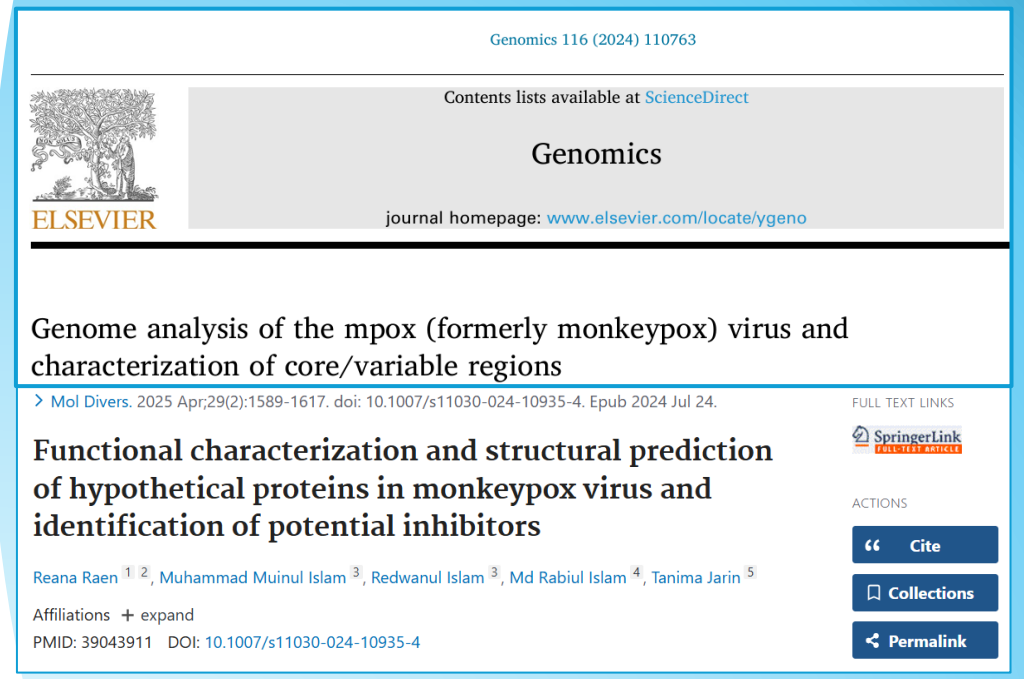
Medical Microbiology

คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
FACULTY OF MEDICINE KHON KAEN UNIVERSITY

Genomic analysis of Mpox, functional characterization of hypothetical proteins and identification of potential inhibitors



Genomics 116 (2024) 110763

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genome analysis of the mpox (formerly monkeypox) virus and characterization of core/variable regions

Mol Divers. 2025 Apr;29(2):1589-1617. doi: 10.1007/s11030-024-10935-4. Epub 2024 Jul 24.

Functional characterization and structural prediction of hypothetical proteins in monkeypox virus and identification of potential inhibitors

Reana Raen [1 2], Muhammad Muinul Islam [3], Redwanul Islam [3], Md Rabiul Islam [4], Tanima Jarin [5]

Affiliations + expand

PMID: 39043911 DOI: 10.1007/s11030-024-10935-4

FULL TEXT LINKS

SpringerLink FULL-TEXT ARTICLE

ACTIONS

Cite

Collections

Permalink

Zwe Win Paing

2nd Year PhD

07.01.2026

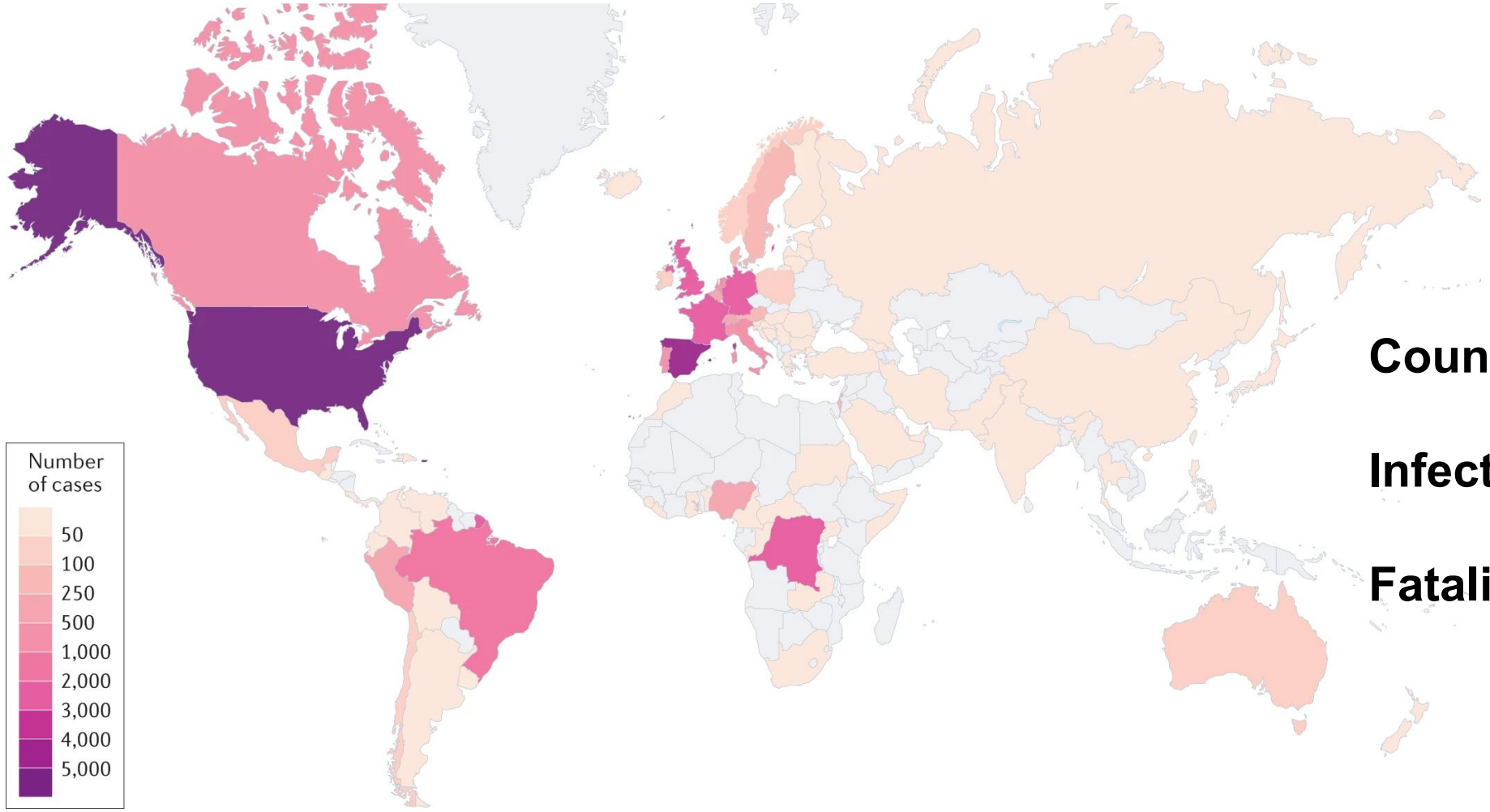
Introduction

One of the greatest achievement in **21 century** is **Smallpox** was declared **eradicated** by the WHO in 1980, however, it poses a threat to human as the disease could reappear due to accidental release of the viruses from repositories or intentional de novo synthesis of infectious viruses for bioterrorism.

Recent global transmission of endemic Mpox, along with cases of human-to-human transmission of poxviruses that belong in Orthopoxvirus, Molluscipoxvirus, and Parapoxvirus genera, indicates a potential for smallpox-like disease outbreaks.

Introduction

Geographical distribution of confirmed and suspected monkeypox cases during the outbreak between May and August 2022.



