVTYRHKNYYALSGIGYESLDLCLEGVGIHHHTLEAGNAVYGKVQHDYSTIKEKAKEMNSLSPGPIIDYH 160
113195363   81 IHDQEGNQISIVTYRHKNYYALSGIGYESLDLCLEGVGIHHHVLETGNAVYGKVQHDYSTIKEKAKEMSALSPGPIIDYH 160
18640413    81 IHDQEGNQISIVTYRHKNYYALSGIGYESLDLCLEGVGIHHHVLETGNAVYGKVQHDYSTIKEKAKEMSTLSPGPIIDYH 160
22164759    81 IHDQEGNQISIITYRHKNYYALSGIGYESLDLCLEGVGIHHHVLETGNAVYGKVQHDYSTIKEKAKEMSALSPGPIIDYH 160
17975078    79 IHDLEGNQISIVTYRHKNYYALSGIGYESLDLCLEGVGIHHHVLETGNAVYGKVQHEYSTIKEKAKEMNALKPGPIIDYH 158

20178557   161 VWIGDCVCQVTAVDVHGKEIMRMRFKKGAVLPIPNLVKVKLGeENDTVNLSTSISALLNSGGGTIEVTSKEERVDYVLMK 240
113195363  161 VWIGDCICQVTAVDVHGKEIMRMRFKKGAVLQIPNLVKVKLG-ENDTENLSTTISALLNSGGGTIEVTSQEDRVNHVLMK 239
18640413   161 VWIGDCICQVTAVDVHGKEIMRMRFKKGAVLQIPNLVKVKLG-ENDSENLSTTISALLNSGGGTIEVTSQEDRVNHVLMK 239
22164759   161 VWIGDCICQVTAVDVHGKEIMRMRFKKGAVLQIPNLVKVKLG-ENDTENLSTAISALLNSGGGTIEVTSKEERVDYVLMK 239
17975078   159 VWIGDCVCQVTTVDVHGKEIMRMRFKRGAVLPIPNLVKVKVGeENDTINLSTSISALLNSGGGTIEVTSKEERVDYVLMK 238

20178557   241 RLESIRHLWSvVYDHFDVVNGKERCYVHMHSSNQSPMLSTVKTNLYMKTMGACLQMDYMEALEYLSELKESGGRSPRPEL 320
113195363  240 RLESIRHMWS-VYDRFNIVNGKECCYIHLHSSNQNLMPSTVKTNLYMRSMTSCIQMDPITALDYLSELKESGGQSPRPEL 318
18640413   240 RLESIRHMWS-VYDRFNIVNGKECCYIHLHSSNQNLMPSTVKTNLYMRSMTSCIQMDPITALYYLSELKESGGQSPRPEL 318
22164759   240 RLESIHHLWSvVYDHLNVVNGKERCYVHMHSSNQHPIPSTVKTNLYMKTMGSCLQMDSMEALEYLSELKESGGRSPRPEL 319
17975078   239 RLESIHHLWSvVYDHLNVVNGEERCYIHMHSSHQSPMLSTVKTNLYMKTMGACLQMDSMEALEYLSELKESGGRSPRPEL 318

20178557   321 PEFEYPDGVEDAGSIERLAEEFFSRSELQADEPVNFCNSINVKHTSVSAKQLRTRIRQQLPSILSSFANTDGGYLFIGVD 400
113195363  319 PEFDYPDGVQDDGSIERFAEEFFNRSELQAGESVEFGNSVNVKNTSVSAKQLRTRIRQQLPSILSSFANTEGGYLFIGLD 398
18640413   319 PEFDYPDGVQDDGSIERVAEEFFNRSELQAGESVEFGNSVNVKNTSVSAKQLRTRIRQQLPSILSSFANTEGGYLFIGLD 398
22164759   320 QKFEYPDGVEDTESIERLAEEFFNRSELQAGESVKFGNSINVKHTSVSAKQLRTRIRQQLPSILSSFANTEGGYLFIGVD 399
17975078   319 QKFEYPDGVKDTESIERLAEEFFNRSELQAGESVKFGNSINVKHTSVSAKQLRTRIRQQLPSILSSFANTKGGYLFIGVD 398