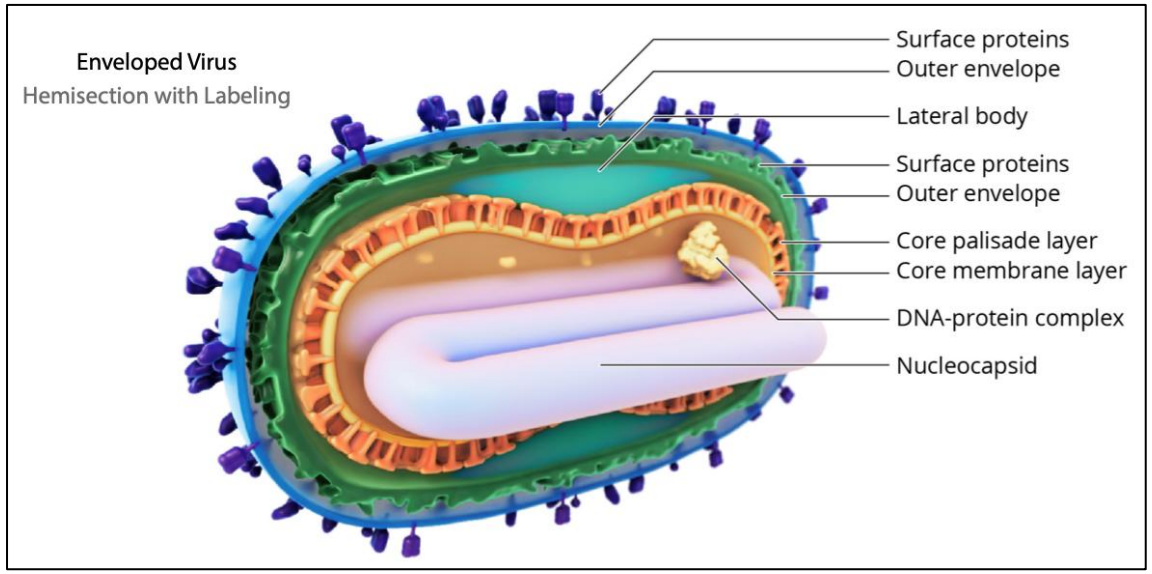
Country: >100

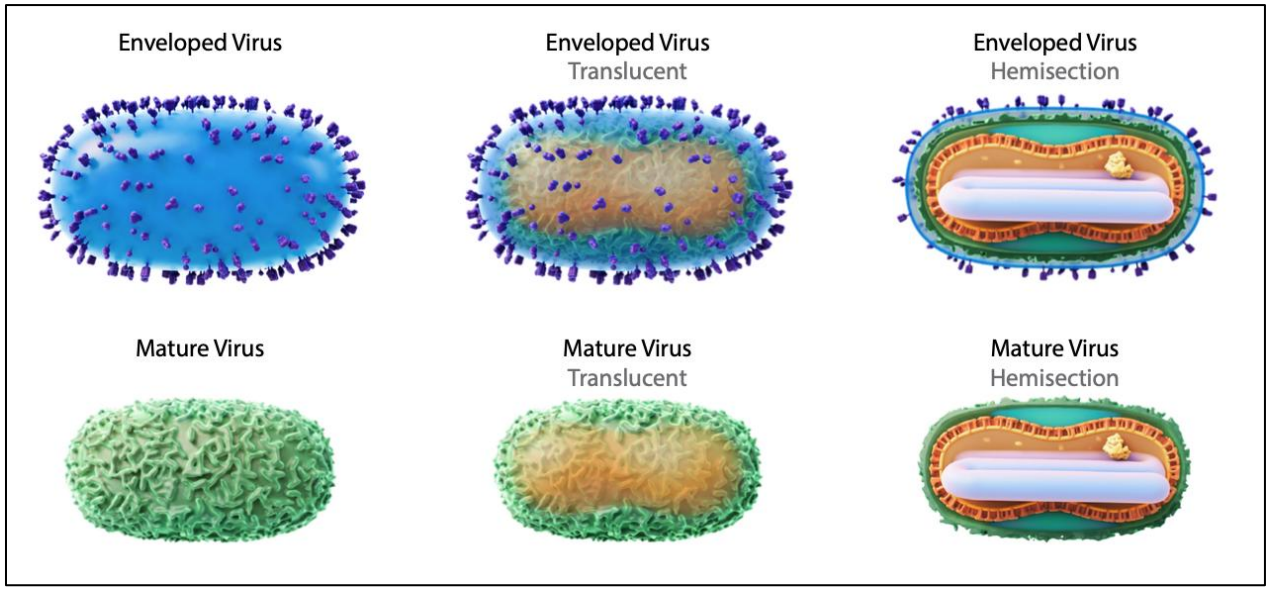
Infected people: >85,000

Fatality: ≈200

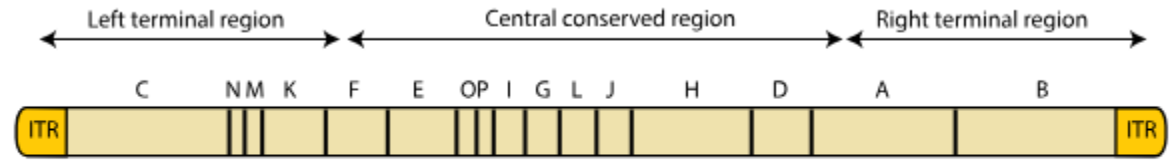
Lum, FM., *et al.* Monkeypox: disease epidemiology, host immunity and clinical interventions. *Nat Rev Immunol* **22**, 597–613 (2022).

Introduction





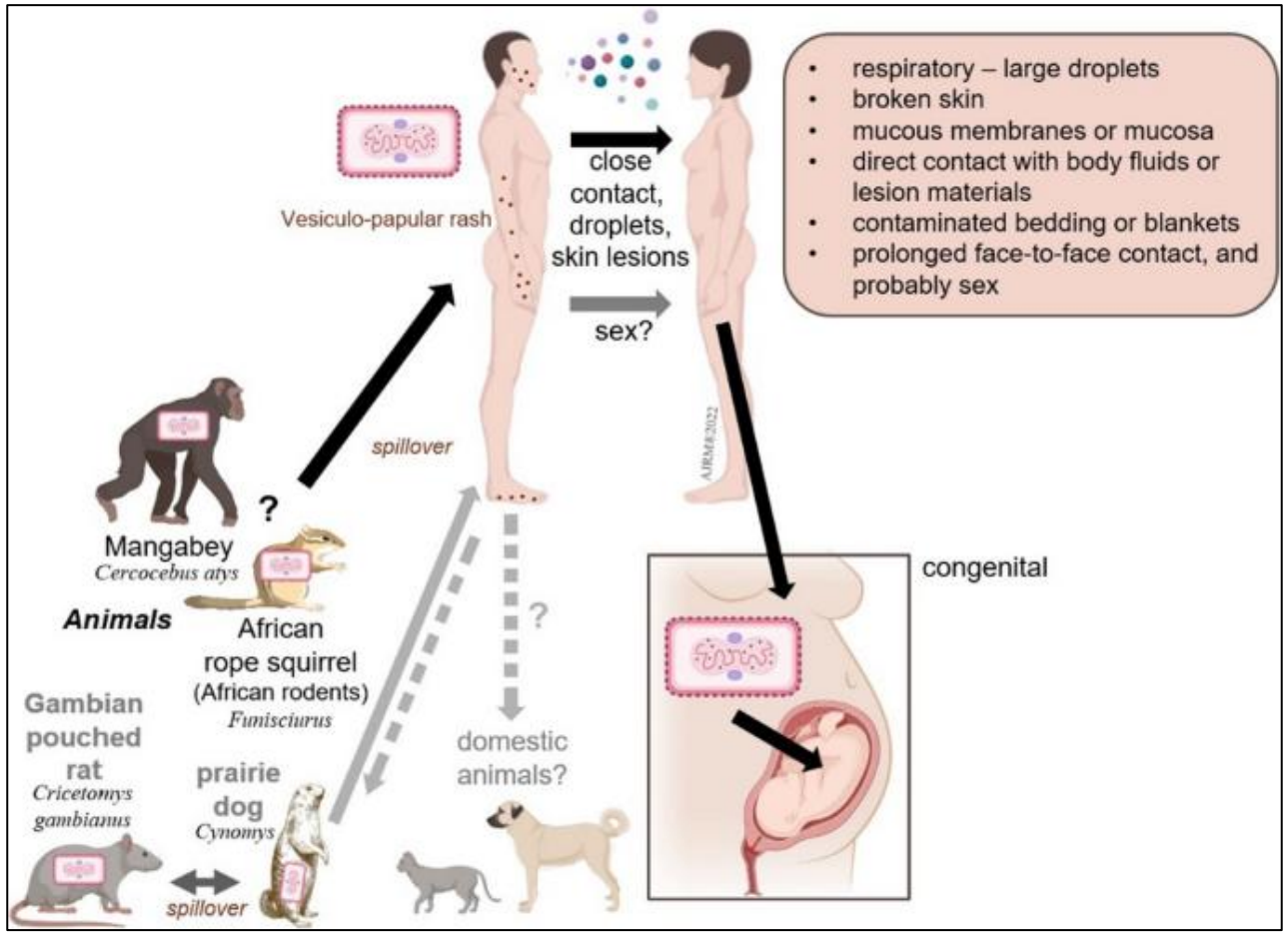
- Enveloped dsDNA virus
- Size: 220 Kb, 223 ORFs
- Genus: *Orthopoxvirus* genus
- Family: *Poxviridae*



ITR= Inverted terminal repeats

https://ictv.global/home

Introduction: Mpox transmission and outbreak



- Natural reservoir of mpox remains unidentified
- Transmitted to humans through rodents and non-human primates

1st outbreak

- 1970
- Democratic Republic of Congo (Central African countries)
- Clade I

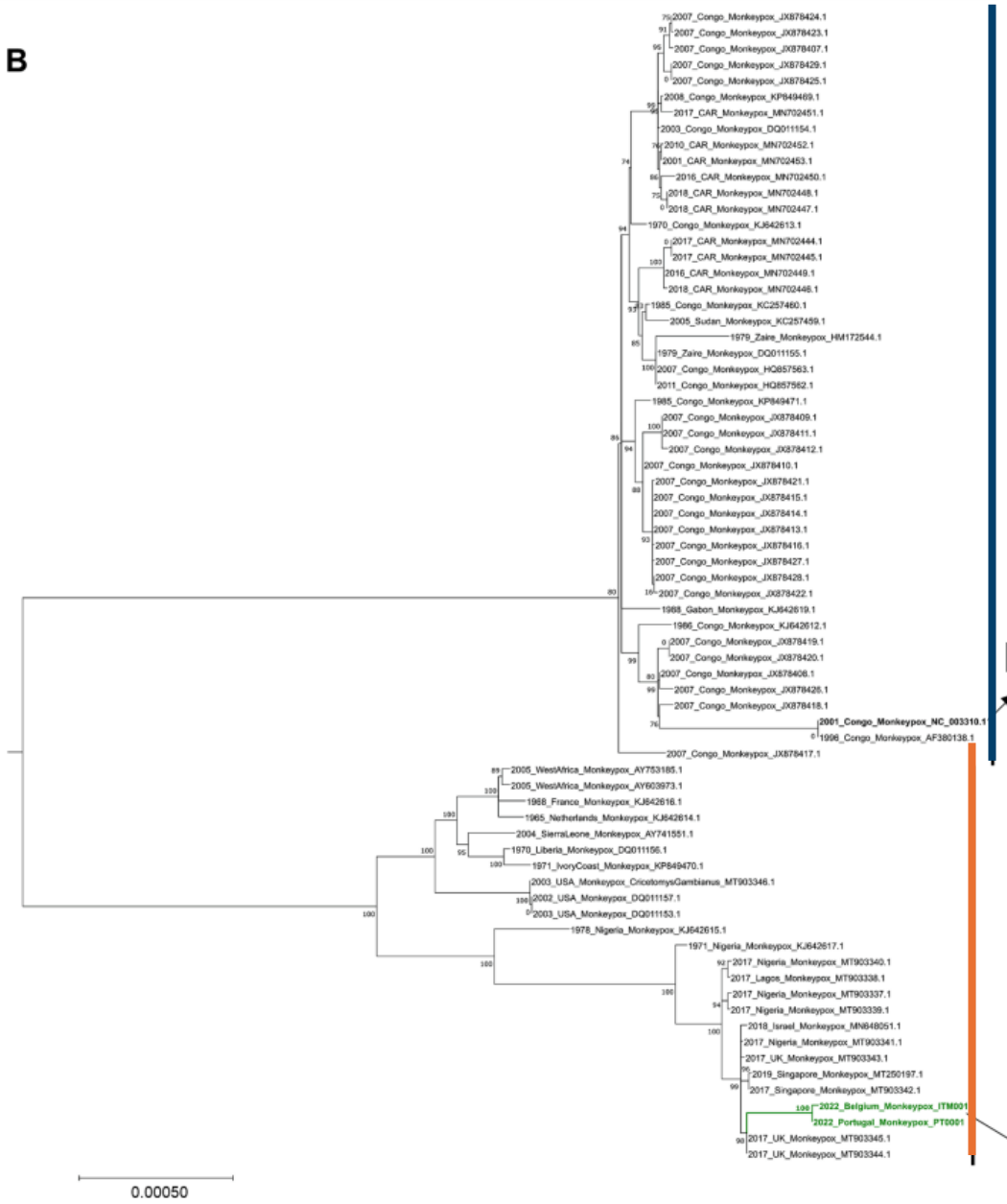
2nd outbreak

- 2003
- Western African countries
- Clade II

Current outbreak

- Clade III (hMpox-1 A, B.1, A.1.1, A.1, and A.2)

Introduction



Central Africa
Severe

Reference Genome

Western Africa
Non-severe

Recent outbreak

MOPICE, B14R

Clade I

Clade II

46 SNPs in coding regions

Clade III

Highly Transmission

Genomics 116 (2024) 110763

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genome analysis of the mpox (formerly monkeypox) virus and characterization of core/variable regions

Q2	Citation Score	Impact Factor
	8.1	3.0

Introduction

Rational of the study: To address the gaps in multinational outbreak, providing further insights into evolutionary trajectory, genetic diversity and phenotypic traits of the virus.

Hypothesis: Variable regions tend to have more mutations, gene changes, repeats, and recombination.

Objective: To identify and characterize the core/variable regions of the currently available Mpox virus whole genome sequences.

Methodology

CNCB

NCBI

Data Retrieving

Nextclade v3.13.1

QC

Prokka

Gene Annotation

NCVOGs

NCBI

Orthofinder v2.5.2

Core-pan analysis

Collinearity analysis

MCscan

Gubbins v3.2.1
TRF and IRF
SNP-sites

Recombination analysis
Repeats sequences detection
SNP analysis

QC and basic characteristics

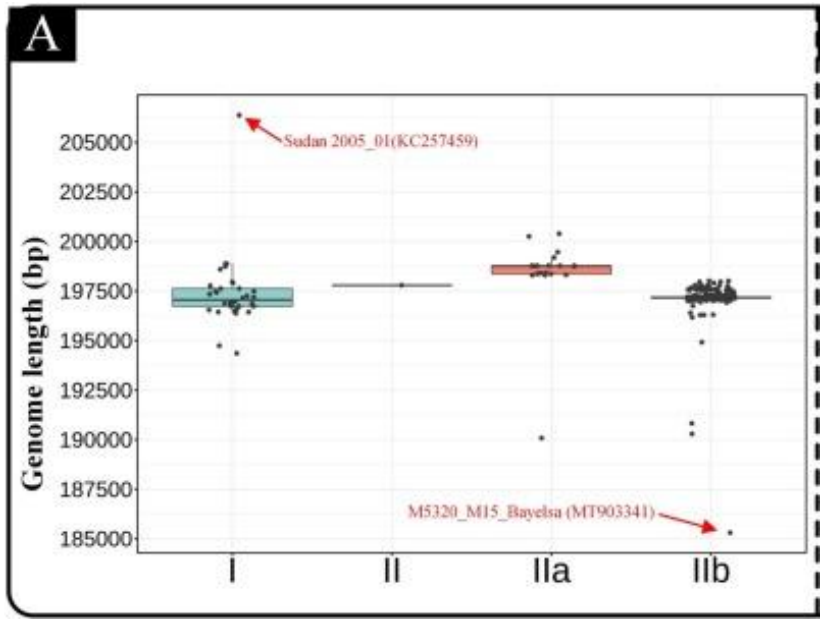
Objective: To assess the quality of the retrieved whole genome sequences

Total WGS: 853

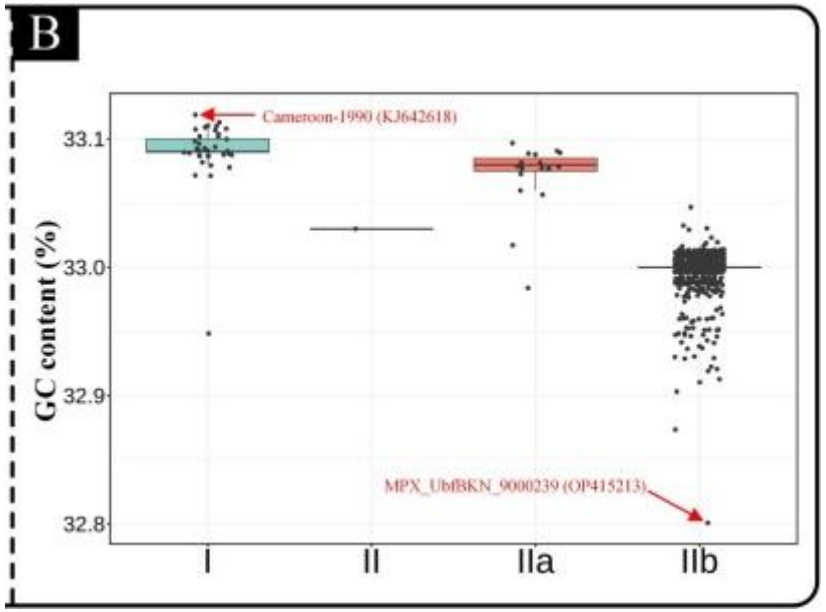
Selected WGS: 825

Genome length ranged from 185,309 bp to 206,372 bp

GC content ranged from 32.80% to 33.12%



10.76%



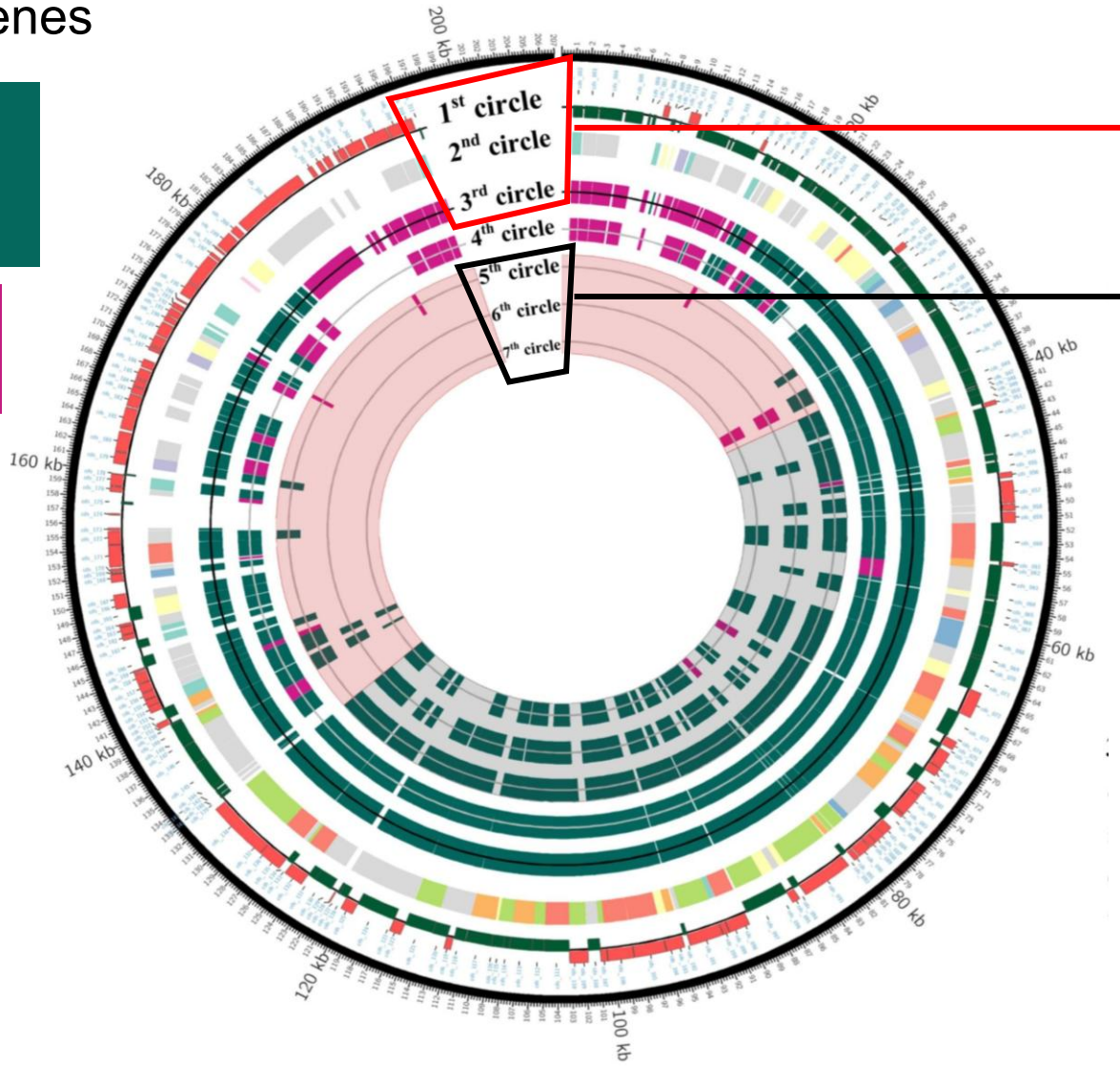
0.97%

Core-pan analysis

Objective: To see Mpox specific genes

Conserved regions: DNA replication, recombination, and repair, as well as virion structure and morphogenesis