20178557   401 NNTHKVVGFTVGQDYLKLVESDIEKYIKRLRVVHFCEKKEDIKYACRFIKVYKPGEETTSTYVCAIKVERCCCAVFADWP 480
113195363  399 NITHKVIGFTVGQDCLKLIENEIEKRIRRLHVVHFCEKKEDIKYACRFIKVYKPGEETTSTYVCAIKVERCCCAVFTNWP 478
18640413   399 NNTHKVIGFTVGQDCLKLIENEIEKRIRRLHVVHFCEKKEDIKYACRFIKVYKPGEETTSTYVCAIKVERCCCAVFTDWP 478
22164759   400 NNTHKVIGFTVGHDYLRLVENDIEKHIKRLRVVHFCEKKEDIKYACRFIKVYKPGDETTSTYMCAIKVERCCCAVFADWP 479
17975078   399 NNTHKVIGFTVGHDYLKLVERDIEKYIQKLPVVHFCKKKEDIKYACRFIKVYKPGDETTSTYVCAIKVERCCCAVFADWP 478

20178557   481 ESWYMDTSGsMKKYSPDEWVSHIKF 505
113195363  479 ESWYMDTSG-IKKYSPDEWVSSIKF 502
18640413   479 ESWYMDTSG-IKKYSPDEWVSSIKF 502
22164759   480 ESWYMDTNG-IKKYSPDEWVSYIKV 503
17975078   479 ESWYMDTSGsMKKYSPDEWVSHIKF 503
```

Figure 10: Conserved Domain Database of NCBI Showing sequence's conserved regions.


AAU01367.1
UVK32022.1
WHP53763.1
UVE14536.1
WAB35350.1
UVF73926.1
WHP54252.1
UZZ60652.1
UTS55304.1
WIK78453.1
WCB61882.1
UYD51645.1
UXJ14004.1
WHP53612.1
WCS72415.1
UWQ29566.1
UUG76435.1
P_010377160.1
QZS35800.1
UWK03881.1
UXP66363.1
WDO85430.1
XCT41122.1
WNN26079.1
WIK70094.1
WDR60305.1
UWK27627.1
WQW77273.1
WDO40586.1
XHB08032.1
AAY97169.1
WZB43160.1
ADK39194.1
WZB41963.1
XED52766.1
AGR34907.1
XHB11645.1
NP_536592.1
ADY23013.1
ADX22817.1
XHB12081.1
AGO89179.1
UVF91383.1
WCS73287.1
UWO42265.1
UUV50547.1
UWO35112.1
UXL62102.1
UWO33862.1
UXL69418.1
UWO44944.1
UXK20013.1
WHT53255.1
WMP37275.1
WDQ82158.1
XBH44711.1
WPJ64625.1
XBJ99146.1
WHP57320.1
UZV18501.1
WVX53064.1
WDE69475.1
WXH62591.1
WMB83587.1
WXU49268.1
WKF31022.1
UXL68171.1
XBJ99675.1
XBJ99321.1
WCB68994.1
XDT98487.1
UYD47710.1
UUV52516.1
UWJ11313.1
UWO61452.1
UWW11974.1
XBH42217.1
AIE40776.1
AAY96971.1
QNP14120.1
WVH08461.1
UXL95778.1
WMM99272.1
WQM96658.1
WTM04342.1
WCS88124.1
WXH63477.1
WYV94009.1
URP85108.1
XBH33426.1
XHB08550.1
XBH26733.1
UVV53053.1
XBH27618.1
WVH08224.1
XBH40812.1
WQS6661.1
UWV21687.1
XDT93857.1
UZL92440.1
taxa
20
40
60
80
100
0.000
0.005
0.010
0.015
0.020
0.025
branch length