Variable regions: Host-virus interaction



1st circle: Coding sequences (cds) of USA_2022_MA001

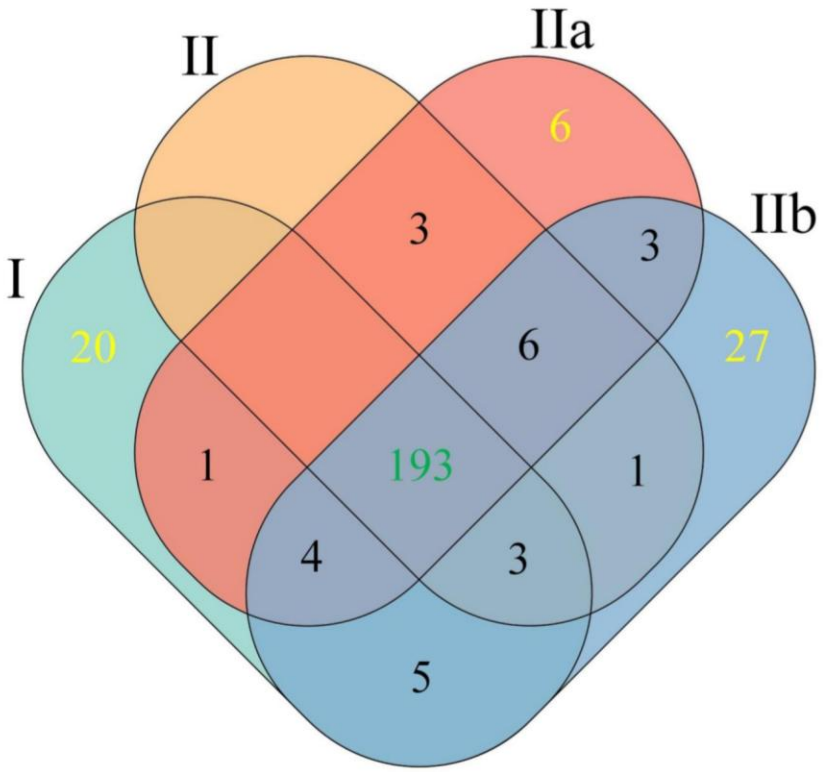
- Sense strand
- Antisense strand

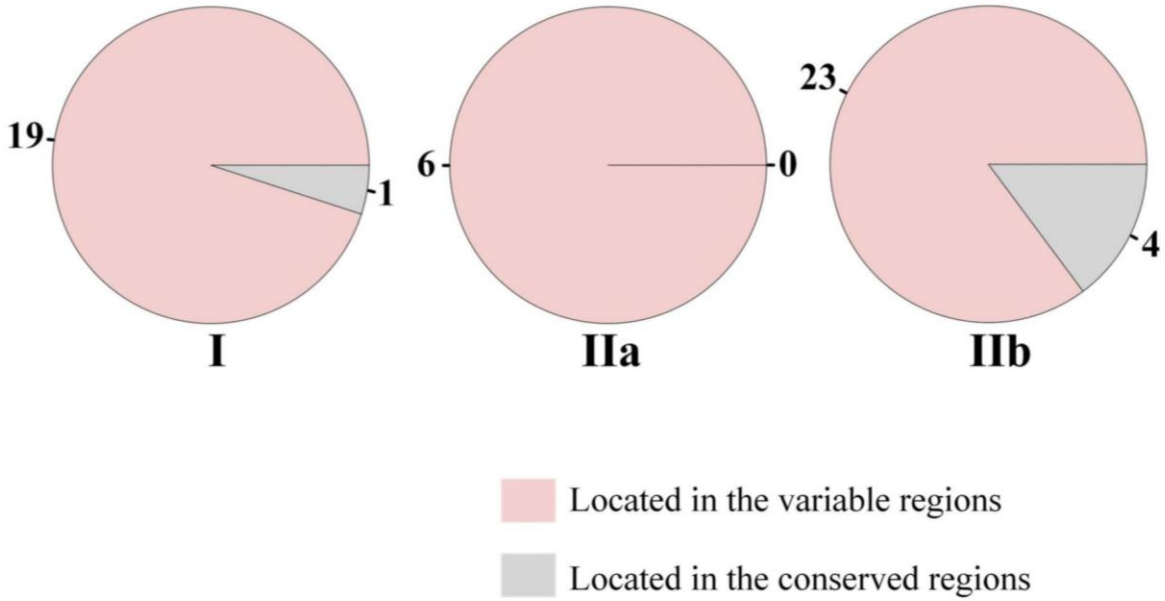
2nd circle: functional categories

- Host-virus interactions
- Other metabolic functions
- Signal transduction regulation
- DNA replication, recombination and repair
- Nucleotide metabolism
- Virion structure and morphogenesis
- Transcription and RNA processing
- Mobile elements
- Uncharacterized

Ref Seq: USA_2022_MA001

Core-pan analysis





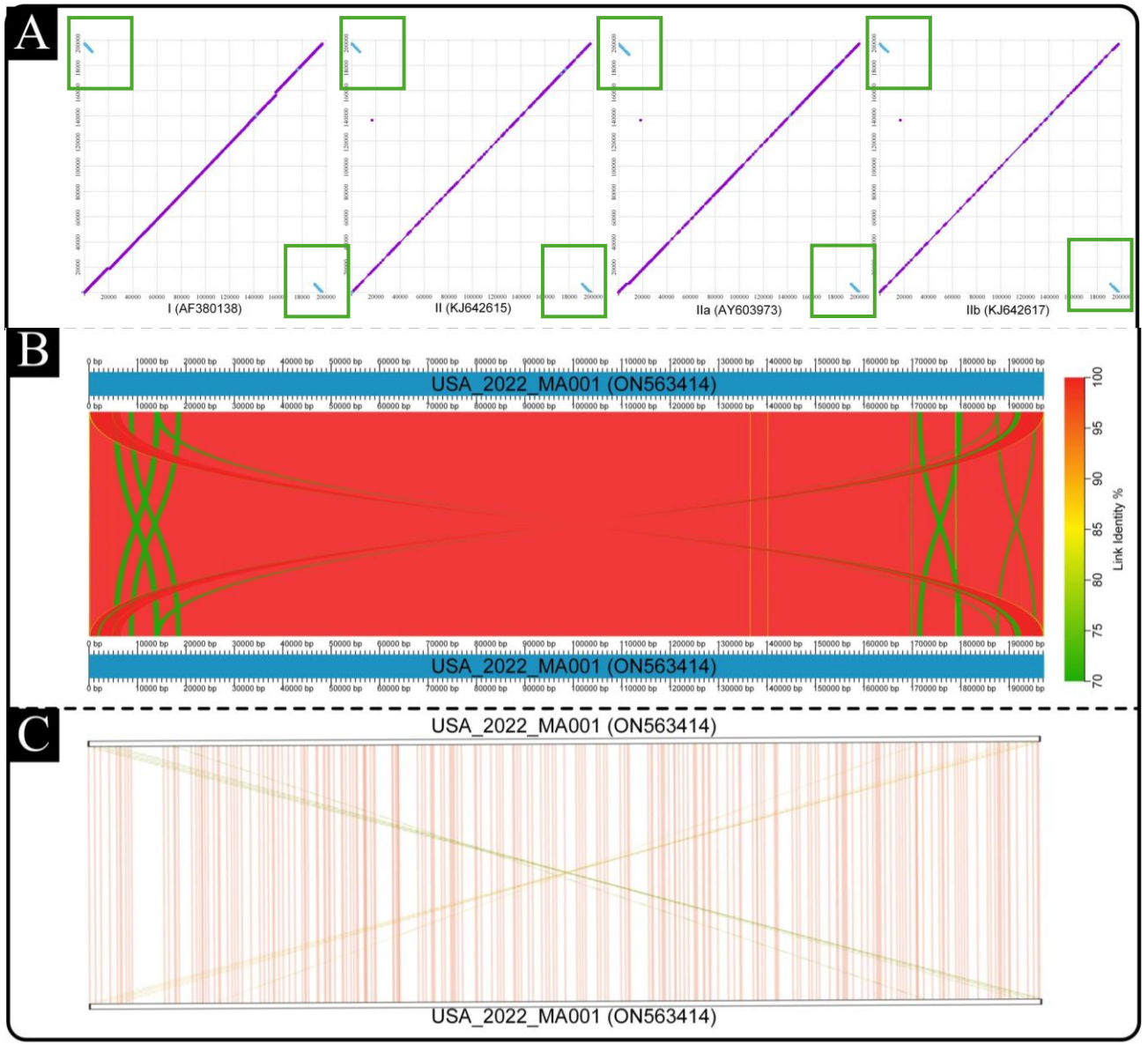
The result showed that multi-copy and specific genes are generally located in the variable regions

Collinearity analysis

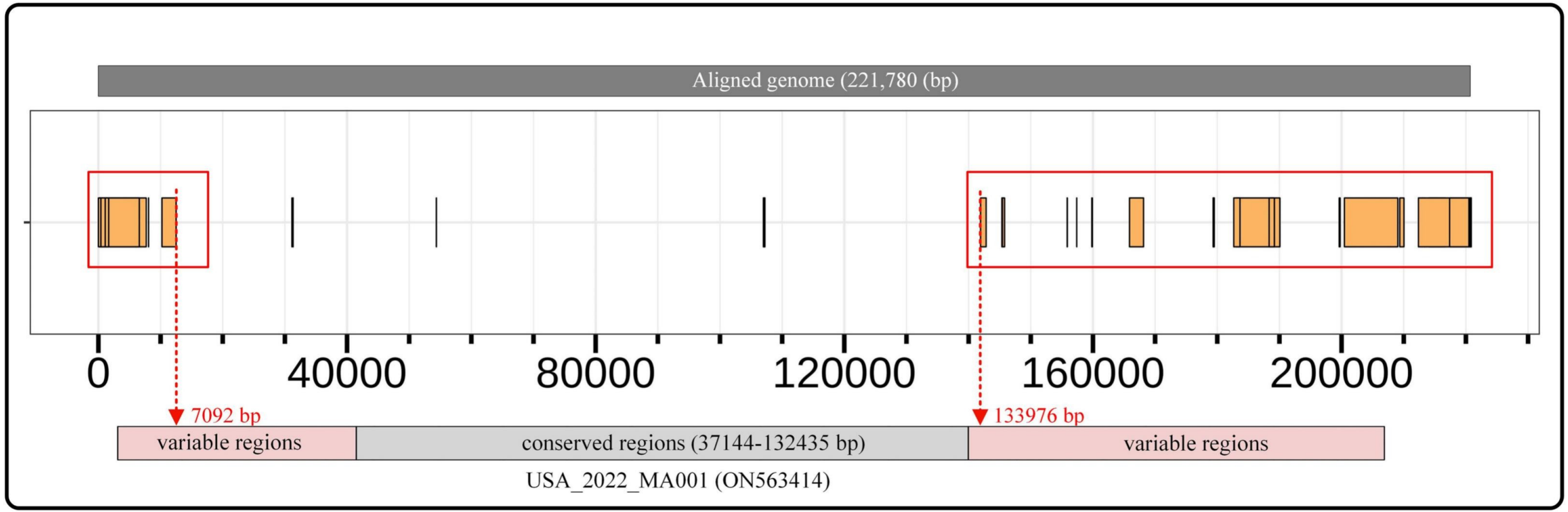
Purple: Well-conserved genome collinearity between mpox interspecies

Blue: MPXV genome contains ITRs

The **ITRs** also lead to homology between the coding gene sequences at the beginning of the genome and the gene sequences at the end

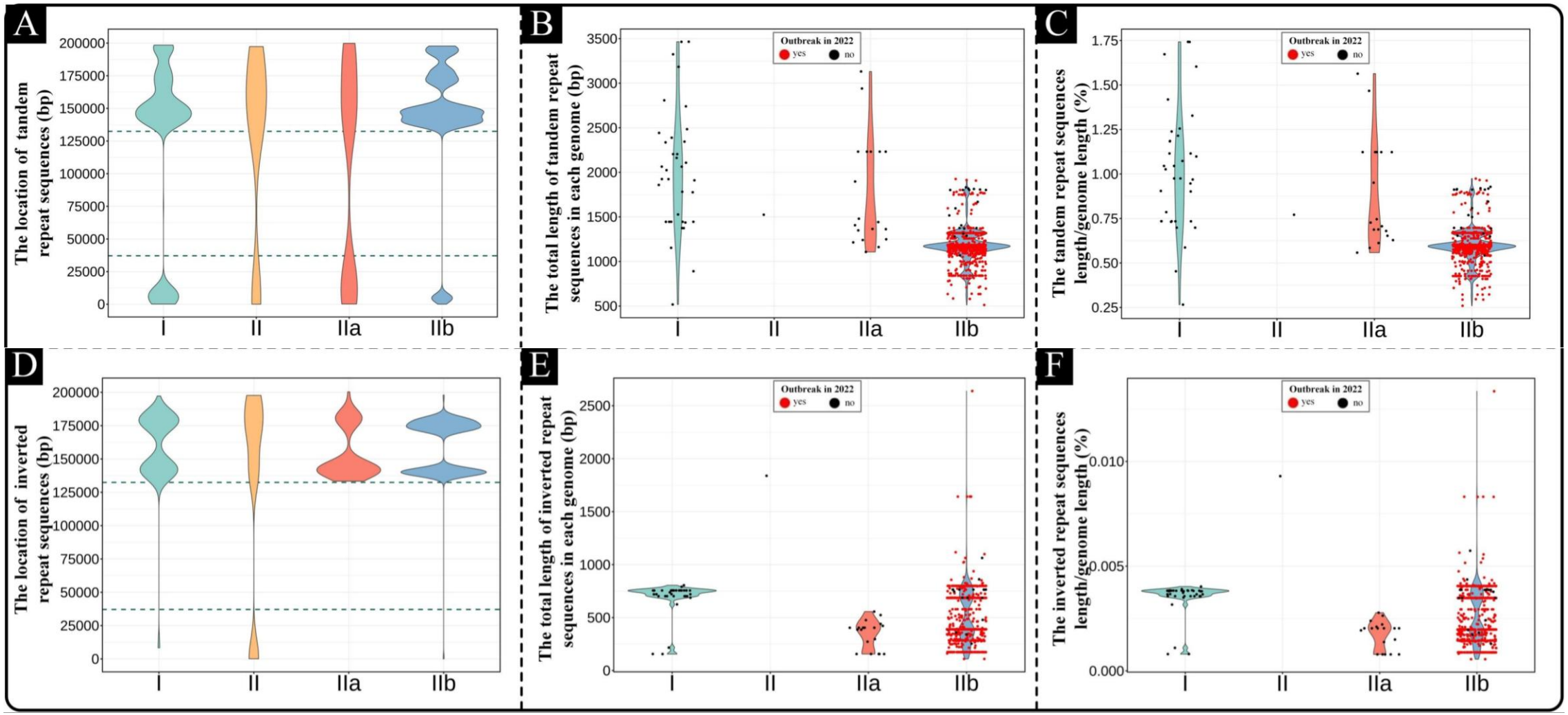


Recombination analysis



Most of the recombination events occurred in the variable regions

Repeat sequences analysis



Length of tandem repeat of MPXV within the IIb clade is overall lower than those in the other three clades

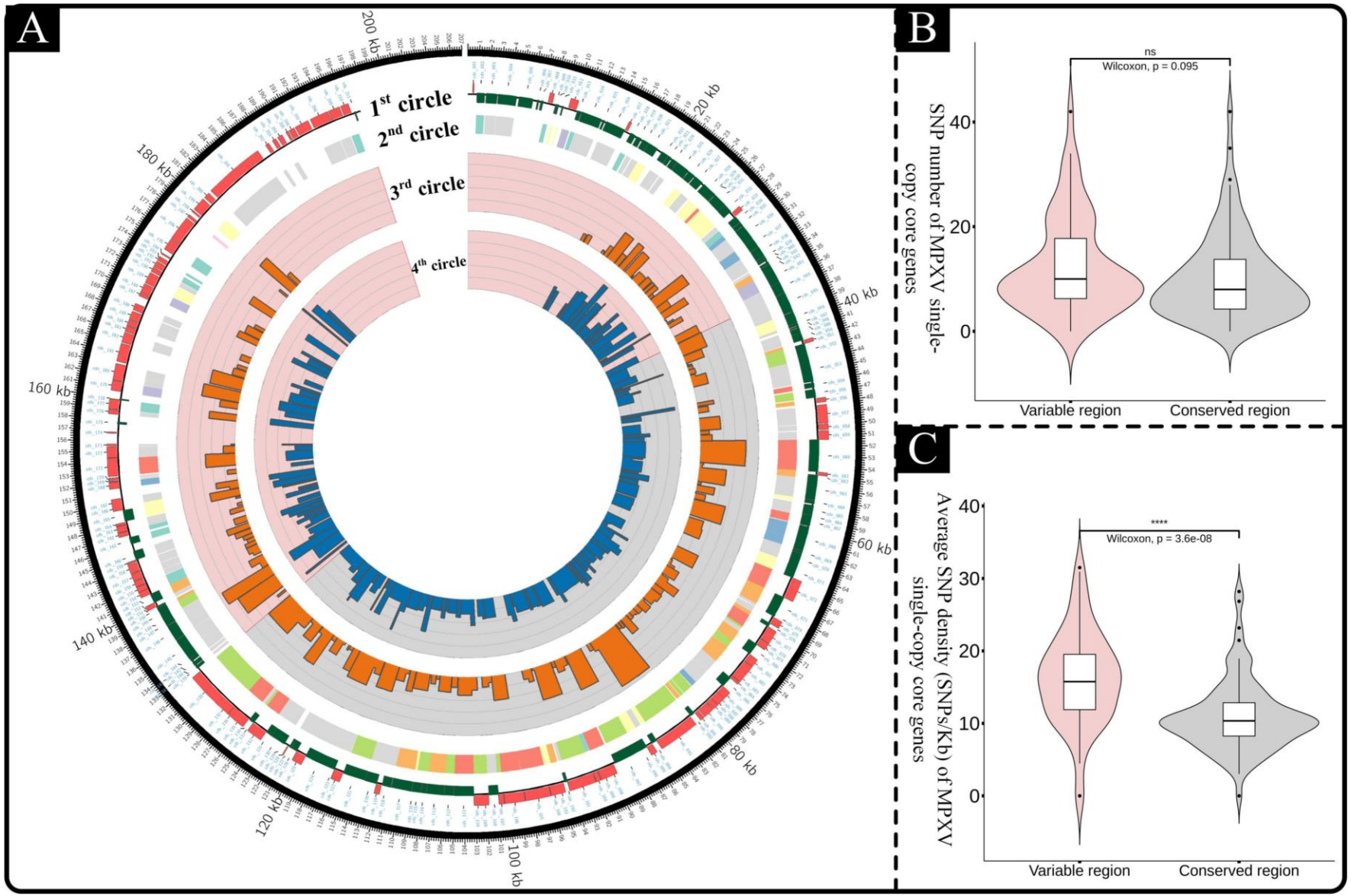
SNP detection

Average **SNP** density in variable regions is higher than in core regions

The average **SNP** density in variable regions **(16.16 ± 6.15 SNPs/Kb)** was significantly higher than that in core regions

The gene with the highest **SNP** density was Kelch repeat and BTB domain-containing protein 2 (**KBTBD2**)

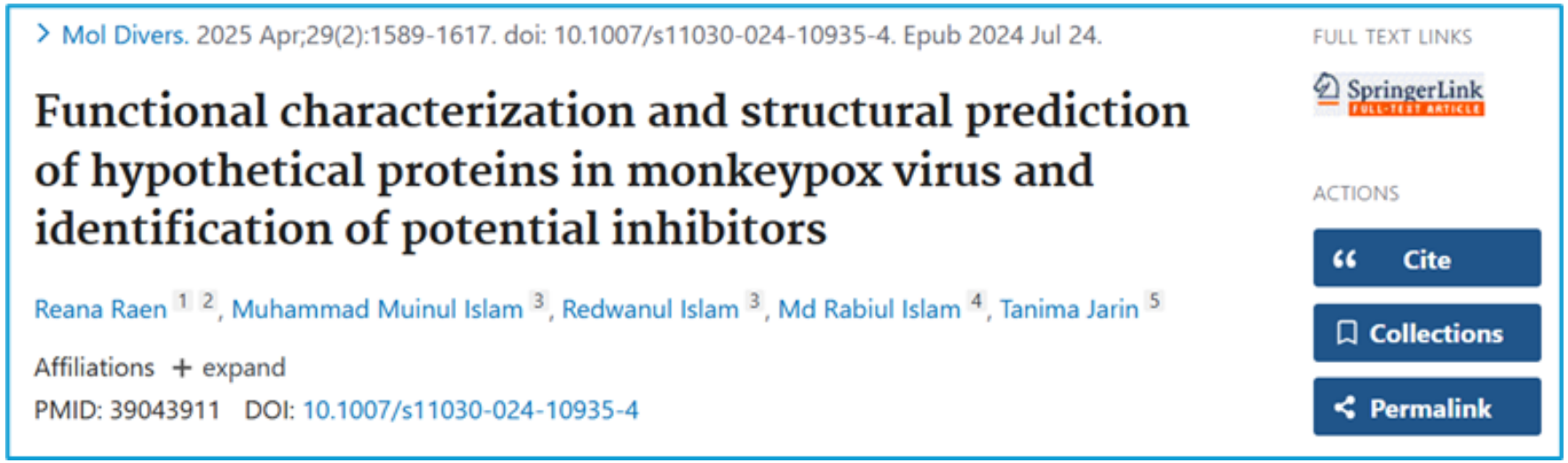
Among all viruses, only the genomes of poxviruses contain genes for **kelch-like proteins**



Conclusion

- The mpox genome's core and variable regions were clearly delineated, with **recombination and repeat** sequences concentrated in **variable regions**.
- Single nucleotide polymorphisms **(SNPs)** are predominantly found in the **variable regions**, indicating higher genetic diversity there.
- **High SNP density** in the **KBTBD2** gene may relate to host adaptation, warranting further experimental investigation.

2nd Paper



Mol Divers. 2025 Apr;29(2):1589-1617. doi: 10.1007/s11030-024-10935-4. Epub 2024 Jul 24.

Functional characterization and structural prediction of hypothetical proteins in monkeypox virus and identification of potential inhibitors

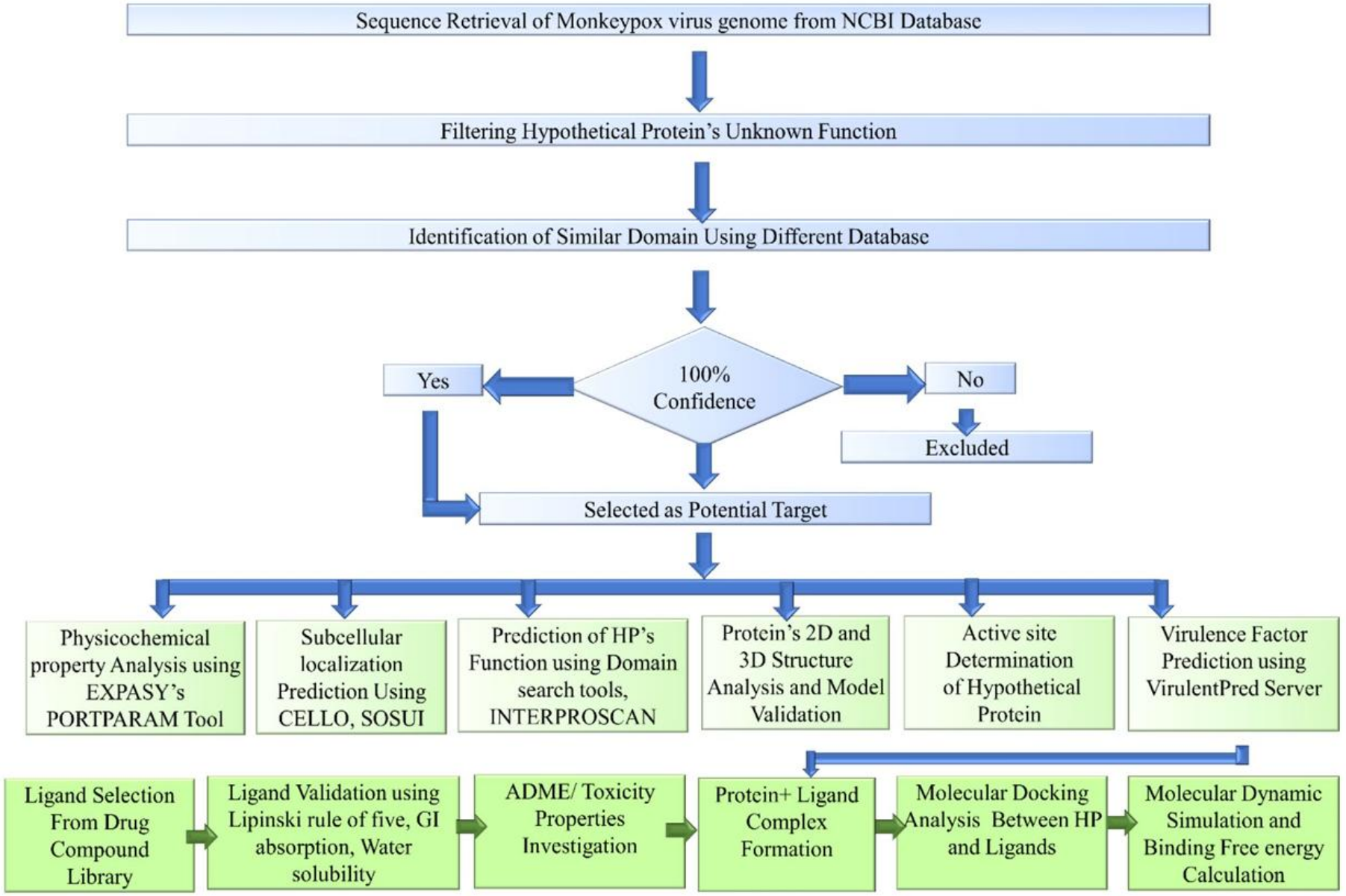
Reana Raen [1 2], Muhammad Muinul Islam [3], Redwanul Islam [3], Md Rabiul Islam [4], Tanima Jarin [5]

Affiliations + expand

PMID: 39043911 DOI: 10.1007/s11030-024-10935-4

Q2	Citation Score	Impact Factor
	4.8	3.9

Methodology



Properties of Hypothetical Protein

Objective: To retrieve the hypothetical proteins and analyze the features of selected HP

GCF 000857045.1 was retrieved from NCBI database
191 proteins in the entire MPXV virus proteome including **28 HPs**

Pfam, CDD-BLAST, ScanProsite, and SMART
28 HPs were categorized into five group.

7 HPs are in high confidence level

AIE40426.1, MPXV-Congo_8-156 were in high confidence level

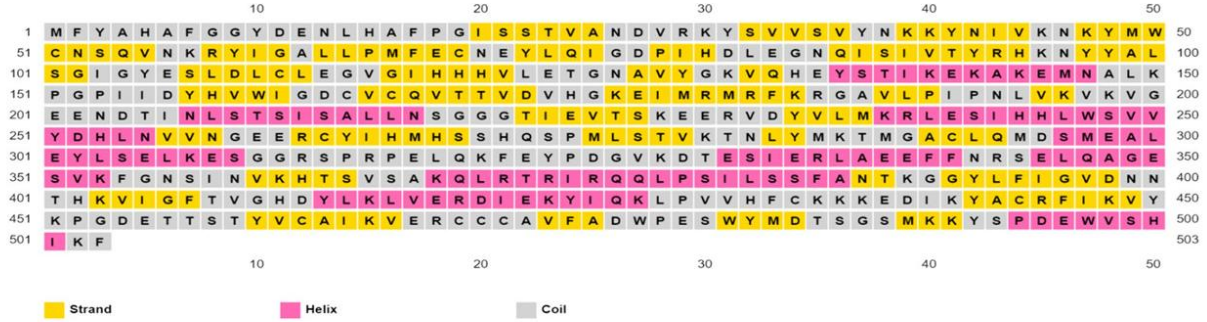
The CELLO program estimated the target protein's subcellular location as "Cytoplasmic"

Table 1 Specific physiochemical properties of the selected Hypothetical protein (MPXV-Congo_8-156)

Characteristics	Finding	Remark
Number of amino acids	503	Suitable
Molecular weight	57.35 kDa	Average
Theoretical pI	7.29	Slightly Basic
Chemical formula	$C_{2562}H_{3983}N_{687}O_{753}S_{28}$	–
Extinction coefficient (at 280 nm in H_2O)	76,960	–
Estimated half-life (mammalian reticulocytes)	30 h	–
Estimated half-life (yeast cell, in vivo)	> 20 h	–
Estimated half-life (E. coli, in vivo)	> 10 h	–
Instability index	39.01	Stable
Stable aliphatic index	82.27	Thermostable
Grand average hydropathicity (GRAVY)	− 0.350	Hydrophilic

Protein structure analysis

Objective: To predict the secondary and tertiary structure of HP



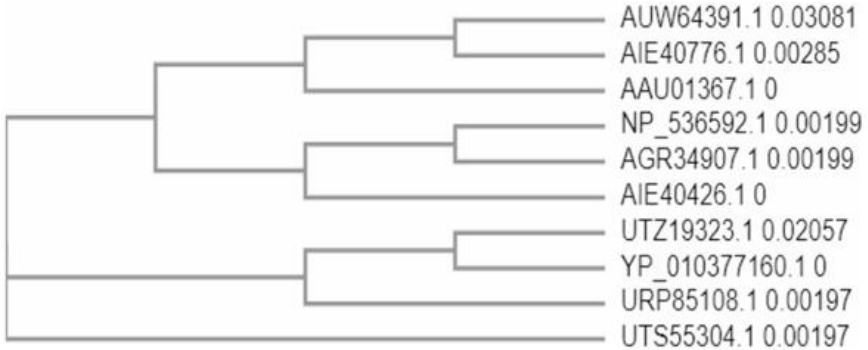
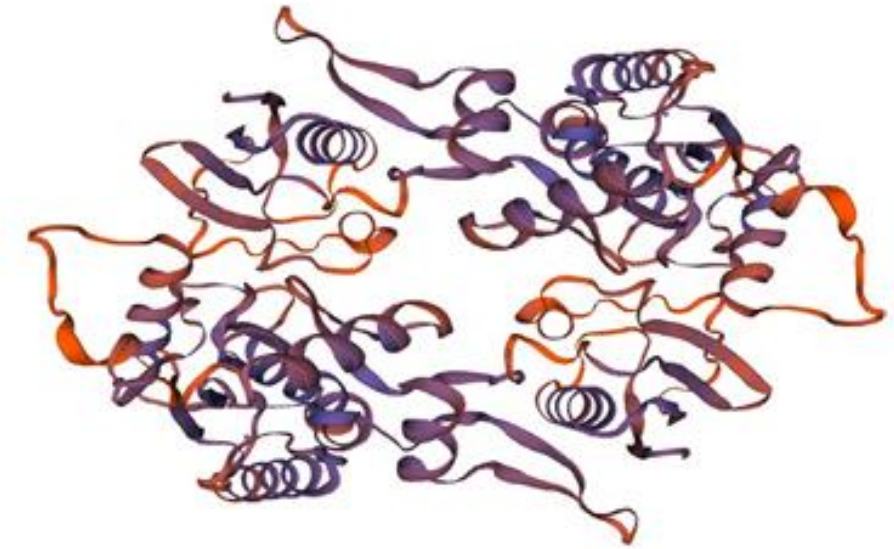


Fig. 2 Analysis of protein phylogeny

Predictions from **NCBI-CD** search indicated the presence of a **PHA02782** superfamily domain



Secondary structure and tertiary structure of HP

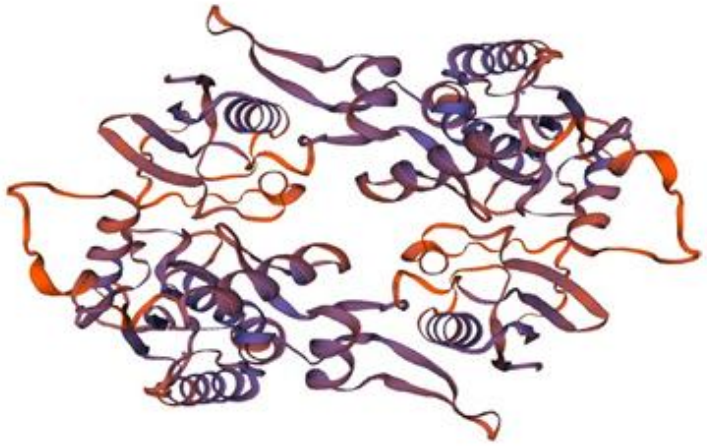
Table 2 3D Structure Generation with Different Servers

Webservers	Methodology	Accuracy
AlphaFold [62]	Neural network/Deep learning	70–75% [62]
I-Tasser (Iterative Threading ASSEmbly Refinement) [63]	Multiple-threading alignments, ab initio, refinement	65–75% [63]
PDB (Protein Data Bank) [64]	Experimental data	60–70% [64]
ORION [65]	Machine learning, hybrid modeling	65–70% [65]
ESyPred3D [66]	Automated homology modeling	60–75% [66]
SWISS-MODEL [37]	Homology modeling	80–90% [37]

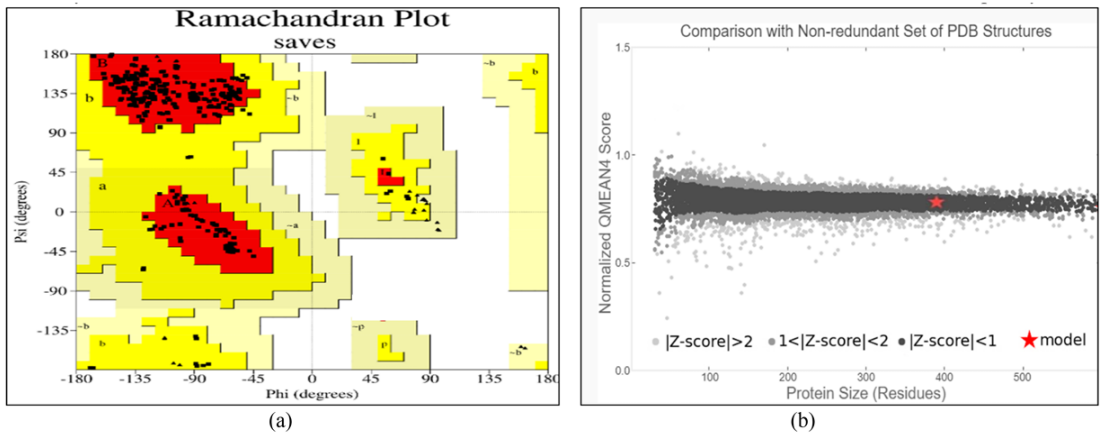
HP 3D structure matched 2D prediction and aligned with expected fold, suggesting similar function to its family proteins

Protein structure analysis

Objective: To evaluate the accuracy of the anticipated three-dimensional structural model of the hypothetical protein



- 92.3% residues in most favored regions
- Remaining residues in additionally allowed regions
- No residues in disallowed regions
- Z-score of −6.79 for the model



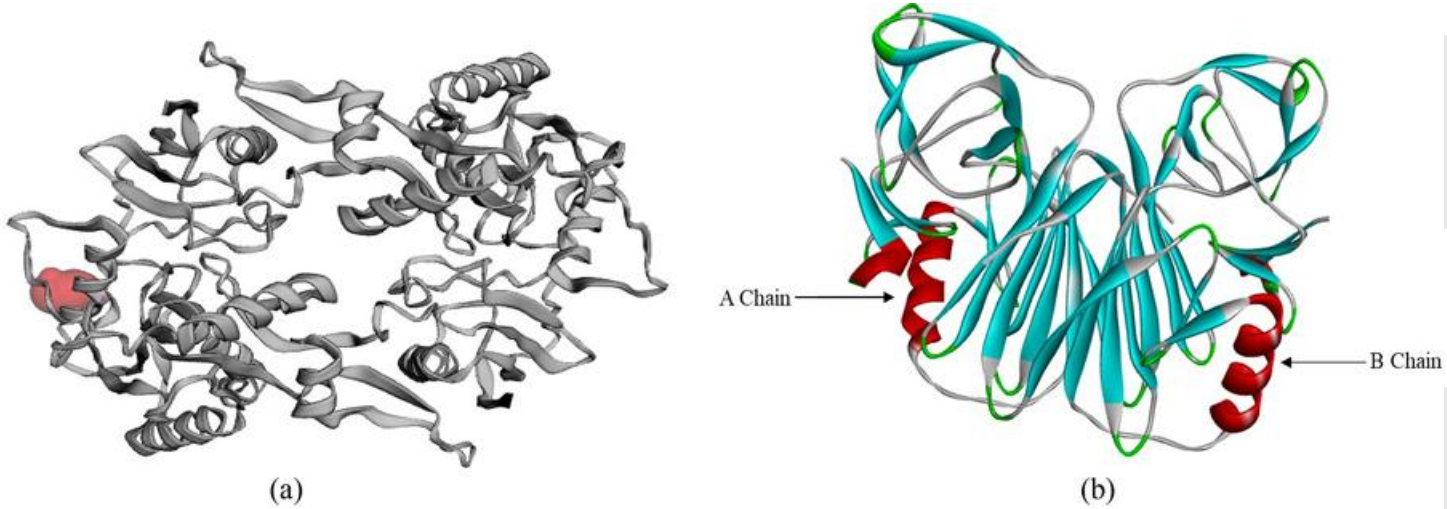
Supports reliability of homology modeling

Model validation for HP

Active site determination

Objective: To analyze the tertiary configurations of protein and identify the binding pocket

Visualization of active sites CASTp server



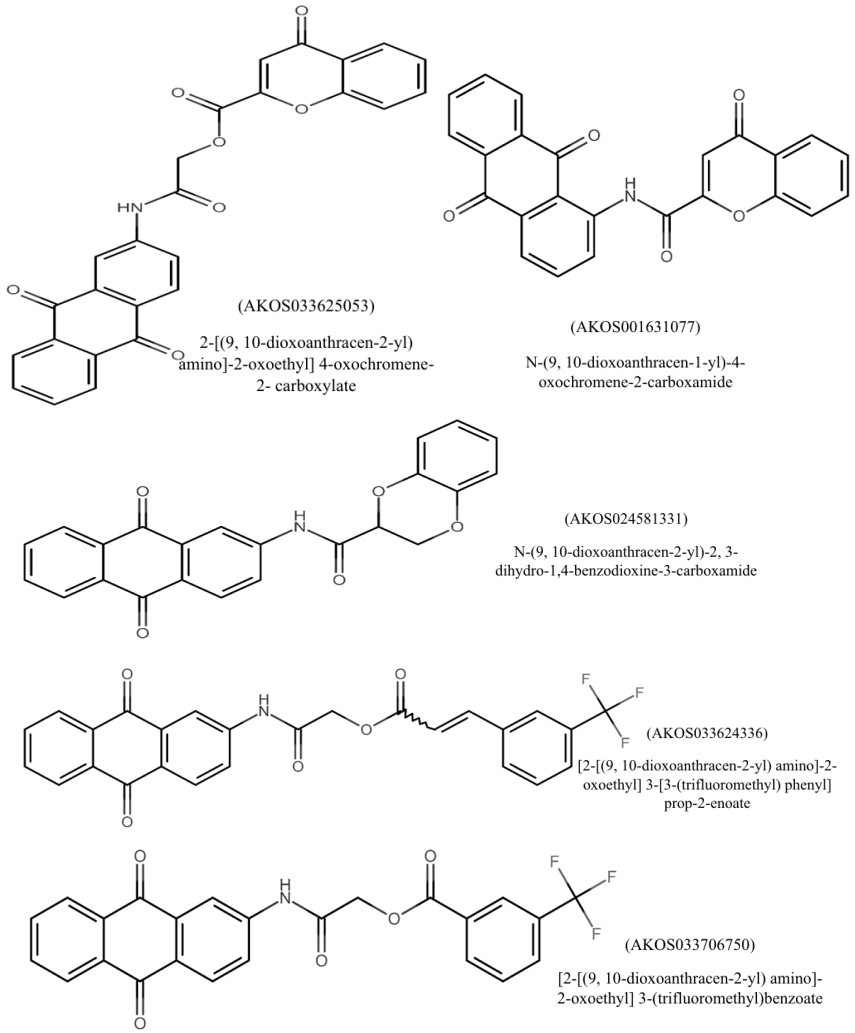
Chain B contains stable, well-defined active sites, making it the primary target for docking and inhibitor design

VaxiJen score of 0.4911
Four conserved peptides passed the score

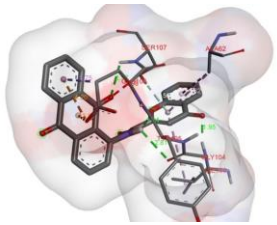
- Largest pocket: surface area 86.6 Å², volume 57.9 Å³
- Key residues in Chain B: TYR72, LEU73, GLN74, ILE75, VAL90, THR91, TYR92, ARG93, LYS95, TYR98, PRO190, ILE191, PRO192
- Chain B showed higher activity due to more stable binding sites
- Active residues align with PHA02782 superfamily domain

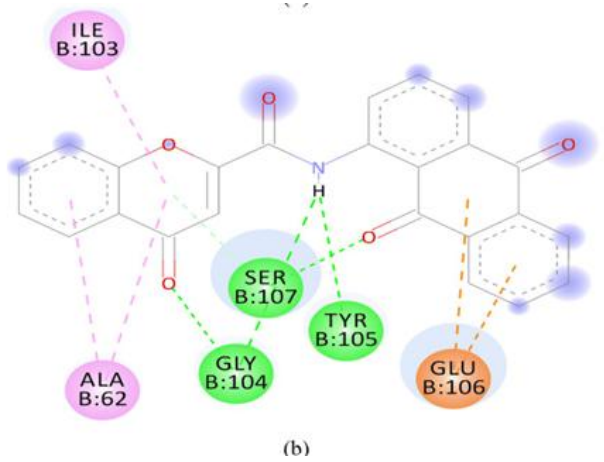
Docking of MPXV-Congo with drugs

Objective: To select the candidate ligands and preparation for docking



- Blide docking and site-specific docking with 500 compounds listed in Pubchem
- 5 lead phytochemicals selected (binding energies −9.2 to −8.8 kcal/mol) depending on docking scores, binding within active sites, and the manner of interactions





Conventional **Hydrogen Bond,** Pi–Anion, Pi–Sigma, Pi–Alkyl bonds were observed between AKOS001631077 and MPXV

Redocking, PK and Toxicity Analysis

Table 6 Binding Energies of Conventional Drugs

No.	Traditional compounds ID	Chemical name	Binding energy (Kcal/mol)
1	60,613	Cidofovir (Control)	−5.2
2	483,477	Brincidofovir	−5.5
3	16,124,688	Tecovirimat	−4.1
4	135,398,740	Ganciclovir	−5.1
5	54,682,461	Tipranavir	−3.8

Displays the binding energies of the conventional medicines indicating lead phytochemicals exhibited stronger binding affinities to the target protein compared to the standard drugs

Table 7 Pharmacokinetics (ADMET) analysis of top drug compounds

	Absorption				Distribution		Metabolism	
	Water solubility (LogS)	GI Absorption	P-gp Substrate	Skin Permeation(cm/s)	BBB Permeation	Log VDss	CYP2D6 Inhib-itor	CYP2D6/CYP3A4 Substrate
Drug compounds								
Threshold value	≥0 Highly soluble; 0 to −2 soluble; −2 to −4 moderately soluble	GI Absorption > 90 High		log Kp > 2.5 Low; log Kp < 2.5 High	BBB > 0.3 cross BBB, < 1 poorly dispersed to the BBB	VDSS: < 0.15 low distribution; VDss > 0.45 High tissue distribution of drug		
AKOS033625053	−2.66	High	No	−2.50	No	1.01	Yes	No
AKOS001631077	−2.88	High	No	−2.39	No	1.23	Yes	No
AKOS024581331	−2.91	High	No	−5.18	No	1.14	Yes	No
AKOS033624336	−4.77	High	No	−3.13	No	1.57	Yes	No
AKOS033706750	−5.43	High	No	−7.25	No	1.32	Yes	No

All the compounds have the good pharmacokinetic properties except of the compound 5 which has lower soluble activity in water and lower skin permeability.

Redocking, PK and Toxicity Analysis

Table 8 Lipophilicity, TPSA, Toxicity and Excretion Analysis of Top Drug Compounds

Drugs	(Log $P_{o/w}$)	TPSA (Å^2)	Toxicity	PAINS	Lipinski rule Violation (5 rules were fulfilled)	Synthetic accessibility	Molecular weight (g/mol)	Excretion Total clearance (log ml/min/kg)
AKOS033625053	3.05	98.31	No	0	No	3.42	453.4	0.65
AKOS001631077	1.51	112.14	No	0	No	3.17	395.4	0.43
AKOS024581331	2.21	119.33	No	0	No	3.31	385.4	0.48
AKOS033624336	2.07	112.18	No	0	No	3.10	479.4	0.46
AKOS033706750	1.54	120.38	Mild	0	No	3.05	453.4	0.47

Table 9 Inhibitory effects and Bioactivity (score from Molinspiration) of Top Drug Compounds

Natural compound	GPCR ligand	Ion channel modulator	Kinase inhibitor	Nuclear receptor ligand	Protease inhibitor	Enzyme inhibitor	Bioactivity score
AKOS033625053	0.21	−0.08	−0.36	−0.01	0.09	0.05	0.16
AKOS001631077	0.04	−0.41	0.04	0.59	−0.17	0.34	0.34
AKOS024581331	0.11	−0.06	0.02	0.12	0.06	0.01	0.28
AKOS033624336	0.08	−0.03	0.01	0.09	0.04	0.02	0.38
AKOS033706750	0.02	−0.10	0.01	0.31	0.10	0.02	0.22

The selected drugs demonstrated significant efficacy in binding to G protein-coupled receptors (GPCRs) and displayed notable inhibitory effects on proteases and enzymes. Compound 1 and 2 were chosen for further analysis.

Conclusion

.

- **Integrated bioinformatics and computational approaches** effectively create and validate the protein model and identify active sites or binding pockets and identify potential antiviral compounds against MPXV.
- **Multiple drug candidates** demonstrated promising binding affinity, pharmacokinetic properties, and drug-likeness criteria.
- In silico analysis supports the **potential repurposing of existing drugs** as effective treatments for monkeypox.
- Highlights the value of combining **structural modeling, virtual screening, and simulation** in accelerated drug discovery.

Criticism and application

	Paper 1	Paper 2
Strong	Establish the bioinformatic workflows for the WGS analysis for the large DNA viruses with high divergent sequences	Use the varieties of comprehensive In-Silico tools to characterize and identify the hypothetical proteins, and screen the potential inhibitors
Weak	The key finding were not approved by laboratory experiment	Not clear how they select HP

Genomics	Identify the variants in highly diverse sequences in large DNA viruses	Properties analysis of the proteins

Acknowledgement

Advisor

Asst. Prof. Supranee Phanthanawiboon

SP lab Member

Khon Kaen University Scholarship for ASEAN & GMS Countries' Personnel

Medical Microbiology

MDKKU
คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
FACULTY OF MEDICINE KHON KAEN UNIVERSITY


MANY
THANKS